Estimating Poverty Over Time and Space: Construction of a Time-Variant Poverty Index for Costa Rica

Romina Cavatassi, Benjamin Davis and Leslie Lipper

ESA Working Paper No. 04-21

November 2004

Agricultural and Development Economics Division
The Food and Agriculture Organization
of the United Nations

www.fao.org/es/esa

ESA Working Paper No. 04-21

www.fao.org/es/esa

Estimating Poverty Over Time and Space: Construction of a Time-Variant Poverty Index for Costa Rica

November 2004

Romina Cavatassi
Agricultural and Development
Economics Division
Economic and Social Department
Food and Agriculture Organization
Italy
e-mail: romina.cavatassi@fao.org

Benjamin Davis
Agricultural and Development
Economics Division
Economic and Social Department
Food and Agriculture Organization
Italy
e-mail: benjamin.davis@fao.org

Leslie Lipper
Agricultural and Development
Economics Division
Economic and Social Department
Food and Agriculture Organization
Italy
e-mail: leslie.lipper@fao.org

Abstract

This paper presents the construction of a spatially explicit, nationally disaggregated measure of poverty over time in Costa Rica. The paper first describes the two possible methods considered for the construction of a poverty map: principal component analysis (PCA) versus small area estimation. Next, reasons for choosing PCA and a description of its application both at one point in time (1973) and over time are presented together with the resulting poverty maps. The methodology applied represents a methodological innovation in that the resulting poverty map is time variant rather than concentrated in a single moment in time. A comparison of the results obtainable using various techniques and a discussion on the relative merits of the various options available concludes the paper.

Key Words: Poverty Mapping, Principal Component Analysis, Time-variant Poverty Index, Small-area Estimation.

JEL: C43, I32, C31.

List of tables and figures

I. Introduction[1]

Abstract: This paper is one of the results of a research project developed by the Agricultural and Development Economics Division at FAO (ESA) in collaboration with the School of International and Public Affairs at Columbia University of New York which addresses a central debate in policy, development and environmental economics: the potential for linking carbon sequestration through land use change to poverty alleviation. The purpose of the project is twofold: to develop a methodology which can be widely applied in estimating potential supply response to environmental service payments among the poor, and to provide an empirical estimate of what this response would be in Costa Rica. Research goals include assessment of the degree to which poverty influences land use change decisions (specifically deforestation) and what implications this has for establishing a carbon emissions baseline as well as the potential supply of carbon under a payment program. Furthermore, the project seeks to determine the degree to which payments for sequestration services could be a potential instrument for poverty alleviation.

Costa Rica was selected for study because it is the site of an on-going research effort to estimate potential carbon offset supply from land use by an interdisciplinary team of researchers, led by Alex Pfaff at the Columbia University, and therefore much of the necessary data were already available. However other features of the country also contribute towards making Costa Rica an ideal setting for our analysis. Land use and the population composition of the country have gone through dramatic transformations since the 1960s. Costa Rica suffered a massive loss of forest cover since the 1960s with peaks in the 70s (Bixby and Palloni, 1996) but experienced also a notable reduction of poverty in the last decades mainly due to the structural adjustment process initiated in mid-1980s and to investment in education (World Bank, 1997).

In their work on estimating the potential for carbon offset supply from land use change, the Pfaff team constructed a dynamic model of land use decision making to predict forest clearing behavior. The data set used in their study consisted of observations of forest cover for Costa Rica over five points in time[2] (Pfaff et al., 2003).

In this three chapter study, we are interested in estimating the degree to which poor people would respond to carbon payments and thus the degree to which such payments may contribute to poverty alleviation. To accomplish this using the methodology developed by Pfaff et. al. (2003), we need a measurement of poverty at a compatible level of analysis in order to distinguish the response to payments among rich and poor.

In recent years new techniques for deriving sub-national level measurements of poverty have been developed. These are referred to as poverty mapping techniques. The primary purpose of poverty mapping is the spatial identification of the poor, which also allows us to create

[1] We would like to thank Alberto Zezza for detailed comments and Irini Maltsoglou, Pierre Vauthier, Dimitra Zarra, Federico Castillo, Juan Robalino and Oswaldo Segura for their help in obtaining and setting up the census data.

[2] 1963, 1979, 1986, 1997, 2000.

variables that can be used in statistical analyses in which poverty is a dependent or explanatory variable. The latter is our primary motive in conducting the analysis presented in this chapter. To accomplish this objective we derive a spatially explicit, nationally disaggregated measure of poverty over time, which can be used as an explanatory variable in the multivariate analysis of land use changes, to assess the impact of poverty on deforestation, and ultimately, the potential supply of carbon from the poor under a carbon offset payment program.

