Ibero-Amerika Institut für Wirtschaftsforschung
Instituto Ibero-Americano de Investigaciones Económicas
Ibero-America Institute for Economic Research
(IAI)

Georg-August-Universität Göttingen
(founded in 1737)

Diskussionsbeiträge · Documentos de Trabajo · Discussion Papers

Nr. 187

Micro Level Estimation of Income
Simulated welfare mapping (poverty maps)
for Paraguay 1992 and 2002

Thomas Otter

April 2009

Platz der Göttinger Sieben 3 · 37073 Goettingen · Germany · Phone: +49-(0)551-398172 · Fax: +49-(0)551-398173
e-mail: uwia@gwdg.de · http://www.iai.wiwi.uni-goettingen.de

Micro Level Estimation of Income

Simulated welfare mapping (poverty maps) for Paraguay 1992 and 2002

1.1. Introduction

Recent theoretical and empirical advances have brought income and wealth distributions back into a prominent position in growth and development theories, and as determinants of specific socio-economic outcomes, such as health or levels of violence and related phenomenon of inequality. To improve empirical investigation, new techniques were required for the simulation of small scale welfare indicators, such as income and its related distribution. Elbers, Lanjouw and Lanjouw (2003) designed a statistical procedure to combine different types of data and take advantage of the detail in household sample surveys and the comprehensive coverage of a census. The method extends the literature on small area statistics (Ghosh and Rao (1994), Rao (1999)) by developing estimators of population parameters which are non-linear functions of the underlying variable of interest (for example per capita income) by deriving them from the full unit level distribution of that variable. The most famous output of these exercises is known as "poverty maps". The use of these poverty maps is an important poverty reduction policy implementation tool used for selecting the poorest villages in the country (or villages where the greatest number of poor people are), such as the programs Bolsa Escuela in Brasil, Progreso in Mexico, Puente in Chile, Bolsa Familia in Argentina, Bono de Desarrollo Humano in Ecuador or Tekopora in Paraguay; all of these conditional cash transfer programs, directly to extremely poor households.[1]

1 The first poverty map for Paraguay was built by Marcos Robles (2000) combining Population Census 1992 with household survey 1997/98 data using the methodology proposed in Hentschel et al (2000), although the Government did not start using this kind of tool until 2003. In 2003, the Government needed to update poverty maps urgently with census and survey data from 2002. The author of this paper carried out this update for the Social Ministry using Elbers et al (2003) methodology based on a 10% sample of census data (the only census sub-sample available by the end of 2003) and the 2002 household survey. The attained results were the input for the "Indice de Priorizacion de Gasto" IPG, a geographic targeting tool for household cash transfer programs. In 2004, the IPG ranking was updated by Marcos Robles and Horacio Santander with the entire census data from 2002 and 2003 household surveys. Although a number of poverty maps in Paraguay already exist, the results shown in this paper are the only ones that

In this paper, the method of Elbers, Lanjouw and Lanjouw (2003) is applied using Paraguayan data from 1992 and 2002 producing estimates with levels of precision comparable to those of commonly used survey based welfare estimates - but for populations down to less than 1,000 people living within the same village. This is an enormous improvement over survey based estimates, which are typically only consistent for areas encompassing hundreds of thousands, or even millions of households. Experience using the method in South Africa, Brazil, Panama, Madagascar, and Nicaragua suggest that the method is reliable (Alderman, *et. al*. (2002), and Elbers, Lanjouw, Lanjouw, and Leite (2004)).

1.2 Methodology and Data

1.2.1 The Basic Methodology

Paraguayan household surveys collect very detailed information on household characteristics, including its income level;[2] however, coverage is limited and only representative at a relatively large geographical unit. Then again, Paraguayan population census has a complete coverage of all households, but collects very limited information on household characteristics and no information on income. The methodology developed by Elbers, Lanjouw and Lanjouw (2003) attempts to combine the advantage of detailed information on household characteristics obtained from a household survey with the complete coverage of a population census.

By combining the respective strengths of survey and census data, the simulated-welfare mapping method aims to estimate welfare indicators for small administrative areas. The approach uses household survey data to estimate a model of per capita income (or any other household or individual-level indicator of wellbeing) as a function of variables that are available in both the household survey and the population census.

The resulting parameter estimates from this estimation procedure are then used in a simulation to predict per capita income for each household in the census. Using the predicted per capita income, household level measures of poverty and inequality are then calculated and aggregated for small areas, such as districts,

---

combine the entire 2002 census with 2002 survey data, and what is more, show the only research results on poverty maps using 1992 census and 1992 survey data.

2 Poverty estimates by income is the official poverty measurement in Paraguay, carried out and updated periodically by National Statistical Office (DGEEC). Official poverty lines (caloric consumption line for extreme poverty and basic family basket for moderate poverty line) are updated by inflation for 4 different regions in the country; Asuncion, Central Urban, Remaining Urban and Rural.

sub-districts, or villages. This explains the origin of the name ‘simulated-welfare mapping’ for the method.

Importantly, the method allows for the calculation of standard errors for either welfare measure estimated. This feature is critical in that it offers a means of assessing the statistical reliability of the estimates as well as of comparisons across estimates for different communities.

1.2.2 The Income Model

Following *Elbers et al.* (2001, 2003), the empirical model of household income is defined as:

$$\ln y_{vh} = E(y_{vh} \mid x_{vh}) + u_{vh} \tag{1.1}$$

where $\ln y_{vh}$ is the logarithm of per capita income of household *h* in village *v*, x_{vh} is a vector of observed characteristics of this household (including village level variables), and u_{vh} is the error term. Note that we assume u_{vh} is uncorrelated with x_{vh}. This model is simplified by using a linear approximation to the conditional expectation $E(y_{vh} \mid x_{vh})$ and decomposing u_{vh} into uncorrelated terms:

$$u_{vh} = \eta_v + \varepsilon_{vh} \tag{1.2}$$

where η_v represents a village level error term common to all households within the village, and ε_{vh} is a household specific error term. It is further assumed that η_v is uncorrelated across villages and ε_{vh} is uncorrelated across households.
With these assumptions, equation (1.1) reduces to

$$\ln y_{vh} = x_{vh}\beta + \eta_v + \varepsilon_{vh}. \tag{1.3}$$

Estimation of the parameters underlying this equation, in particular the vector of parameters β and the distributional characteristics of the error terms, can be done by using standard tools from econometric analysis (Elbers *et al.*, 2003).

1.2.3 The Implementation Procedure

The standard procedure to implement the simulated-welfare mapping method for creating a map of mean income by sub-national administrative unit consists of five steps:

Step 1: Matching Variables in the Survey and the Census

In order to obtain rigorous estimates of income levels of the households in the census, the explanatory variables selected in the income determination model have to exist and be measured in the same way in both the household survey and in the census. If the sample of the household survey was randomly selected and is nationally representative, the distribution of each explanatory variable in the household survey can be expected to be the same as its distribution in the census.[3]

Step 2: Selecting Explanatory Variables for the Income Model

The selection of the explanatory variables in the income model starts by running a regression of log per capita income, using the survey data base, on the matched variables identified in Step 1, as well as some variables that can be created from other variables such as the square and cube of household size. In order to obtain a robust specification, variables are only selected for inclusion in the model if they contribute significantly to the explanation of per capita income. Hence variables with low statistical significance are dropped from the model.[4]

After a promising set of variables has been selected in this way, the regression is run again and the residuals of this regression are saved. These residuals need to be scrutinized to check if there are some outliers in the observation. If indeed there are some residual values which are far out of the range of most residual values, then these observations must be checked for coding or other errors. Ultimately, it may be necessary to delete them from the data.

The village level variables are obtained from either the population census data aggregated at the village level (for example the total population or age means of household heads in each village) or from other administrative data sources. These survey and census data can be completed with other data sources, mostly administrative data, such as the existence of public schooling (number of schools in a district) or infrastructure (kilometers of asphalt roads). These variables are then grouped into several sets such as demographic variables, village infrastructure variables, and village economic variables.

3 As a matter of fact, only variables that have the same distribution in census and survey are selected for inclusion in the income prediction models.

4 There are two kinds of dropped variables. First there are dummy variables whose frequencies are < 0.03 or > 0.97, to be dropped (even if most of them are expected to be insignificant since they would show low variance). This is carried out in order to make sure that the values of the variables included in the model show some variance which can influence in the variance of predicted income. Second, all other variables which are not significant in regression are dropped in order to make the models as robust as possible.

The residuals of the last regression are then aggregated at the village level to calculate the mean of these residuals for each village. The variable selection is then carried out by running separate regressions of the village-level mean of residuals on each set of the village-level variables. The variables with significant t-values are selected as the candidates for inclusion in the income model.

The feasibility of including these candidate village-level variables in the income model is tested by running regressions of village dummy variable on these variables. One regression is run for each village dummy variable. If the coefficient of a certain variable in a regression is one, it shows that there is a perfect multicollinearity between this variable and the village dummy variable. This will happen if, for example, a village has a certain infrastructure while no other villages have, or on the other hand, all villages except one have a certain infrastructure. Such variables are necessarily excluded from the model.

Step 3: Estimating the Income Model

The result of step 2 is a complete specification of the income model, incorporating both household-level and village-level independent variables of the model. The next step is to test whether there is heteroscedascity in the data. This will determine the method to be employed to estimate the model. The first step to accomplish this is to estimate the model of equation (1.3) using Ordinary Least Squares (OLS) and save the residuals as variable $\hat{u}_{vh}$.

Based on equation (1.2) the residuals $\hat{u}_{vh}$ are then decomposed into uncorrelated components as:

$$\hat{u}_{vh} = \hat{u}_{v\bullet} + \left(\hat{u}_{vh} - \hat{u}_{v\bullet} \right) = \hat{\eta}_v + e_{vh} \tag{1.4}$$

To investigate the presence of heteroscedasticity in the data, a set of potential variables that best explain the variations in e_{vh}^2 are used to estimate the following logistic model:

$$\ln\left[\frac{e_{vh}^2}{A - e_{vh}^2} \right] = z_{vh}^T \hat{\alpha} + r_{vh} \tag{1.5}$$

where we take A as being equal to $1.05 * \max\{e_{vh}^2\}$, as in Elbers *et al.*, (2003). This specification puts bounds on the predicted variance of ε_{vh}^2.