The construction of a time variant poverty map represents a methodological innovation. While the building of poverty maps has gained increased interest among development practitioners and policy makers, most methods concentrate on maps of a single moment in time. Time series spatially explicit data are relatively difficult to come by, constraining the degree to which poverty maps over time can be developed. In addition, accounting for changes in spatial groupings over time creates complications in the analysis, as we demonstrate in this chapter.

Different methodologies are available for the construction of poverty maps (see Davis, 2003, Henninger, 2002, and Snel and Henninger, 2002 for reviews of alternative methodologies). In the present research, the choice of a poverty-mapping indicator was constrained by time, budget, access to data, as well as research objectives. The project required a technique that was inexpensive and relatively quick to construct, did not require travel to Costa Rica, was based on existing and accessible data, and could be used as an explanatory variable in a multivariate framework. The technique also needed to be comparable over time, as the analytical strategy involves time series (one observation per decade) multivariate regression.

This paper first discusses the two candidate methods we considered for the construction of the poverty map over time, and provides the reasons why we chose the principal components method. Next, we describe the estimation of the principal component method, both at one point in time (1973) and over time, as well as the resulting poverty maps. We compare the results of the poverty mapping using various techniques and base dataset, and conclude with a discussion on the relative merits of the various options available.

II. Choosing a method

Two methods were considered in order to disaggregate poverty by district for four decades: principal components analysis and community-level small area estimation.

Principal components

Principal components is a type of factor analysis, based on a statistical technique for reducing a given number of variables by extracting a linear combination which best describe these variables and transforming them into one index. This index of poverty or marginality, as it is often called, depending on the variables employed can provide a multidimensional community-level poverty indicator. The first principal component, the linear combination capturing the greatest variation among the set of variables, can be converted into factor scores, which serve as weights for the creation of the marginality index. For a national poverty map the method requires census data at any level of political or geographical aggregation (from the household to the state or provincial level)[3]. The desired level of disaggregation in our case is the third administrative level, or district (after province and canton).

The poverty index is based on the formula (from Filmer and Pritchett, 1998):

(1) $$A_j = \sum_{i=1}^{n} F_i \left[(a_{ji} - a_i) / s_i \right]$$

where F_i is the factor score for asset *i*, a_{ji} is the j^{th} district's value for asset *i* and a_i and s_i are the mean and standard deviation of asset *i* variable over all districts. By construction the mean value of the index is zero.

Small area estimation

Small area estimation is a statistical technique which combines survey and census data to estimate welfare or other indicators for disaggregated geographic units such as municipalities or rural communities. Small area estimation applies parameters from a predictive model to identical variables in a census or auxiliary database, assuming that the relationship defined by the model holds for the larger population as well as from the original sample. Small area estimation is currently the most popular methodology for the creation of national poverty maps. Two principal approaches have emerged. The first, using household unit level data from a census, has been developed principally by staff at the World Bank and is the principal methodology utilized and promoted by the Bank (World Bank, 2000; Hentschel et al. (2000) and Elbers, Lanjouw and Lanjouw (2001)). The second uses community level averages instead of household unit level data, and has been employed by researchers at both the World Bank and various international agricultural research centers (Bigman et al., 2000 and Minot, 2000).

The community-level small area estimation method requires two sets of data at a minimum: census data averaged at a given level and a representative household survey corresponding approximately to the same time period as the census. The first step is to estimate a model of consumption based household welfare using household survey data. This model should be estimated by statistically representative regions or areas (such as urban/rural), with explanatory variables limited to those found in both data sets. The second step is to apply these parameter estimates to average values taken at the chosen level of disaggregation. A predicted level of average consumption is then obtained from the consumption equation, and from this the incidence of poverty at the chosen level of disaggregation is constructed (see Bigman et al., 2000).

The method of choice

Principal components has been used in a number of countries. The Mexican government has used principal components for decades to create a marginality index for planning purposes. More recently, it has been employed as part of the targeting mechanism of the PROGRESA rural anti-poverty program, which dispenses almost $2 billion to over 4 million households annually. The Mexican application of principal components has been compared to a method similar to community level small area estimation (Skoufias, Davis and de la Vega, 2001). While both methods are highly correlated, the community-level small area estimation resulted in a stricter categorization of poverty implying that the small area estimation method would be more appropriate if avoiding leakage (including the non poor as beneficiaries) is more important than avoiding undercoverage (excluding the poor). The correlation between the two methods tends to break down in the middle of the marginality spectrum, which suggests that principal components is sharpest at high levels of marginality. This result, however, cannot be assumed to be true for all contexts.

Filmer and Pritchett (1998) used principal components in order to construct a household level asset index as a proxy for wealth. They evaluated their application to India by comparing it with other estimates of state level poverty, and they found a high level of correlation. They did find, however, a systematic bias against rural wealth as compared to conventional poverty measures. Many of their asset variables depend on infrastructure, and thus urban households are more likely to look better off then poorer households. However, standard poverty measures may be biased since real incomes/consumption are not adjusted by these implicit price differentials. Filmer and Pritchett also compared the asset index to consumption data on the same households using data from Nepal, India, and Pakistan, and found the measures produced similar rankings. Overall, they found the asset index, as a measure of long-term wealth, was more stable and had less measurement error than traditional consumption expenditures, and thus performed better as an explanatory variable (in their case in predicting school enrollment differences).

For a review of some applications of the small area estimation method see Davis (2003). Minot and Baulch (2002) look into the issue of how much precision is lost when using census data aggregated to community level or any other level. They conclude that while the best option is to use household-level data, community-level census data can be used to generate reasonably accurate poverty estimates.

We choose principal components for a number of reasons. First, it is a cheap and relatively easy method to compute, once data are obtained. Second, it has been utilized in practice in a number of countries and has provided acceptable results. Third, principal components have been shown to compare favorably with consumption based measures, particularly as an explanatory variable/proxy for long term marginality (or wealth) in multivariate analysis. Fourth, the necessary data are available over four decades. This last element provides the key advantage over community-level small area estimation, since household survey data is available only from 1987 to 2000. Linking 1990's survey data to census from 1960s and 1970s would be risky considering the important changes in the Costa Rican economy over this period (for a discussion of the potential problems of such an analysis, see Elbers et al., 2000).

III. Estimating Poverty with Principal Components

A. Data Sources and Availability

In order to create a poverty index comparable over time and space to the dataset on deforestation trends and sequestration supply constructed by Pfaff et. al. (2003), we required socio-economic data over the same time periods as their land use data set. That is as many time points possible from at least 1963 onwards. The unit of measurement for the poverty variable also needed to be appropriate for the scale of analysis utilized in the land use dataset constructed by Pfaff et. al. (2003). Their dataset is based upon district level information, as well as pixel level data which can be aggregated to the district level. We, thus, selected the district as the appropriate scale of analysis for our poverty analysis.

From 1973 to 2000 census data, variables aggregated at the district level are available electronically from the Centro Centroamericano de Población[4]. Census data from 1963 were not available electronically, and thus were collected from Dirección General de Estadística y Censos in Costa Rica in hard copy format and entered into a database. Unfortunately, the 1963 census data to which we had access did not include information on all the variables reported in the 1973 and later censuses. In the 1963 dataset, information on education, type of remuneration, dependency ratio, literacy and telephone service were not available at the district level.

We selected a group of variables from the census datasets which are typically associated with poverty. We excluded variables which in our judgment had no clear economic meaning as well as variables playing a small role in explaining the variance, such as type of job occupation or houses with heating system. Ultimately we developed a list of 17 variables from the 1973 and later censuses, and a smaller set of 12 variables from the 1963 data. The final list of variables is shown in Table 1. Most of these variables have been utilized and found to be significant in explaining poverty in Costa Rica in previous studies (World Bank 1997 and 2000b and Bixby and Palloni, 1996).

B. Estimating Poverty Indices for 1973

The difference between the data available for 1963 and 1973 onwards required the estimation of a different set of variables in each case. In this section we focus on the estimation of a poverty map using only 1973 data – that is the first year for which a full complement of explanatory variables were available. In later sections we take up the estimation of a time-variant poverty map for the 1973-based pooled dataset, and in the appendix the estimation for the 1963 dataset-based pooled dataset.

[4]http://censos.ccp.ucr.ac.cr/. The Centro is a collaborative effort between University of Costa Rica, Dirección General de Estadística y Censos of Costa Rica, Public Data Queries, Inc. of Ann Arbor, Michigan, and Population Study Center of University of Michigan.

Table 1. Variables utilized

Variable	Definition
1. male*	percentage of men total population
2. no bathroom*	percentage of dwellings without bathroom
3. no hot water*	percentage of dwellings without access to hot water
4. use coal or wood*	percentage of families who cook with coal or wood or
5. dirt floor*	percentage of dwellings with earth floor
6. dependency ratio	dependency ratio (children under 15 and people over 65/total household)
7. house in bad conditions*	percentage of dwellings in bad condition
8. no washing machine*	percentage of families without washing machine
9. no electricity*	percentage of household dwellings without electricity
10. no telephone	percentage of household dwellings without telephone
11. no refrigerator*	percentage of families without refrigerator
12. employed	percentage of people who are employed and get a salary as job remuneration
13. illiterate	percentage of illiterate population aged 12 or more
14. no water*	percentage of household dwellings without connection to private or public water system
15. no sewage*	percentage of household dwellings without sewers
16. occupants per room*	average number of occupants per bedroom
17. years of education*	average number of years of education per adult

* Available in the 1963 dataset

The results from principal components analysis applied to the 1973 census data can be found in Table 2 in which the eigenvalues of the correlation matrix are ordered from the largest to the smallest[5]. A sudden drop in the eigenvalue between the components (such as between component 1 and 2) suggests that subsequent eigenvalues are just sampling noise. As shown in the table, the first principal component explains over 63 percent of the variance in the 17 variables. This is a relatively high percentage, almost double of that found by Filmer and Pritchett in their study of India.