In the case where homoscedasticity is rejected, a household specific variance estimator for ε_{vh} is calculated as:

$$\hat{\sigma}^2_{\varepsilon,vh} = \left[\frac{AB}{1+B}\right] + \frac{1}{2}\hat{\mathrm{V}}\mathrm{ar}(r)\left[\frac{AB(1-B)}{(1+B)^3}\right] \qquad (1.6)$$

where $B = \exp\left\{ z_{vh}^T \hat{\alpha} \right\}$.

The income model is then re-estimated using the Generalized Least Squares (GLS) method, employing the estimated variance-covariance matrix, $\hat{\Sigma}$ with a structure shown in (1.7), resulting from equation (1.6) and weighted by the population weight, l_{vh}. The estimated parameters, $\hat{\beta}_{GLS}$, and their variance,

$$\mathrm{Var}\left(\hat{\beta}_{GLS}\right),$$

are saved for use in the simulation.

$$\begin{bmatrix} \mathrm{var}(n_s)+\mathrm{var}(e_{sh}) & \mathrm{var}(e_{sh}) & \mathrm{var}(e_{sh}) & \mathrm{var}(e_{sh}) \\ \mathrm{var}(e_{sh}) & \mathrm{var}(n_s)+\mathrm{var}(e_{sh}) & \mathrm{var}(e_{sh}) & \mathrm{var}(e_{sh}) \\ \mathrm{var}(e_{sh}) & \mathrm{var}(e_{sh}) & \mathrm{var}(n_s)+\mathrm{var}(e_{sh}) & \mathrm{var}(e_{sh}) \\ \mathrm{var}(e_{sh}) & \mathrm{var}(e_{sh}) & \mathrm{var}(e_{sh}) & \mathrm{var}(n_s)+\mathrm{var}(e_{sh}) \end{bmatrix} \quad (1.7)$$

Step 4: Simulations on Census Data

The purpose of this step is to apply the parameters estimated in the previous step to the census data. However, since the values of these parameters are obtained through estimations, they are not the precise values of these parameters and subject to sampling error. This needs to be taken into account when applying the parameters to the census data by taking into account the sampling error of the coefficient estimates. For a start, recall that the purpose is to calculate the simulated version of equation (1.3):

$$\ln y_{vh}^s = x_{vh}\beta^s + \eta_v^s + \varepsilon_{vh}^s \qquad (1.8)$$

where the superscript *s* refers to simulated version of each parameter or variable and now x_{vh} refers to characteristics of the households in the population census data.

Simulation of β

The simulated value of β is attained through a random draw, assuming

$$\beta \sim N\left(\hat{\beta}_{GLS}, \mathrm{Var}\left(\hat{\beta}_{GLS}\right)\right).$$

Note that the draw has to take into account the covariance across *β's*. The randomly drawn parameter is defined as β^s. The next step is to then apply this simulated parameter to each household in the census data to calculate the value of $x_{vh}\beta^{s}$.

Simulation of η_v

The process of obtaining the simulated value of η_v requires two steps of simulations. This is because the variance of η itself is estimated with error. Hence, the first step is to obtain the simulated variance of η, σ_{η}^{2s}. Elbers *et al.* (2003) propose to draw σ_{η}^{2s} from a gamma distribution:

$$\sigma_{\eta}^{2} \sim G\left(\hat{\sigma}_{\eta}^{2}, \hat{\mathrm{V}}\mathrm{ar}\left(\sigma_{\eta}^{2}\right)\right).$$

Consequently, a random draw of the variance for the whole sample is exercised and its mean is defined as σ_{η}^{2s}. Then the second step is to randomly draw η_{v}^{s} for each village in the census data, assuming $\eta_v \sim N\left(0, \sigma_{v}^{2s}\right)$.

Simulation of ε_{vh}

The process of obtaining the simulated value of ε_{vh} requires the use of the estimation results of equation (1.5). Assuming

$$\alpha \sim N\left(\hat{\alpha}, \mathrm{Var}\left(\hat{\alpha}\right)\right),$$

a random draw of α is made and defined as α^{s}. As in the case of β, the draw has to take into account the covariance across α's. The simulated parameter is then used to simulate the household specific variance estimator for ε_{vh} as defined in equation (1.6) for each household in the census data. Finally, the simulated value of household specific idiosyncratic error, ε_{vh}^{s}, for every household in the census data is obtained by taking a random draw, assuming $\varepsilon_{vh} \sim N\left(0, \sigma_{vh}^{2s}\right)$.[5]

Collecting

Now all three components of equation (1.7) have been simulated, the value of $\ln y_{vh}^{s}$ for all households in the census data can be calculated by summing up the values of $x_{vh}\beta^{s}$, η_{v}^{s}, and ε_{vh}^{s} that have been obtained. The whole set of simulations is then repeated a number (150 in our case) of times, so that in the end a database of 150 simulated values of (log) per capita household income of all the households in the census data is created. This is mainly to see if there

5 Elbers *et al.* (2003) mention alternatives for the assumption that the error component terms follow normal distributions. In separate sets of simulations we have experimented with these alternative assumptions. In no case did this lead to significantly different results.

variance within these 150 simulations in this fixed effects exercise is acceptably small.

Step 5: Calculation of Poverty and Inequality Indicators

The final output of Step 4 is a database of 150 simulated values of household income of all households in the census data. This database is used as the basis for calculating point estimates and standard errors of various poverty and inequality measures at the department, district and village levels. The point estimate of each measure is the mean of the calculated measure over the 150 simulated household incomes. Meanwhile, the standard error of this estimate is equal to the standard deviation of the calculated measure over the 150 simulated household incomes. The welfare indicators of a region – at any level – are calculated directly from the data of all individual households residing in that region.[6]

1.2.4 Data Sources

Four sources of data were used: (i) Encuesta de Condiciones de Vida (ECV) 1992 (ii) Censo Nacional de Población y Vivienda (CNPV) 1992, (iii) Encuesta Permanente de Hogares (EPH) 2002, and (iv) CNPV 2002. Both census and EPH 2002 were carried out by the Paraguayan National Statistical Office DGEEC (Direccion General de Estadistica, Encuestas y Censos), while the 1992 household survey was carried out by National University and the Inter American Development Bank.

Both surveys are representative household surveys, covering all areas of the country, with representative results for four different regions, Asuncion, Central Urban, Remaining Urban and Rural.[7] In the 1992 survey, 5,059 households (22,257 individuals) were interviewed, while 3,789 households (17,600 individuals) were interviewed in 2002. In general, both surveys follow the general format of a World Bank Living Standards Measurement Survey. The

---

6 The application of this poverty mapping exercise from step 3 to 5 is implemented using a computer program called PovMap (Version 1.2.4, February 2005), developed by Qinghua Zhao at the World Bank. All other steps were carried out using SPSS 13.0.

7 The Asuncion region only includes the city of Asuncion, while Central Urban covers the urban areas of the most populated department of Paraguay, called “Departamento Central”. Most of these urban areas are direct neighbors of Asuncion, together forming a kind of metropolitan region, excluding Asuncion. Remaining Urban include all other urban areas except Asuncion and Central Urban. Their common characteristic is to be urban, although not building a continuous geographic area. Rural includes all the rural areas. 1992 and 2002 household survey exclude in their sampling the departments Boqueron and Alto Paraguay in Chaco region, both remote rural areas; on the one hand due to budget constraints and, on the other, because less than 3% of the population live within these two departments.

population censuses of 2002 and 1992 are respectively the sixth and fifth population census carried out in Paraguay, both by DGEEC, in a systematic and comparable way. The previous censuses were carried out in 1950, 1962, 1972 and 1982. All censuses are carried out during the month of August, and cover the entire population living within Paraguayan territory, including foreign residents.

1.3 Results

1.3.1 Regression Results

In order to estimate per capita income for every household in the census, a set of household head individual characteristics, such as age, years of schooling or economic activity, characteristics of the spouse and the family group, such as number of children, their schooling or economic activity, characteristics of the habitat, access to basic services and assets within the household are used. We also tried some local infrastructure data such as kilometers of asphalt road, number public schools, the existence of a post office, public transport or government organized market places in the district.[8] To reinforce empirical evidence on spatial effects we controlled the number of direct neighboring districts[9] and percentage of economic activity by different sectors in neighboring districts.

As the aim of the regression models is to predict as precise as possible per capita income for every household in the census, using coefficients from regressions based on household surveys, including only common variables from survey and census, household survey regressions results need not be understood as regressions on determinants of income, but as regressions on variables which are correlated with income. Variables that are correlated with income, and not only variables which determine income, are used in order to achieve good results.

---

8 Most of these happened to be insignificant and were excluded from the final models (see footnote 4). Only for the "Remaining Urban Area in 2002" variable did the Kilometer of asphalt roads in each district (ROAD) happen to be significant and was included in the final model (see Table 1.9).

9 The number of direct neighboring districts (NVD) for Paraguayan economy can be understood as a proxy for closeness to areas of higher economic dynamics. Four out of the five most economically important cities are border cities, three of these with twin cities on the other side of the border. For being border districts the number of direct neighbor districts is smaller than for other districts within the country. The NVD variable can be a proxy to measure local effects of these more dynamic border cities.