[5] Eigenvalues and eigenvectors are essentially a linear algebra tool to simplify complex matrices. For details see Weintraub, 1982.

Table 2. Principal components, 1973 district-level census data.

Component	Eigenvalue	Difference	Proportion	Cumulative
1	10.82863	9.11637	0.637	0.637
2	1.71226	0.82997	0.1007	0.7377
3	0.88229	0.17545	0.0519	0.7896
4	0.70684	0.02783	0.0416	0.8312
5	0.67901	0.15135	0.0399	0.8711
6	0.52766	0.20271	0.031	0.9022
7	0.32495	0.04467	0.0191	0.9213
8	0.28028	0.04602	0.0165	0.9378
9	0.23426	0.02127	0.0138	0.9515
10	0.21299	0.06902	0.0125	0.9641
11	0.14396	0.00345	0.0085	0.9725
12	0.14052	0.01829	0.0083	0.9808
13	0.12223	0.02213	0.0072	0.988
14	0.1001	0.0519	0.0059	0.9939
15	0.0482	0.01652	0.0028	0.9967
16	0.03168	0.00752	0.0019	0.9986
17	0.02416	.	0.0014	1

The eigenvector associated with the first component can be found in Table 3. In principal components, the eigenvector provides the factor score for each variable, which indicates, as understood by equation (1), the direction and weight of the impact of each variable in the poverty index. The signs on all variables are as expected. Higher values of most variables (such as share of households with a dirt floor, or share without refrigerators) are associated with higher levels of poverty. Two variables have a negative sign as expected: wage labor remuneration and average education level. Higher values of these variables are associated with lower levels of poverty.

Table 3. Eigenvectors, 1973 estimation

Variables	Eigenvector
Male	0.22398
no bathroom	0.26164
no hot water	0.2509
use coal or wood	0.26469
dirt floor	0.17918
dependency ratio	0.26293
house in bad conditions	0.16004
no washing machine	0.2698
no electricity	0.24997
no telephone	0.23071
no refrigerator	0.2799
Employed	-0.24224-
Illiterate	0.24647
no water	0.15804
no sewage	0.23694
occupants per room	0.26562
years of education	-0.28908

Thus from equation (1) we derive a district level poverty index for each of 406 districts at a specific point in time – in this case 1973. The index ranges from approximately (–13) for the wealthiest districts, to (7) for the poorest. Districts are then ranked by this index.

In **Table 4** we look at the results from the principal components analysis to see if they make sense. We rank by index value deciles the mean of each of the variables in the index. Moving from the first (best off districts) to the 10th decile (worst off) values change in a logical fashion, confirming the validity of the index.

Table 4. Mean values by poverty index deciles, 1973.

Variable	I	II	III	IV	V	VI	VII	VIII	IX	X
male	.47	.50	.50	.51	.51	.52	.52	.52	.53	.53
no bathroom	.02	.08	.13	.20	.27	.34	.43	.50	.60	.68
no hot water	.72	.90	.95	.96	.98	.99	.99	1.00	1.00	1.00
use coal or wood	.16	.34	.51	.65	.74	.79	.86	.86	.92	.96
dirt floor	.03	.11	.14	.18	.19	.17	.22	.24	.32	.44
dependency ratio	.41	.45	.47	.49	.51	.51	.52	.53	.53	.54
house in bad conditions	.08	.10	.12	.13	.13	.16	.15	.18	.20	.21
no washing machine	.62	.78	.86	.89	.94	.95	.98	.99	.99	1.00
no electricity	.02	.07	.11	.17	.22	.36	.49	.59	.73	.85
no telephone	.75	.93	.97	.99	.99	.99	1.00	1.00	1.00	1.00
no refrigerator	.50	.69	.81	.84	.91	.93	.95	.96	.96	.98
employed	.29	.26	.24	.24	.21	.19	.18	.16	.15	.11
illiterate	.04	.07	.08	.10	.12	.14	.15	.17	.21	.28
no water	.00	.01	.02	.04	.03	.10	.17	.21	.27	.31
no sewage	.01	.03	.04	.06	.08	.11	.17	.25	.33	.50
occupants per room	1.47	1.97	2.18	2.43	2.60	2.62	2.73	2.94	3.09	4.05
years of education	6.12	4.70	4.25	3.88	3.56	3.37	3.28	3.07	2.87	2.44