Table 1.1 Variable definitions for 1992 estimates

AECON	Years of schooling spouse	RADIO	Household has radio (dummy)
AEDUJ2	Years of schooling head squared	SANITA1	Household has latrine (dummy)
AEJEFE	Years of schooling household head	TECHO1	Roof of cement (dummy)
AEM12S2	Sum of years of schooling > 12 years sqrd	TIPO2	Habitat is a house (dummy)
AEM15S2	Sum of years of schooling > 15 years sqrd	VIVI1	Own habitat (dummy)
AEM18S	Sum of years of schooling > 18 years	VIVI2	Rented habitat (dummy)
AEM18S2	Sum of years of schooling > 18 years sqrd	AEDUJ2_1	Cluster mean of years of schooling head sqrd
AGUA1	Household connected to public network of water supply (dummy)	AEDUJ3_1	Cluster mean years of schooling head cubed
D3	Department Cordillera (dummy)	AEM18S_1	Cluster mean of sum of years of schooling > 18 years
D6	Department Caazapa (dummy)	AEM18S_2	Cluster mean of sum of years of schooling > 18 years squared
D7	Department Itapua (dummy)	AGUA1_1	Cluster mean of household connected to public network of water supply (dummy)
D8	Department Misiones (dummy)	CONAGR_1	Cluster mean of spouse employed or self-employed in agricultural sector
D10	Department Alto Parana (dummy)	D3_1	Cluster mean of Department Cordillera
EDADJE	Age of household head	D4_1	Cluster mean of Department Guaira
EDJ2	Age of household head squared	FUENTE_1	Cluster mean of water supply inside the house
EDJ3	Age of household head cubed	HELA_1	Cluster mean of household has refrigerator
FUENTE1	Water supply inside the house (dummy)	JAGRO_1	Cluster mean of head working in agriculture
HELA	Household has refrigerator (dummy)	JCOM_1	Cluster mean of household head working in commercial sector
JAGRO	Head working in agriculture sector (dummy)	JTRAN_1	Cluster mean of household head working in transport sector
JEFEOC	Head employed or self-employed (dummy)	LUZ_1	Cluster mean of household has electricity
JSERV	Household head working in service sector (dummy)	MAT182_1	Cluster mean of number of household members > 18 enrolled in education squared
LNTOTP	Log total number of persons in household	MAT183_1	Cluster mean of number of household members > 18 enrolled in education cubed
LUZ	Household has electricity (dummy)	MATM18_1	Cluster mean of number of household member > 18 years enrolled in education
MATTOT	Total number of household member enrolled in any educational institution	MATTOT_1	Cluster mean of total number of household member enrolled in education
MCOSER	Household has sewing-machine (dummy)	NDOR2_1	Cluster mean of number of bedrooms squared
NDOR	Number of bedrooms	NDOR3_1	Cluster mean of number of bedrooms cubed
NVD	Number of direct neighbor districts	OCUM18_1	Cluster mean of number of household members > 18 employed or self-employed
OCUM122	Number of household member > 12 years employed or self-employed squared	PEPS_1	Cluster mean of percentage of secondary employment in neighbor districts
OCUM123	Number of household member > 12 years employed or self-employed cubed	PET_1	Cluster mean of percentage of tertiary sector employment in district
OCUM183	Number of household member > 18 years employed or self-employed cubed	TECHO1_1	Cluster mean of roof of cement
PEA	Number of household member in economic activity	TIPO1_1	Cluster mean of habitat is a house
PEPT	Percentage of tertiary sector employment in neighbor districts	TIPO2_1	Cluster mean of habitat is an apartment
PES	Percentage of secondary sector employment in district	TOT3_1	Cluster mean of total number of household members cubed

Source: CNPV and ECV 1992

Table 1.2 Regression results Asuncion 1992

Var.	Coef.	Std.E.	t	Prob>|t|
Intercept	13.444	0.551	24.41	<.0001
AEDUJ2	0.002	0.000	9.600	<.0001
PEA	0.238	0.017	13.61	<.0001
LNTOTP	-1.262	0.051	-24.85	<.0001
VIVI1	0.250	0.034	7.454	<.0001
NDOR	0.152	0.016	9.415	<.0001
AECON	0.024	0.003	7.488	<.0001
FUENTE1	0.197	0.048	4.103	<.0001
EDJ2	0.000	0.000	2.390	0.017
RADIO	0.314	0.065	4.790	<.0001
MATTOT	0.056	0.014	4.138	<.0001
TECHO1	-0.332	0.085	-3.918	<.0001
AEM18S	0.014	0.002	5.570	<.0001
JSERV	-0.088	0.032	-2.708	0.0069
MCOSER	0.100	0.030	3.340	0.0009
AEM18S2	0.000	0.000	-4.541	<.0001
TIPO2	0.211	0.074	2.833	0.0047
AEDUJ3_1	0.000	0.000	3.916	<.0001
JAGRO_1	2.185	0.466	4.687	<.0001
FUENTE_1	-0.455	0.108	-4.204	<.0001
AGUA1_1	0.315	0.106	2.962	0.0031
AEM18S_2	0.000	0.000	-3.088	0.0021
NDOR2_1	0.030	0.007	3.983	<.0001
LUZ_1	-1.692	0.559	-3.025	0.0025
MAT183_1	0.417	0.169	2.474	0.0135
JCOM_1	-0.365	0.117	-3.117	0.0019
MATM18_1	1.975	0.573	3.449	0.0006
MAT182_1	-1.989	0.652	-3.051	0.0023
Nr. of obs.		1,346		
Adj. R-sqrd.		0.636		
Cluster		89		
$var(\hat{\eta}_s)$		0.084		
$var(var(\hat{\eta}_s))$		<0.001		
$var(\hat{e}_{sh})$		5.987		

Source: Author's calculations based on ECV 1992

Table 1.3 Regression results Central Urban 1992

Var.	Coef.	Std.E.	t	Prob>|t|
Intercept	11.351	0.371	30.599	<.0001
AEDUJ2	0.002	0.000	9.742	<.0001
PEA	0.381	0.027	14.230	<.0001
LNTOTP	-0.981	0.051	-19.268	<.0001
AGUA1	0.154	0.046	3.331	0.0009
HELA	0.208	0.054	3.835	0.0001
VIVI1	0.215	0.045	4.804	<.0001
NVD	-0.040	0.016	-2.517	0.0121
TECHO1	-0.309	0.086	-3.577	0.0004
NDOR	0.079	0.025	3.185	0.0015
OCUM123	-0.015	0.003	-4.924	<.0001
JAGRO	-0.453	0.126	-3.587	0.0004
OCUM183	0.013	0.003	3.773	0.0002
MCOSER	0.077	0.038	2.021	0.0437
AEM18S_1	0.080	0.019	4.192	<.0001
PET_1	1.014	0.308	3.290	0.0011
OCUM18_1	-0.372	0.096	-3.881	0.0001
TOT3_1	0.001	0.000	2.396	0.0168
HELA_1	-0.591	0.218	-2.717	0.0068
NDOR3_1	0.015	0.004	3.590	0.0004
MATTOT_1	-0.196	0.060	-3.255	0.0012
CONAGR_1	10.986	3.456	3.179	0.0015
AEM18S_2	-0.001	0.000	-2.103	0.0359
Nr. of obs.		679		
Adj. R-sqrd.		0.617		
Cluster		46		
$var(\hat{n}_s)$		0.056		
$var(var(\hat{n}_s))$		<0.001		
$var(\hat{e}_{sh})$		5.676		

Source: Author’s calculations based on ECV 1992

Table 1.4 Regression results Remaining Urban 1992

Var	Coef.	Std.E.	T	Prob>|t|
Intercept	10.721	0.167	64.175	<.0001
AEJEFE	0.036	0.004	8.757	<.0001
PEA	0.370	0.023	16.047	<.0001
LNTOTP	-0.990	0.039	-25.44	<.0001
HELA	0.183	0.031	5.897	<.0001
AECON	0.026	0.004	6.850	<.0001
EDADJE	0.006	0.001	6.250	<.0001
PEPT	0.837	0.233	3.595	0.0003
JAGRO	-0.403	0.047	-8.527	<.0001
D3	-0.171	0.047	-3.650	0.0003
VIVI2	-0.196	0.039	-5.085	<.0001
AGUA1	0.056	0.028	1.984	0.0474
JSERV	-0.176	0.034	-5.235	<.0001
D7	0.340	0.041	8.352	<.0001
D10	0.294	0.051	5.798	<.0001
SANITA1	-0.130	0.030	-4.322	<.0001
OCUM122	-0.032	0.007	-4.430	<.0001
NDOR	0.070	0.014	4.947	<.0001
AEM15S2	0.000	0.000	-2.385	0.0172
MATTOT	0.034	0.011	3.056	0.0023
AEM18S	0.010	0.003	3.723	0.0002
PES	0.731	0.165	4.429	<.0001
D6	0.478	0.199	2.402	0.0164
AEDUJ2_1	0.002	0.000	5.173	<.0001
TIPO1_1	0.437	0.134	3.252	0.0012
TIPO2_1	0.937	0.346	2.710	0.0068
PEPS_1	-0.532	0.334	-1.594	0.111
Nr. of obs.		1997		
Adj. R-sqrd.		0.604		
Cluster		157		
$var(\hat{n}_s)$		0.079		
$var(var(\hat{n}_s))$		<0.001		
$var(\hat{e}_{sh})$		5.870		

Source: Author's calculations based on ECV 1992

Table 1.5 Regression results Rural 1992

Var	Coef.	Std.E.	t	Prob>|t|
Intercept	10.795	0.126	85.638	<.0001
NDOR	0.100	0.025	4.010	<.0001
LUZ	0.452	0.059	7.626	<.0001
PEPT	2.975	0.352	8.462	<.0001
LNTOTP	-0.930	0.047	-19.815	<.0001
HELA	0.227	0.056	4.057	<.0001
D8	-0.480	0.106	-4.527	<.0001
PEA	0.183	0.031	5.874	<.0001
JAGRO	-0.403	0.057	-7.123	<.0001
OCUM122	-0.026	0.006	-4.585	<.0001
AEDUJ2	0.002	0.001	2.662	0.0079
JEFEOC	0.358	0.084	4.267	<.0001
EDJ3	0.000	0.000	3.557	0.0004
MCOSER	0.113	0.046	2.448	0.0146
AEM15S2	-0.001	0.000	-3.945	<.0001
AEM18S	0.022	0.005	4.123	<.0001
AEM12S2	0.000	0.000	3.853	0.0001
TECHO1_1	0.715	0.147	4.861	<.0001
D4_1	-0.330	0.060	-5.544	<.0001
NDOR3_1	0.009	0.003	3.219	0.0013
JTRAN_1	-3.436	0.971	-3.537	0.0004
D3_1	-0.188	0.089	-2.122	0.0341
Nr. of obs.		978		
Adj. R-sqrd.		0.574		
Cluster		31		
$var(\hat{n}_s)$		0.075		
$var(var(\hat{n}_s))$		<0.001		
$var(\hat{e}_{sh})$		5.012		