In 1973, in the richest districts only 2 percent of the households did not have at least a latrine, while in the poorest districts the percentage goes up to 68 percent. Similarly in wealthiest districts only 16 percent of the households used coal or wood to cook, as opposed to 96 percent in poorest districts. Access to electricity and to a sewage system are near universal in the richest districts, while in the poorest decile 85 percent of the population lacked access to electricity and 50 percent lacked access to sewage facilities. The average number of occupants per room for the wealthiest districts was 1.47 as compared to 4.05 for the poorest. Following the same trend, the average number of years of education per person is 6.12 in richest districts and 2.44 in the poorest.

Overall, the poorest live in low quality dwellings, lack access to water and electricity, do not have a bathroom, use coal or wood to cook, have a lower level of education, have lower levels of employment in wage labour and have a higher number of occupants per room. The relationship of these variables with poverty is similar to that found in other studies, with the exception of the role of gender, where a larger share of women within the household has generally been associated with a greater level of poverty (World Bank 1997 and 2000). The gender variable used in this study is somewhat different - the share of males at the district level - which may explain the ambiguous result.

In Table 5, the ten wealthiest districts in 1973 are ranked according to their index scores. Not surprisingly, nine out of ten of the districts are located in the province of San Josè, and six of them are located in the canton of San Josè, the capital.

Table 5. Wealthiest 10 districts in 1973

PROVINCE	CANTON	DISTRICT	Poverty index
SAN JOSÈ	San Josè	Carmen	-13.4494
SAN JOSÈ	San Josè	MataRedo	-11.4891
SAN JOSÈ	San Josè	Catedral	-9.8910
SAN JOSÈ	Montes de Oca	SanPedro	-9.4676
SAN JOSÈ	San Josè	Mercedes	-8.7696
SAN JOSÈ	San Josè	San Francisco Dos Rio	-8.3663
SAN JOSÈ	Goicoechea	Guadalupe	-7.9786
SAN JOSÈ	Tibas	San Juan	-7.7678
SAN JOSÈ	San Josè	Zapote	-7.7660
HEREDIA	Heredia	Heredia	-7.7605

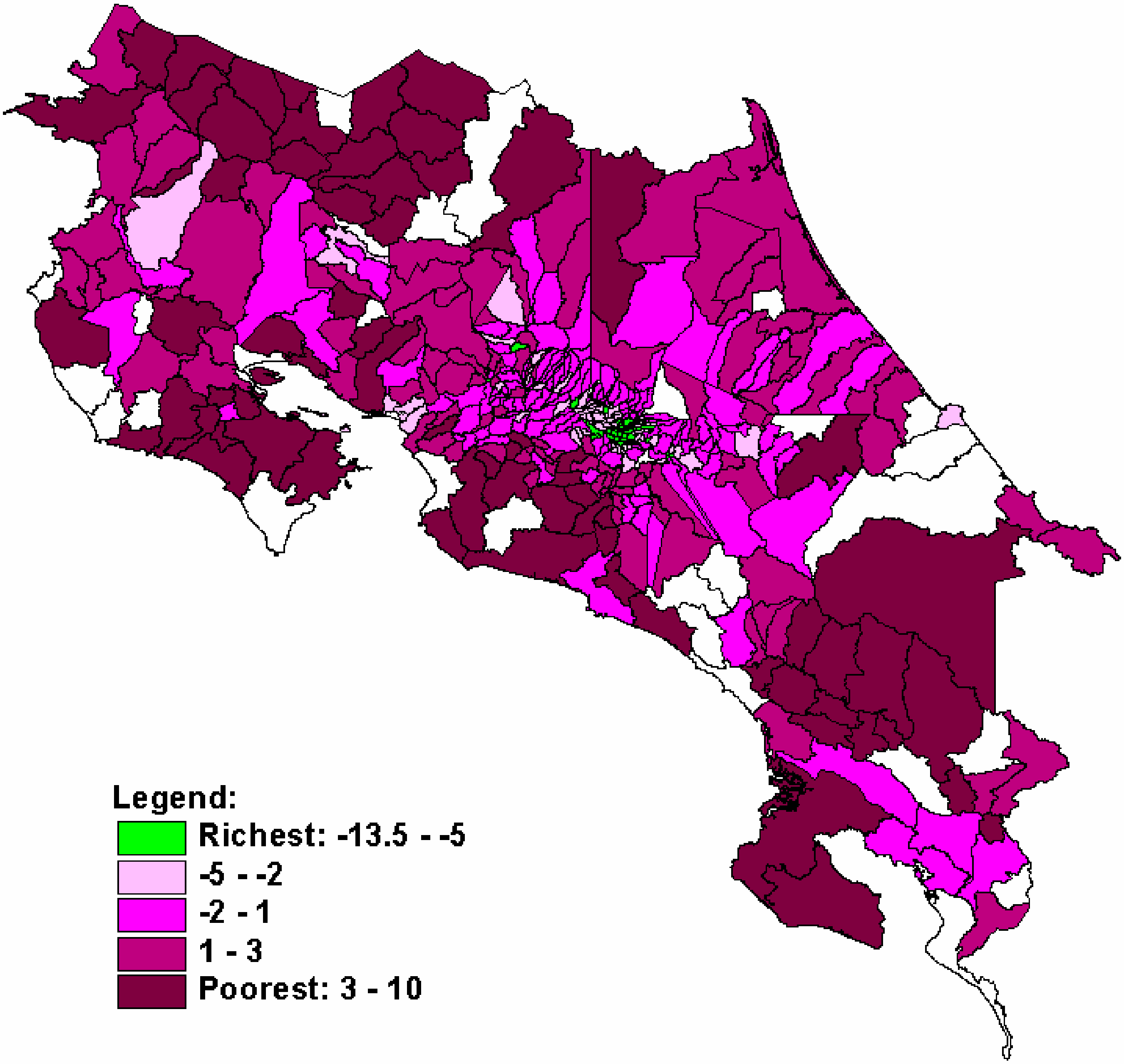
In Table 6 we list the 10 poorest districts in 1973 ranked according to the index. Four of the ten are located in the province of Puntarenas, and all of these except for one are located in the Canton of Buenos Aires

Table 6. Poorest 10 districts in 1973

PROVINCE	CANTON	DISTRICT	Poverty index
SAN JOSÈ	Acosta	Sabanilla	6.7002
PUNTARENAS	Buenos Aires	Colinas	6.6145
ALAJUELA	Los Chiles	El Amparo	6.1917
ALAJUELA	Upala	San Josè	6.1082
PUNTARENAS	Buenos Aires	Potrero Grande	5.8940
PUNTARENAS	Buenos Aires	Boruca	5.6805
GUANACASTE	La Cruz	Santa Cecilia	5.6293
ALAJUELA	Los Chiles	Canyo Negro	5.5568
PUNTARENAS	Osa	Sierpe	5.5302
SAN JOSÈ	Tarrazú	San Carlos	5.4328

The results of the 1973 static poverty map are shown in Figure 1. The richest districts are those with the smallest value of the poverty index (negative value). The wealthiest are given a different color, as they are smaller and located near the center of the country and thus harder to distinguish. Thus the wealthier districts are represented first by green, and then by the lighter shades of purple. Poorer districts have larger values of the poverty index, and are represented by darker shades of purple. White areas represent missing data. The poorest districts are those located far from the Central Valley. This distribution of poverty confirms results of other studies (World Bank 1997 and 2000, Bixby 1996).

Figure 1. District-level poverty map, 1973.



C. A Time-Variant Measure of Poverty

In the previous discussion we focused on the construction of a static poverty index for one point in time, using the example of 1973 to illustrate the technique. Static indices, however, are not comparable over time. Each index is based on an eigenvector, or scale, which is relevant only to that particular estimation. In other words, the units in which the indexes are constructed vary in each estimation, precluding comparison of the index among different census years. This presents a problem for the regression analysis described in the introduction, which requires a poverty index comparable over time.

To overcome this limitation, we pool the 1963, 1973, 1984 and 2000 census data and estimate principal components over the combined data. The resulting eigenvector is than applied to the variable values from each census year using equation (1). The principal limiting assumption is that we have averaged the impact of the included variables over the four decades of data we have available. Change in the marginality index is thus limited to changes in the levels of variables, and not changes in the relative importance (or impact) of each variable in determining the index. For instance changes in social or economic structure may alter the importance of education over the period 1963 to 2000, but we have essentially averaged these changes over all years.

We face two data limitations in making this estimation. First, the number of districts changes each census year. Over time, as the population grows, districts split and the overall number increases. The numbers change from 333 in 1963, to 406 in 1973, 420 in 1984 and 459 in 2000. The second limitation is the lack of availability of a complete set of 17 variables for estimating poverty in the 1963 census dataset as was discussed in section A above. Below we describe how we overcame these limitations.