Source: Author's calculations based on ECV 1992

Table 1.6 Variable definitions 2002

AEJ	Years of schooling of household head
AG_POCB	Water source is a well with pump (dummy)
AIRE	Household has air conditioning (dummy)
ALB	Household with at least one illiterate (dummy)
ANTENA	Household has an antenna for satellite TV (dummy)
AUTO	Household has a car (dummy)
BA_REC	Household with public removal of garbage (dummy)
CABLE	Household has cable TV dummy)
CEL	Household has at least one cell phone (dummy)
CO_GAS	Household cooks with gas (dummy)
CO_LENA	Household cooks with firewood (dummy)
DEP_AMA	Department Amambay (dummy)
DEP_CAAG	Department Caaguazu (dummy)
DEP_SANP	Department San Pedro (dummy)
EDADJ	Age of household head
ESPV	Percentage of primary sector employment in neighboring districts
HELA	Household has refrigerator (dummy)
JOCAGR	Household head working in agriculture sector (dummy)
JOCMAC	Household head working as machine operator (dummy)
LNTOTP	Log total number of persons in household
LUZ	Household has electricity (dummy)
MOTO	Household has a motorcycle (dummy)
NBIVIVI	Household with unmet basic needs in housing (dummy)
NHIJOS	Total number of children in household
NM15OC	Number of household members > 15 year occupied
NPER_PIE	Number of individuals per bedroom
NPIEZ	Number of rooms in habitat
OCJ	Household head employed or self employed
PA_LAD	Habitat walls of brick (dummy)
PC	Household has a PC (dummy)
PI_LAD	Habitat with floor of brick (dummy)
PI_TIER	Habitat with floor of earth
ROAD	Km of asphalt roads in district
TE_ZIN	Habitat with roof of zinc (dummy)
TEL	Household with telephone line (dummy)
VDVD	Household has a DVD player (dummy)
ANTENA_1	Cluster mean of household with an antenna for satellite TV
ASI18A_1	Cluster mean of number of household members 18 to 24 enrolled in education
BA_HOY_1	Cluster mean of households which trough their garbage in a hole
NBIVIV_1	Cluster mean of households with unmet basic needs in housing
TE_ZIN_1	Cluster mean of households with a roof of zinc

Source: CNPV and ECV 1992

Table 1.7 Regression results Asuncion 2002

Var.	Coef.	Std.Err.	t	Prob>|t|
Intercept	3.382	0.095	140.276	<.0001
TEL	0.370	0.061	6.050	<.0001
NPER_PIE	0.173	0.024	-7.132	<.0001
VDVD	0.266	0.061	4.391	<.0001
CABLE	0.247	0.061	4.064	<.0001
PC	0.259	0.067	3.872	0.0001
CO_LENA	-0.468	0.139	-3.370	0.0008
AEJ	0.012	0.006	2.161	0.0312
LNTOTP	-0.284	0.058	-4.934	<.0001
ALB	-0.140	0.058	-2.421	0.0159
Nr. of obs.		476		
Adj. R-sqrd.		0.612		
Cluster		51		
$var(\hat{n}_s)$		0.145		
$var(var(\hat{n}_s))$		<0.001		
$var(\hat{e}_{sh})$		4.938		

Source: Author's calculations based on EPH 2002

Table 1.8 Regression results Central Urban 2002

Var.	Coef.	Std.Err.	t	Prob>|t|
Intercept	12.677	0.095	132.741	<.0001
NPIEZ	0.113	0.013	8.365	<.0001
LNTOTP	-0.781	0.067	-11.517	<.0001
TEL	0.270	0.060	4.498	<.0001
AIRE	0.255	0.066	3.843	0.0001
AEJ	0.027	0.005	5.119	<.0001
NM15OC	0.209	0.018	11.164	<.0001
OCJ	0.356	0.048	7.415	<.0001
NHIJOS	-0.043	0.016	-2.613	0.0093
BA_REC	0.142	0.0456	3.158	0.0017
Nr. of obs.		495		
Adj. R-sqrd.		0.632		
Cluster		68		
$var(\hat{n}_s)$		0.087		
$var(var(\hat{n}_s))$		<0.001		
$var(\hat{e}_{sh})$		4.890		

Source: Author's calculations based on EPH 2002

Table 1.9 Regression results Remaining Urban 2002

Var.	Coef.	Std.Err.	t	Prob>|t|
Intercept	13.344	0.288	46.286	<.0001
PI_TIER	-0.226	0.061	-3.712	0.0002
PI_LAD	-0.197	0.051	-3.811	0.0001
NPIEZ	0.097	0.012	7.634	<.0001
LNTOTP	-0.702	0.034	-20.62	<.0001
HELA	0.240	0.042	5.705	<.0001
AUTO	0.221	0.043	4.947	<.0001
TEL	0.202	0.051	3.971	<.0001
CABLE	0.200	0.053	3.705	0.0002
PC	0.268	0.077	3.455	0.0006
AEJ	0.041	0.005	8.063	<.0001
OCJ	0.231	0.040	5.738	<.0001
JOCAGR	-0.281	0.059	-4.725	<.0001
ROAD	0.001	0.000	4.143	<.0001
EDADJ	0.005	0.001	3.799	0.0002
ESPV	-0.529	0.112	-4.73	<.0001
ANTENA_1	0.667	0.183	3.633	0.0003
ASI18A_1	-0.621	0.203	-3.049	0.0023
NBIVIV_1	-0.712	0.237	-3.005	0.0027
Nr. of obs.		1158		
Adj. R-sqrd.		0.576		
Cluster		136		
var($\hat{\eta}_s$)		0.083		
var(var($\hat{\eta}_s$))		<0.001		
var($\hat{e}_{sh}$)		5.416		

Source: Author's calculations based on EPH 2002

Table 1.10 Regression results Rural 2002

Var.	Coef.	Std.Err.	t	Prob>|t|
Intercept	12.306	0.112	109.256	<.0001
PA_LAD	0.164	0.036	4.527	<.0001
PI_TIER	-0.273	0.037	-7.236	<.0001
TE_ZIN	0.085	0.047	1.804	0.0715
LUZ	0.157	0.047	3.603	0.0003
AG_POCB	0.188	0.044	4.218	<.0001
NPIEZ	0.094	0.015	6.241	<.0001
LNTOTP	-0.564	0.054	-10.347	<.0001
CEL	0.173	0.045	3.783	0.0002
AIRE	0.972	0.130	7.475	<.0001
AUTO	0.270	0.062	4.324	<.0001
MOTO	0.155	0.044	3.519	0.0004
ANTENA	0.404	0.085	4.728	<.0001
CO_GAS	0.0731	0.050	1.438	0.1507
NHIJOS	-0.039	0.010	-3.912	<.0001
OCJ	0.321	0.050	6.356	<.0001
JOCAGR	-0.306	0.038	-7.881	<.0001
JOCMAC	0.281	0.094	2.968	0.003
NBIVIVI	-0.084	0.041	-2.023	0.0432
DEP_SANP	-0.228	0.053	-4.271	<.0001
DEP_CAAG	-0.215	0.046	-4.669	<.0001
DEP_AMA	-0.404	0.107	-3.749	0.0002
ESPV	-0.329	0.109	-3.015	0.0026
BA_HOY_1	0.645	0.171	3.76	0.0002
TE_ZIN_1	0.537	0.111	4.817	<.0001
Nr. of obs.		1537		
Adj. R-sqrd.		0.632		
Cluster		207		
$var(\hat{n}_s)$		0.125		
$var(var(\hat{n}_s))$		<0.001		
$var(\hat{e}_{sh})$		4.568		

Source: Author's calculations based on EPH 2002

All eight models produce acceptable results with most of the adjusted R sqrd. > 0.6; $var(\hat{n}_s)$ between 0.15 and 0.06; $var(var(\hat{n}_s))$ < 0.001 and $var(\hat{e}_{sh})$ between 4.6 and 5.9. For 1992 a higher number of significant variables (13 to 22) were identified. Additionally, 1992 models include between 4 and 11 cluster means variables, identifying local effects in the error term. In 2002, only 2 models include cluster means and the number of significant variables from census is much smaller. For 1992 and 2002, all coefficients from census variables have the expected signs, but not all the cluster means do. Interestingly, variables included to produce additional evidence on more extended spatial effects (passing district borders) produce the expected results. In 1992, NVD has a negative coefficient for Central Urban and percentage of tertiary sector employment in neighboring districts has a positive sign for Remaining Urban and Rural models. For 2002, the percentage of primary sector employment in neighboring districts has a negative sign for the Remaining Urban and Rural areas. Infrastructure data such as availability of public transport, education and health care institutions, post offices and public market places did not produce any significant results for income estimation. It was only for the Remaining Urban area 2002 that we found a significant and positive effect of kilometers of asphalt roads.

1.3.2 Poverty Estimates

The poverty measures calculated are the poverty headcount index (P0), poverty gap index (P1), and poverty severity index (P2) from the FGT family of poverty measures.[10] Meanwhile, the inequality measure calculated is the Gini ratio and General Entropy measures (GE0; GE1; GE2), as deciles mean incomes, but only P0 and Gini results are reported in this paper. In addition to the estimates of poverty and inequality indicators as usually presented, the results of the simulated-welfare mapping exercise also provide the standard errors of these estimates as a measure of their precision.

Tables 1.11 and 1.12 compare the estimated headcount poverty rate as reported in the household surveys in 1992 and 2002 and those estimated from the Population Census data (standard error of simulations in brackets, standard error as percentage of estimated poverty in squared brackets).

Table 1.11 Percentage of Poverty – 1992

	Survey				
	ASU	**CU**	**RU**	**RUR**	**TOTAL**
Extreme Poverty	4.4	11.2	16.1	28.2	20.4
Moderate Poverty	19.0	29.6	25.8	18.3	22.2
Total Poverty	23.4	40.8	41.9	46.5	42.6
Mean income	318,869	175,631	156,957	69,243	136,116
	Simulations				
Extreme Poverty	6.8	12.9	21.8	31.9	23.5
	(0.7)	(2.7)	(0.8)	(1.3)	(1.4)
	[10.3]	[20.7]	[3.8}	[4.1]	[5.8]
Moderate Poverty	16.6	27.2	25.1	17.6	20.8
Total Poverty	23.4	40.1	46.9	49.5	44.3
	(1.2)	(5.3)	(0.9)	(1.3)	(1.9)
	[5.1]	[13.2]	[1.9]	[2.6]	[4.3]
Mean income	335,400	188,143	145,187	79,710	142,519

Notes: Extreme Poverty Line 54,286 (ASU); 53,333 (CU); 47,500 (RU); 39,840 (RUR)
Poverty Line: 110,738 (ASU); 108,000 (CU); 89,000 (RU); 42,258 (RUR)

Source: Author's calculations based on ECV 1992 and CNPV 1992.

10 Foster *et al.* (1984).

For 1992, extreme poverty is higher in the Asuncion, Remaining Urban and Rural simulations, even when considering standard errors. Simulated mean income is higher than that observed in the Asuncion, Central Urban and Rural regions. This seems to be a consequence of differences in the income distribution between observed incomes in the survey and simulated incomes with census data. Overall poverty is almost the same except for the Remaining Urban area. We overestimate extreme poverty and underestimate moderate poverty but have an almost exact result for overall poverty except in Remaining Urban where simulated overall poverty exceeds observed poverty by 12%. General over-estimate is 4%. Since we overestimate mean income but nevertheless get higher poverty rates, there seem to be differences in the observed income distribution in survey and the simulated income distribution in census estimates.[11]

Table 1.12 Percentage of Poverty – 2002

	ASU	CU	RU	RUR	TOTAL
Survey					
Extreme Poverty	7.7	15.9	16.2	31.1	21.7
Moderate Poverty	21.5	37.9	22.7	19.4	24.6
Total Poverty	29.1	53.8	38.9	50.5	46.3
Mean income	664,097	374,642	353,579	188,998	317,063
Simulations					
	6.7	16.2	16.6	34.4	24.1
Extreme Poverty	(1.1)	(1.3)	(0.9)	(0.9)	(0.9)
	[16.4]	[8.0]	[5.4]	[2.6]	[3.7]
Moderate Poverty	23.2	39.6	23.9	19.9	24.5
	29.8	55.7	40.5	54.3	48.5
Total Poverty	(1.9)	(1.4)	(1.0)	(0.9)	(1.1)
	[6.4]	[2.5]	[2.5]	[1.7]	[2.3]
Mean income	725,053	375,413	355,013	186,634	322,467

Notes: Extreme Poverty Line 142,308 (ASU); 140,717 (CU); 106,802 (RU); 73,501 (RUR)
Poverty Line: 321,229 (ASU); 317,998 (CU); 197,895 (RU); 118,483 (RUR)

Source: Author's calculations based on EPH 2002 and CNPV 2002.

Considering standard errors, the estimates for extreme poverty for 2002 fit for all regions except the Rural area. In this model, simulated total poverty exceeds observed poverty in all areas by an average of about 5%. We are still overestimating mean income, although not as much as in 1992.

11 One possible source for these distribution differences is the fact that household observed household income in the survey "is not continuous" (this is that several of the higher centiles [>80] are empty [no observations]). Nevertheless, in simulation exercises there will be estimates for a "continuous distribution".

To understand what could be the possible reasons for the differences between observed results in the surveys and estimated results in census database, several tests were carried out. There is almost no difference between observed (survey) and simulated (census) covariate levels. Only in one (Central Urban 2002) out of the eight models in use, simulated census covariate levels differed from the observed ones in the survey in more than 1%. Furthermore, there is no problem with the residual assumption in survey regressions. In all cases, residuals have a mean zero and a low variance. Consequently, if there are no structural problems in the models themselves, one possible source for biases could be a sampling problem in the actual surveys, for example classification of households in urban and rural areas. In fact, for 2002 the maximum difference for population share by region between survey and census is 0.2%. However, in 1992 there is a classification problem for the Asuncion and Rural areas (Asuncion has 11.6% of population in census and 14.9% in survey; Rural area has 48.6% in census and 45.6% in survey). The maximum difference for Central Urban and Remaining Urban is 0.5%. Recall that the 1992 survey was the first nationwide household survey ever carried out in Paraguay, so there may have been some lack of experience in paying attention to all the details. This hypothesis is consistent with differences between 1992 and 2002 simulations, where biases in 2002 are much smaller than in 1992. The way sampling problems can introduce biases in the results seems to be through the variance of household specific error (e_{sh}). The variance of 1992 errors exceeds that of 2002 errors between 9 and 23 percent.

Table 1.13 Inequality measures

	Survey					Simulations				
	ASU	CU	RU	RUR	TOTAL	ASU	CU	RU	RUR	TOTAL
Gini income 1992	0.519	0.408	0.464	0.499	0.558	0.493 (0.021) [0.043]	0.439 (0.017) [0.039]	0.480 (0.007) [0.015]	0.547 (0.017) [0.031]	0.589 (0.015) [0.025]
Gini income 2002	0.437	0.419	0.446	0.534	0.561	0.444 (0.015) [0.034]	0.394 (0.011) [0.028]	0.460 (0.009) [0.019]	0.538 (0.011) [0.020]	0.510 (0.011) [0.022]

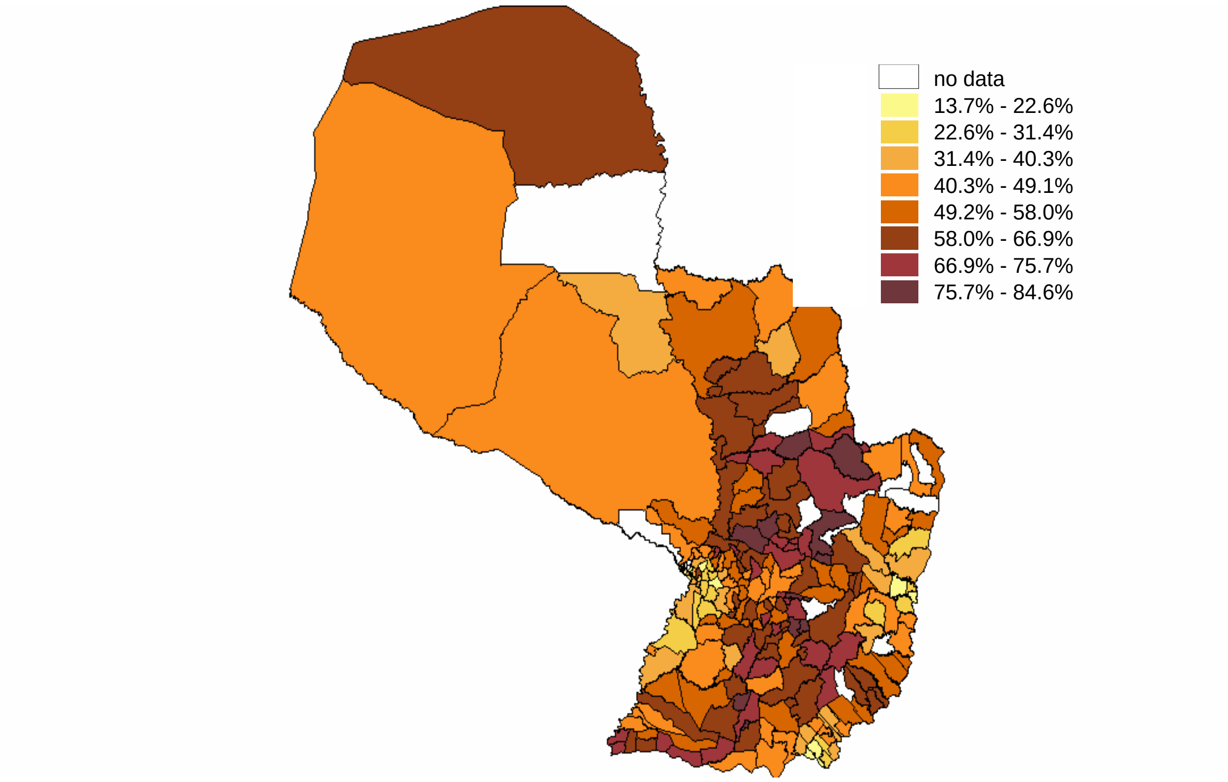
Source: Author’s calculations based on ECV 1992, CNPV 1992, EPH 2002 and CNPV 2002.

For 1992 we underestimate inequality for Asuncion but overestimate inequality in all other regions. These differences also seem to be a consequence of allocation differences between rural and urban areas in 1992, between census and survey, for the Asuncion and Rural areas. For 2002, the estimated inequality in the Asuncion and Rural areas are slightly higher than those observed and we underestimate inequality in the Central Urban and Remaining Urban areas.