Overcoming the first limitation: Changes in district areas over time

Typically new districts are created by splitting larger districts into parts, in which case the original name of the district remains with one section, while the other(s) receive a new name. We managed to obtain information about the evolution of districts from 1980 onwards but not for changes that occurred between 1963 to 1980. To re-run the poverty index consistently with new districts or district re-coding we essentially faced two options – either reaggregate data from later years back into the 1963 configuration of districts, or disaggregate early year data based on the 2000 district configuations. Specifically:

1. Re-aggregate: If in year (t+1) we have districts A and B which were derived from district A in year t, we re-aggregate them back into district A. Re-aggregating means that the value of each variable for district B is aggregated back with the value of the same variable of district A (i.e. if we have the number of men and women of district A and B (in t+1) we add them together and have the total number of men and women in the district that now we call A as it was in year t). In this way we can build our analysis on the "base year" which is 1973 and remain consistent with the oldest districts having all details we need on variables to use although at a more aggregate level of analysis (fewer districts).

2 Disaggregate: District A in earlier years is divided into parts consistent with the district configuration of later years. Here, A in t became A and B in t+1, therefore we disaggregate A into A and B also at time t.

The second method, disaggregation, would create changes in poverty resulting from changes in area. There would be more districts and thus more observations as a result of having this information, since in the earlier years aggregated values for old districts would be assigned to both sections of the disaggregated estimate. However, this could create a serious bias in the poverty estimates if in fact the disaggregated sections differ significantly in their incidence of poverty. The rich ones will show declines in poverty over time and the poor ones increases, due to the changes in the areas over which averages were calculated.

The first method, re-aggregating, uses the information we have on district changes over time not to create more districts, but instead to eliminate the changes in poverty level that arise solely from the changes in area. With this method we end up with a data set of only those districts that persist through time. However, the spatial averaging done in the early years, within the official districts, would be the same as the spatial averaging done by re-aggregating the districts in the later years.

In our estimation we applied the first method, reasoning that the gain in error reduction was more important than the loss of information it entails. We use the number of districts at the earliest point in time for the analysis as the base and re-aggregate all data from later years to these same district boundaries. Thus for the estimates of poverty indices by district for 1973-2000, we use 406 districts, which was the number of districts in 1973. For the estimates done for 1963-2000, we use 333 districts, which was the number of districts in 1963.

Overcoming the second limitation: differences in data availability between census years

To overcome the second problem we decided to create two different indexes: one which would use the full set of 17 variables for the years it was available 1973-2000, and a second index with the smaller set of 12 variables, which are available for all time periods beginning in 1963.

3. Time-Variant Poverty Map Results

District level, 17 variables, for 1973, 1984 and 2000 data (406 districts)

In this section we estimate the poverty index with the 1973 base year which includes the full set of 17 variables. The results for the poverty index with 1963 as the base year can be found in the Appendix. We begin by pooling the 1973, 1984 and 2000 data, and estimating the principal components over the pooled data. **Table 7** contains the eigenvalues of the correlation matrix ordered from largest to smallest. As in our estimation of the static poverty index, the first factor explains over 64 percent of the variance. Table 8 contains the factor scores obtained the first principal component. The signs on these factor scores are as expected. Higher values of most variables (dependency ratio, share of households without access to electricity or without washing machine etc) lead to higher level of poverty. Similarly higher number of people without wage remuneration or with lower level of education lead to higher level of poverty.

Table 7. Principal components, pooled 1973, 1984, and 2000 district-level census data.

Component	Eigenvalue	Difference	Proportion	Cumulative
1	11.00786	9.55335	0.6475	0.6475
2	1.45451	0.47916	0.0856	0.7331
3	0.97535	0.15211	0.0574	0.7905
4	0.82324	0.24999	0.0484	0.8389
5	0.57325	0.0602	0.0337	0.8726
6	0.51304	0.12657	0.0302	0.9028
7	0.38647	0.14879	0.0227	0.9255
8	0.23768	0.03105	0.014	0.9395
9	0.20663	0.01516	0.0122	0.9516
10	0.19147	0.02166	0.0113	0.9629
11	0.16982	0.02579	0.01	0.9729
12	0.14403	0.04016	0.0085	0.9814
13	0.10386	0.01415	0.0061	0.9875
14	0.08971	0.02056	0.0053	0.9928
15	0.06915	0.03028	0.0041	0.9968
16	0.03887	0.02383	0.0023	0.9991
17	0.01505	.	0.0009	1

Table 8. Eigenvectors, principal components estimated over 1973-2000

Variables	Eigenvector
male	0.1912
no bathroom	0.26221
no hot water	0.23956
use coal or wood	0.26299
dirt floor	0.22362
dependency ratio	0.27125
house in bad conditions	0.16554
no washing machine	0.26132
no electricity	0.25822
no telephone	0.24961
no refrigerator	0.2762
employed	-0.21707
illiterate	0.26199
no water	0.17728
no sewage	0.23722
occupants per room	0.25321
years of education	-0.27569

Applying factor scores from Table 8 to the 1973, 1984 and 2000 data, and applying equation (1), we get poverty indexes for 1973, 1984 and 2000 at the district level.