1.3.3 Poverty and Inequality Maps

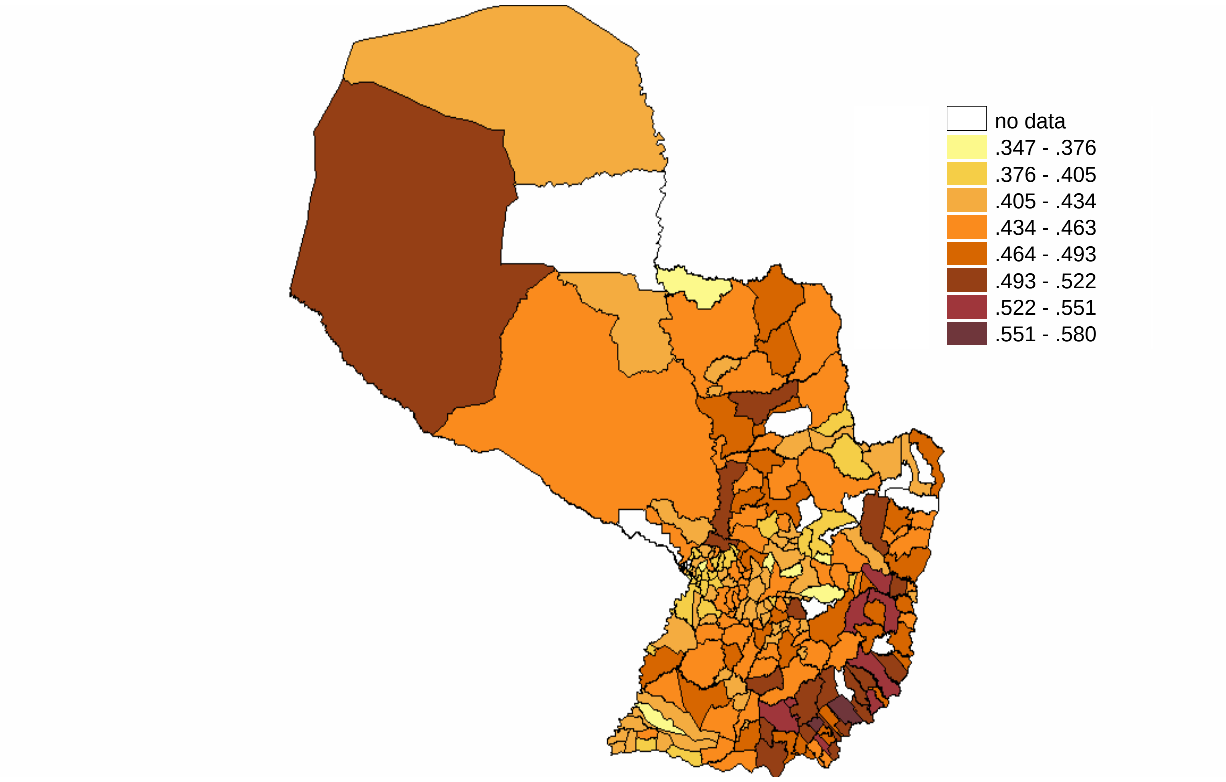
When examining these maps, it should be kept in mind that they have been created using the *expected* headcount. The *true* headcount for a location will differ from the expected headcount because of sampling and modeling errors. The maps do not take the errors into account.

Figure 1.1 FGT0 Per capita income 1992 at district level



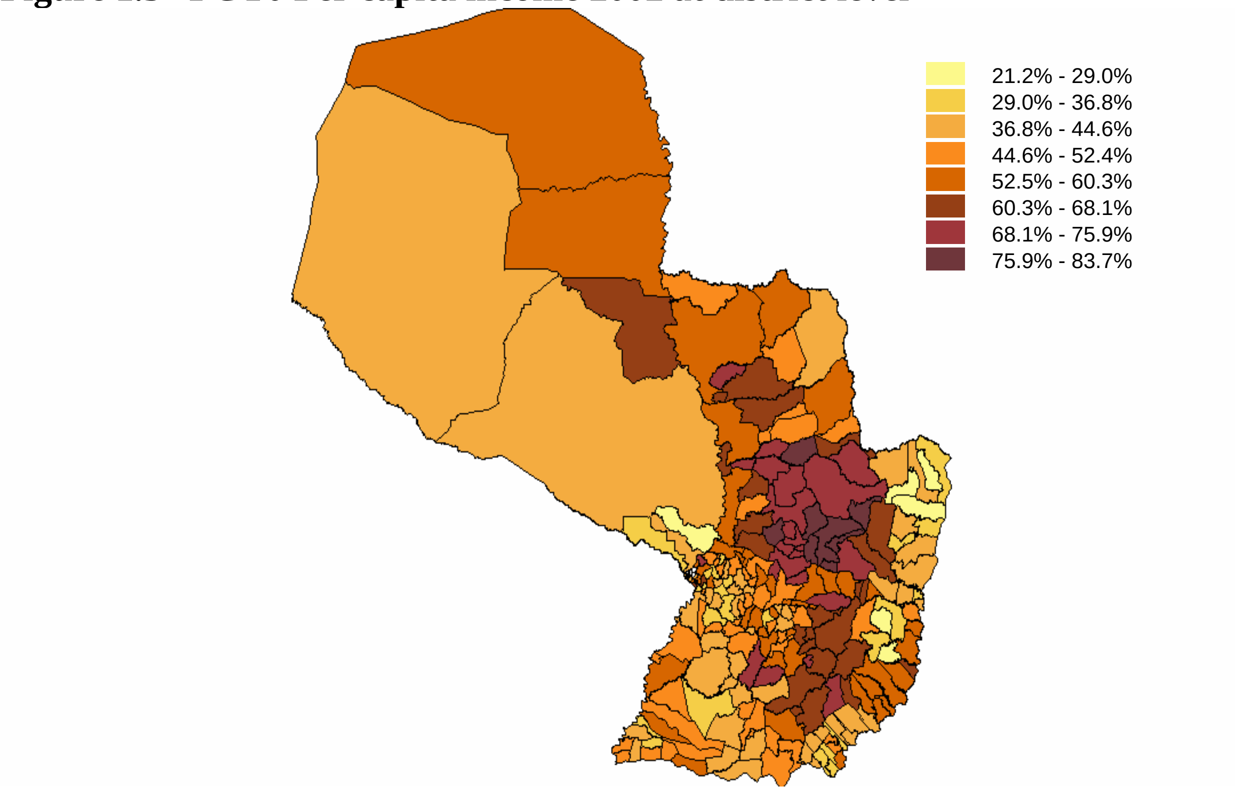
Source: Author's calculations based on ECV 1992 and CNPV 1992

Figure 1.2 Gini Per capita income 1992 at district level



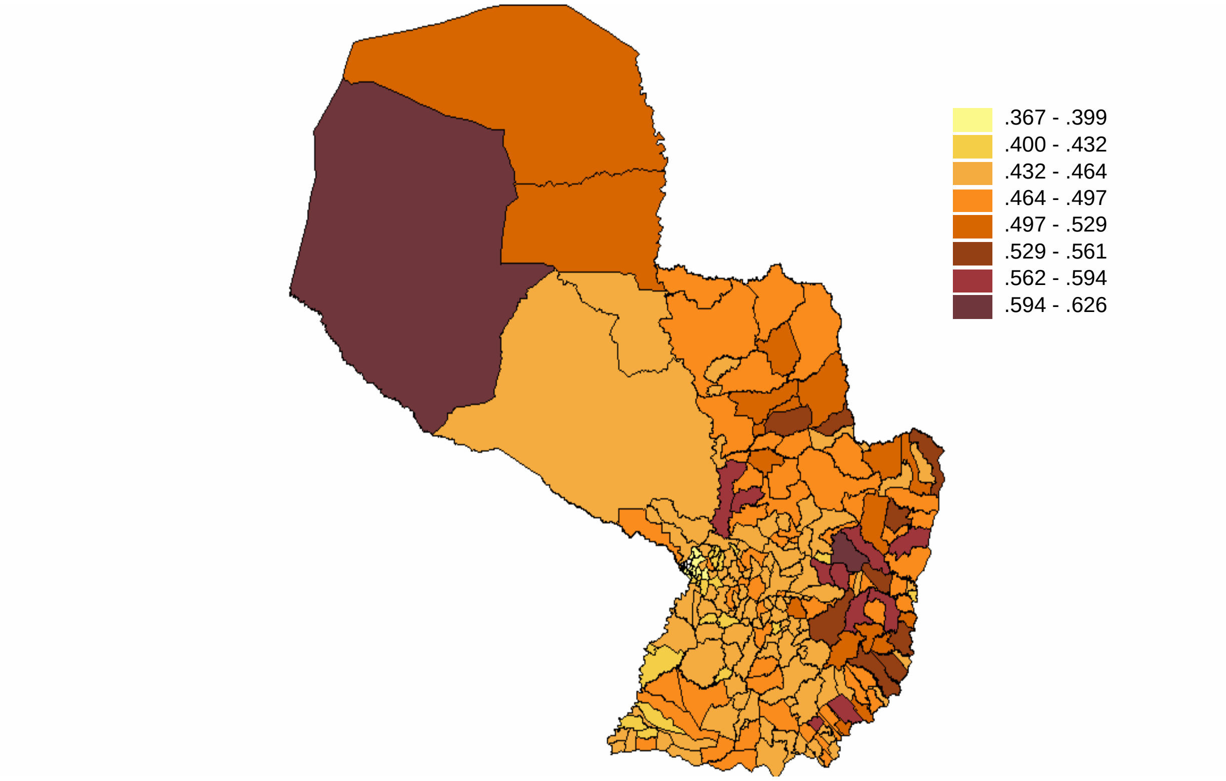
Source: Author's calculations based on ECV 1992 and CNPV 1992

Figure 1.3 FGT0 Per capita income 2002 at district level



Source: Author's calvulations based on EPH 2002 and CNPV 2002

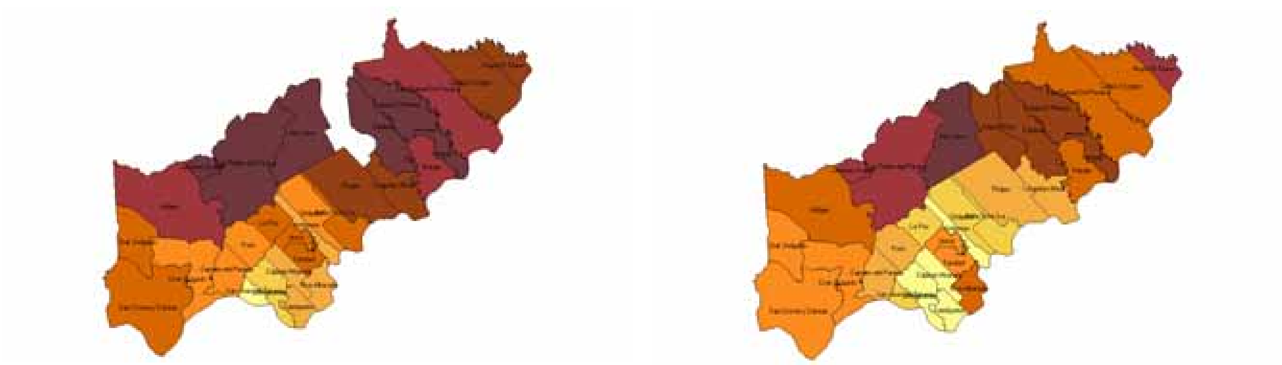
Figure 1.4 Gini Per capita income 2002 at district level



Source: Author's calculations based on EPH 2002 and CNPV 2002

Figure 1.5 FGT0 Per capita income Itapua department 1992 and 2002 at district level

1992 **2002**

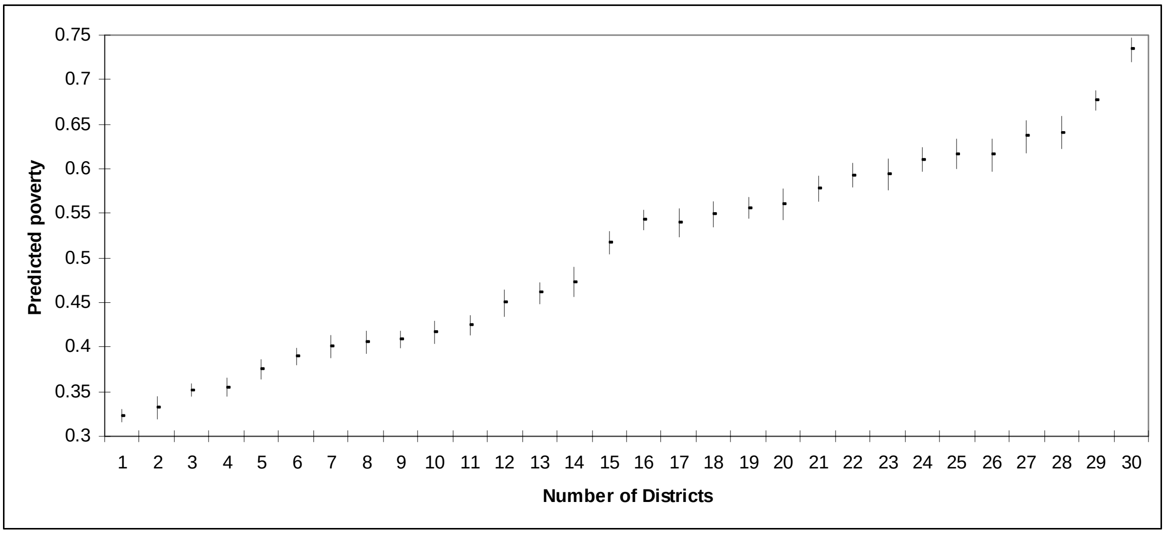


Source: Author's calculations based on ECV 1992, CNPV 1992, EPH 2002 and CNPV 2002.

The maps in Figure 1.5 show the heterogeneity of poverty levels in small areas, using the Itapua department example. Itapua is one of the most prosperous departments in Paraguay, considering its economic performance (important GDP growth driven by mechanized Soya agro-industry) between 1992 and 2002. Regarding departmental mean poverty levels in 2002, Itapua ranked 11 out of 18 departments, about 1 percentage point above the national poverty rate, despite its high GDP growth. What is behind this apparent contradiction can be seen by mapping Itapua's poverty levels by district. In 1992, there was a belt of poor, up to extremely poor districts in the north of the department (darker colors), with poverty rates of up to 68%, versus a more prosperous zone in the south (brighter colors) with poverty rates down to 14%. Until 2002, and as a consequence of GDP growth, poverty generally decreased in Itapua (more districts with bright colors), but the poor district belt in the north still remained. Some of these poor districts even increased their poverty levels, now up to 72% of the population. The lowest levels in the south were 29% en 2002. So, considering a pro poor policies intervention, maybe by conditional cash-transfer programs to extremely poor households, Itapua might not be selected for program participation, considering department mean levels of poverty. Nevertheless, disaggregating poverty estimates by districts, it happens to be that some of the poorest districts of the country are located in Itapua, being direct neighbors of some of the most prosperous districts in the country. Small welfare estimates help to improve the targeting of pro-poor policies.

To show an example of what precision can be achieved at the district level, Figure 1.6 shows the predicted poverty headcount in rural Itapua for 2002, along with a confidence interval from one standard error below to one standard error above the point estimate. The department of Itapua was selected because covers almost the complete range of standard errors for point estimates observed for the 1992 and 2002 exercise, varying from 0.015 to 0.075.

Figure 1.6 Rural poverty estimates Itapua 2002



Source: Author's calculations based on EPH 2002 and CNPV 2002

Apart from the standard errors for point estimates, regression models have structured and unstructured errors as seen above. To check for spatial patterns of these kinds of errors they are mapped in the annex, as well as relative changes in poverty and inequality.

As mentioned above, there is a set of former poverty map exercises in Paraguay. Robles (2000) combined the 1992 census with 1997/98 survey using the Hentschel et al (2000) method, which differs slightly from the method applied in this paper, so they can not easily be compared. The second poverty map exercise carried out by Otter (2003), using the same method applied in this paper, combines the 2002 survey with a 10% sub-sample of the census and combines the estimated poverty levels with weighted unmet basic needs percentages. Since the data bases are not the same, there may be some difficulties in comparing the results of this paper with the Otter (2003) exercise. Finally, Robles and Santanders (2004) poverty map exercise is most similar to this paper. Using the same method, they combine 2002 census with 2003 household survey data, mostly because the 2003 survey sample allows to run a separate regression model for every department (18 models) and not only 4 different models by region as in 2002. Since poverty rates changed considerably (dropping by 6 percentage points in the national mean) between 2002 and 2003, the best way to compare the results of these two exercises is to compare rankings of districts by poverty level, which should not change strongly, even if poverty percentages decrease considerably. When comparing the rankings, we observe that 64% of all districts are ranked within the same deciles, meanwhile the standard deviation of ranking differences is only 0.94 points. Consequently, the results of both poverty mapping exercises are consistent between each other, and differences should be a consequence of the more detailed estimates by Robles and Santander and poverty changes between these two years.

1.3.4 Pro-poor growth evidence

Although this paper is not about pro-poor growth, its results provide empirical evidence on such growth from poverty map exercises. Even if the existence of pro-poor growth evidence could easily be confirmed from the household survey data, doing this with simulated incomes based on census data will allow the identification of whether there are any spatial patterns in pro-poor growth, for example the concentration of a huge number of households benefiting from pro-poor growth which could be concentrated in a small and limited geographic area.

According to international organizations pro-poor growth is simply defined as economic growth that benefits the poor (e.g., UN 2000a; OECD 2001). This definition, however, provides little information on how to measure or how to implement it. What remains to be specified are, first, whether economic growth benefits the poor and, second, if this is the case, to what extent. Klasen (2004) provides more explicit requirements that a definition of pro-poor growth needs to satisfy. The first requirement is that the measure differentiates between growth that benefits the poor and other forms of economic growth, and it must answer the question by how much the poor have been benefited. The second requirement is that the poor must have benefited disproportionately more than the non-poor. The third requirement is that the assessment must be sensitive to the distribution of incomes amongst the poor. The fourth requirement is that the measure must allow an overall judgement of economic growth and not only focus on the gains of the poor.

To identify the existence of pro-poor growth of per capita income according to point estimates from the poverty map exercise, specific inflation rates by region and income deciles were calculated. To obtain these measures as realistic as possible they are based on consumption profiles by deciles, built up as a mean of 1997/98 and 200/01 consumption profiles[12] (only during these two periods did Paraguayan household surveys include a consumption module). Table 1.14 shows the deciles specific inflation rates. In general, inflation is lower for lower income deciles and lower for less urban or more rural areas. These results seem to drive the specific results of pro-poor growth.

Table 1.14 Mean inflation rates by deciles, Paraguay 1997/98 – 2000/01 (%)

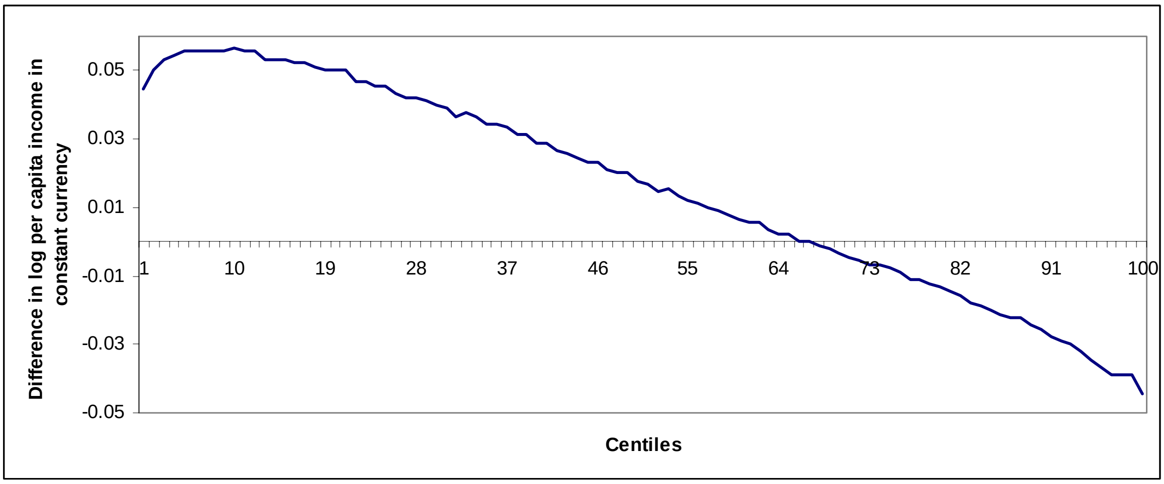
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
ASU	112.5	115.7	115.2	115.3	116.4	117.5	121.7	122.1	123.5	123.5
CU	114.4	114.9	114.4	114.7	115.3	115.7	117.7	118.0	120.7	120.5
RU	111.1	111.1	112.8	111.9	112.6	115.5	114.5	116.9	116.4	119.5
RUR	107.5	108.0	108.2	108.5	109.4	110.1	110.3	111.0	111.8	113.6
TOTAL	111.4	112.4	112.7	112.6	113.4	114.7	116.1	117.0	118.1	119.3

Source: Author’s calculations based on EIH 1997/98 and EIH 2000/01.

Figure 1.7 shows the growth incidence curve of log per capita income for constant currency in 1992 values. There is a clear pro-poor growth pattern for deciles 5 to 25. Growth incidence curves produced separately for the four different regions show that there is almost no pro-poor growth in the Asuncion and Central Urban areas, very few in Remaining Urban, but mostly in rural areas.

12 Carried out by groups of goods and services: food, clothing, housing, health, transport, education, various.

Figure 1.7 Growth incidence curve of log per capita income



Source: Author's calculations based on ECV 1992, CNPV 1992, EPH 2002 and CNPV 2002.

1.4 Discussion

In the regression results, all adjusted R squared are not very high; approximately 0.6. On the one hand, this is due to the fact that only variables whose coefficients are highly significant were included in order to make the models as robust as possible, as explained above. On the other hand, reduced levels or adjusted R squared may result from a considerable number of dummy variables in the models, which may have reduced the power of explanation for probably having lower variance than other kind of variables.

In regression results, it seems that the less homogeneous a population is, the higher the probability of identifying locational components in the error term. If there are locational components in the error term, it is easier to identify them in smaller geographic areas (this is why the Asuncion model produces more significant cluster effects than the other regions in 1992).

If this is the case, the question is now why the 2002 models produce much less significant cluster effects? Could it be true that the population became more homogeneous? Even if poverty and inequality changes between 1992 and 2002 are not big, the real story household survey data tell us that there was a poverty reduction between 1992 and 1997 and an increase between 1998 and 2002. Urban inequality tended to decrease while rural inequality tended to increase over the whole period. Additionally, we have a growing urban migration of poor and a growing urban poverty.

There are two main differences between the 1992 and the 2002 results. First, the 1992 results include much more significant cluster effects than those for 2002, which capture part of a locational effect. In 2002, this fact seems to have disappeared. Noticeably, the locational effect is not only or directly related to the geographic location, but is in relation to the population group and their characteristics, living in the observed area, at any moment. Even if the 2002 regression models include less variables and less significant cluster effects, their prediction power for poverty is higher than 1992 while their prediction power for inequality is almost the same as in 1992. Consequently, geographic location seems to be less important for 2002 than for 1992.

Although the income regression models are not modeling determinants, but variables which are correlated with income, most of these have the expected signs. This is also true for the cluster means. The interpretation of these cluster means, which try to capture part of the locational error, is difficult. Nevertheless, for some cluster mean variables there can be a kind of intuitive understanding. For example, in some models the percentage of primary sector employment appears as a significant variable with a negative sign for the household (individual) level. However, its cluster mean also has a positive sign. This may be understood as a positive effect at community level *ceteris paribus* and for the given mean level of income in the region.

At least concerning empirical evidence from Paraguay, the methodology seems to work better for the prediction of higher incomes, since extreme poverty is overestimated (by underestimating lower incomes), as are mean incomes. Consequently, the associated distributions are not the same as those observed in household surveys. Several reasons can be attributed for this. For example, rural incomes (where most of the low incomes are located) depend strongly on climate and market price changes not captured by variables included in a census. Lowest incomes in urban areas may be difficult to simulate correctly due to sampling and measurement problems in household surveys.

Poverty maps show that there is a concentration of poverty in the center of eastern Paraguay (where 98% of the population is living). Changes of poverty during the observation period neither altered significantly the spatial distribution of poverty nor of inequality. In general, structured and unstructured errors are higher in more rural areas. All these results are consistent with Paraguayan rural economic history over the period, with a crisis of small scale cotton cash crop farming and an increase of large scale Soya bean mechanized farming, deepening poverty and inequality in rural areas.

1.5 Conclusions

As shown, the method of Elbers et al is a reliable method for small area welfare estimates, producing poverty point estimates for sub-national levels. Obviously, there are several sources of errors in the methodology and other errors from sampling and measurement problems in the household surveys. Nevertheless, the income estimates are consistent with economic history in Paraguay and, since most of the errors made during the estimation procedure can be quantified, it is possible to determine their reliability. In any case, the gain in additional information is crucial for politics and policies design and implementation.

Poverty analysis is often based on national level indicators that are compared over time or across countries. The broad trends that can be identified using aggregate information are useful for evaluating and monitoring the overall performance of a country. For many policy and research applications, however, the information that can be extracted from aggregate indicators is not sufficient, since these hide significant local variation in living conditions within countries.

The detailed poverty maps of small administrative areas, that are the ultimate output of the simulated-welfare mapping method, provide benefits that help address the shortcoming of aggregate poverty analysis in the following ways:

(i) Poverty maps capture the heterogeneity of poverty within a country.
Almost all countries in the world have regions that are better off and others that are left behind. Such differences are often lost in national level statistics. Poverty maps can reveal the variation in local poverty levels when small area information is available. As shown, seemingly homogeneous regions can actually have a large degree of local heterogeneity.

(ii) Poverty maps improve targeting interventions.
In designing poverty reduction programs, resources can be used more effectively if the most needed groups can be better targeted. This reduces the leakage of benefits from a poverty reduction program to non-poor households and, on the other hand, reduces the risk that poor households will be missed by a program. This requires an adequate targeting to poor areas, but also a correct beneficiary selection.

(iii) Poverty maps can help governments – national and local – to articulate their policy objectives.
Basing allocation decisions on observed geographic poverty data, rather than subjective rankings of regions, increases the transparency of government decision making. Such data can thus help limit the influence of special interests in allocation

decisions. There is a related role for well-defined poverty maps to lend credibility to government and donor decision-making. By increasing transparency, poverty maps can help prevent the regional autonomy policy from being hijacked by the local elite.

(iv)Poverty maps have an important role in communicating information on welfare distribution to the civic population within a country.
Poverty maps are not only useful to governments and decision makers, but also to local communities. Compiling disaggregated information on human welfare generates locally relevant information. This provides local stakeholders with the facts that are required for local decision making and for negotiation with government agencies. Poverty maps thus become an important tool for local empowerment and decentralization.

(v) Poverty maps are useful for evaluating the impact of various programs.
Poverty maps offer opportunities to undertake detailed empirical research on the causal relationships between local poverty, income inequality, and various other social outcomes, both at the individual and community levels. Until now, scarcity of welfare indicators for small areas has prevented researchers from studying the relationship between various programs, poverty, inequality, and various outcomes, such as health, education, crime, and the environment. Poverty maps open up more opportunities for researchers to examine these relationships.

(vi) Estimation of small area indicators of poverty allows their incorporation into geographical information systems (GIS).
This feature of poverty maps facilitates the combination of poverty information with other indicators from policy-relevant subject areas. Examples are geographic databases of transport infrastructure, public service centers, access to input and output markets, or information on natural resources quality and vulnerability. Using geographic overlay techniques and spatial analysis methods, the newly constructed databases on poverty can thus be used to address a range of multidisciplinary questions. The databases can also be used by the private sector to guide them in determining the locations for new investment opportunities.

Bibliography

Bigman, D. and Fofack, H. (2000), Geographical Targeting for Poverty Alleviation: An Introduction to the Special Issue, *The World Bank Economic Review*, 14(1), pp. 129-145.

Demombynes, G., Elbers, C., Lanjouw, J., Lanjouw, P., Mistiaen, J. and Özler, B. (2002). Producing an Improved Geographic Profile of Poverty: Methodology and Evidence from Three Developing Countries. World Institute for Development Economics Research, United Nations University, discussion paper no. 2002/39. [

Elbers, C., Lanjouw, J. and Lanjouw, P. (2001), Welfare in Villages and Towns: Micro-Level estimation of Poverty and Inequality, Vrije Universiteit, Amsterdam, mimeo.

Elbers, C., Lanjouw, J. and Lanjouw, P. (2002), Micro-level Estimation of Welfare, *Econometrica* **71**(1): 355–364.

Foster, J., Greer, J., and Thorbecke, E. (1984), A Class of Decomposable Poverty Measures, *Econometrica*, **52**, pp. 761-66.

Ghosh, M. and Rao, J.N.K. (1994) Small Area Estimation: An Appraisal. *Statistical Science*. **9**(1): 55-93

Henninger, N. (1998), Mapping and Geographic Analysis of Human Welfare and Poverty – Review and Assessment, World Resources Institute, Washington, D.C.

Hentschel, J., Lanjouw, J., Lanjouw, P. and Poggi, J. (2000), Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador, *The World Bank Economic Review*, **14**(1), pp. 147-165.

Klasen, S. (2004) In Search of the Holy Grail: How to Achieve Pro Poor Growth?, in Tungodden, B., N. Stern, and I. Kolstad (eds). Toward Pro Poor Policies-Aid, Institutions, and Globalization. New York: Oxford University Press. Reprinted in Krakowski, M. (ed.) Attacking Poverty: What makes growth pro-poor? Baden-Baden: Nomos (2004).

Otter, Th. (2003), Estimación de Pobreza por Ingreso y por Método Integrado de PLIPEX a nivel Distrito en Paraguay a Diciembre de 2002, Secretaria de Accion Social, Asuncion, mimeo.

Pradhan, M., Suryahadi, A., Sumarto, S. and Pritchett. L. (2001), Eating Like Which 'Joneses'? An Iterative Solution to the Choice of Poverty Line Reference Group, *The Review of Income and Wealth*, **47**(4), pp. 473-487.

Pritchett, L., Sumarto, S. and Suryahadi, A. (2000), Quantifying Vulnerability to Poverty: A Proposed Measure, Applied to Indonesia, *World Bank Policy Research Working Paper No. 2437*, The World Bank, Washington, DC.

Robles, M. (2000), Indicadores para la focalizacion, DGEEC, Fernando de la Mora.

Robles, M. and Santander, H. (2004), Mapas de Pobreza para Paraguay 2004, DGEEC, Fernando de la Mora, mimeo.

van de Walle, D. (1998), Targeting Revisited, *The World Bank Research Observer*, **13**(2), pp. 231-248.