To examine the robustness of the method applied we calculated eigenvectors for each point in time and for sub periods (i.e 1973-1984 and 1984-2000). In all calculations the first factor explains over 64 percent of the variance with a range that goes from 63.7% to 68%. Similarly, the correlation indices between eigenvectors are all above 0.88 which is considered a very strong correlation and in our case validate the robustness of the methodology used.

The mapped results of the estimations are shown in Figures 4-6. As before, the darker shades refer to the poorer districts, whilst the wealthiest districts are given a lighter blue color.

Figure 2. Time-variant poverty map, 1973 (pooled index, 1973 base year).

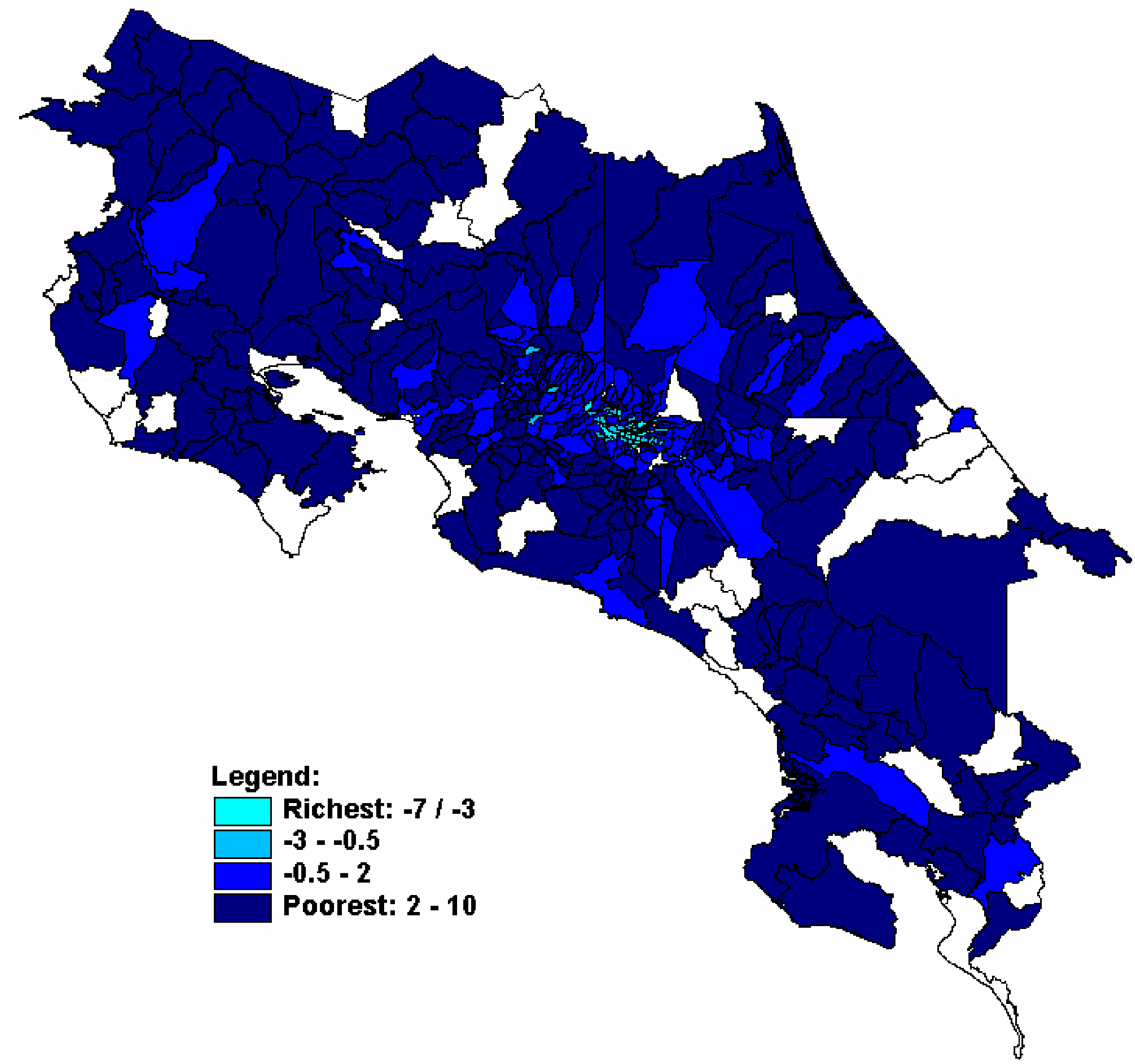


Figure 3. Time-variant poverty map, 1984 (pooled index, 1973 base year).

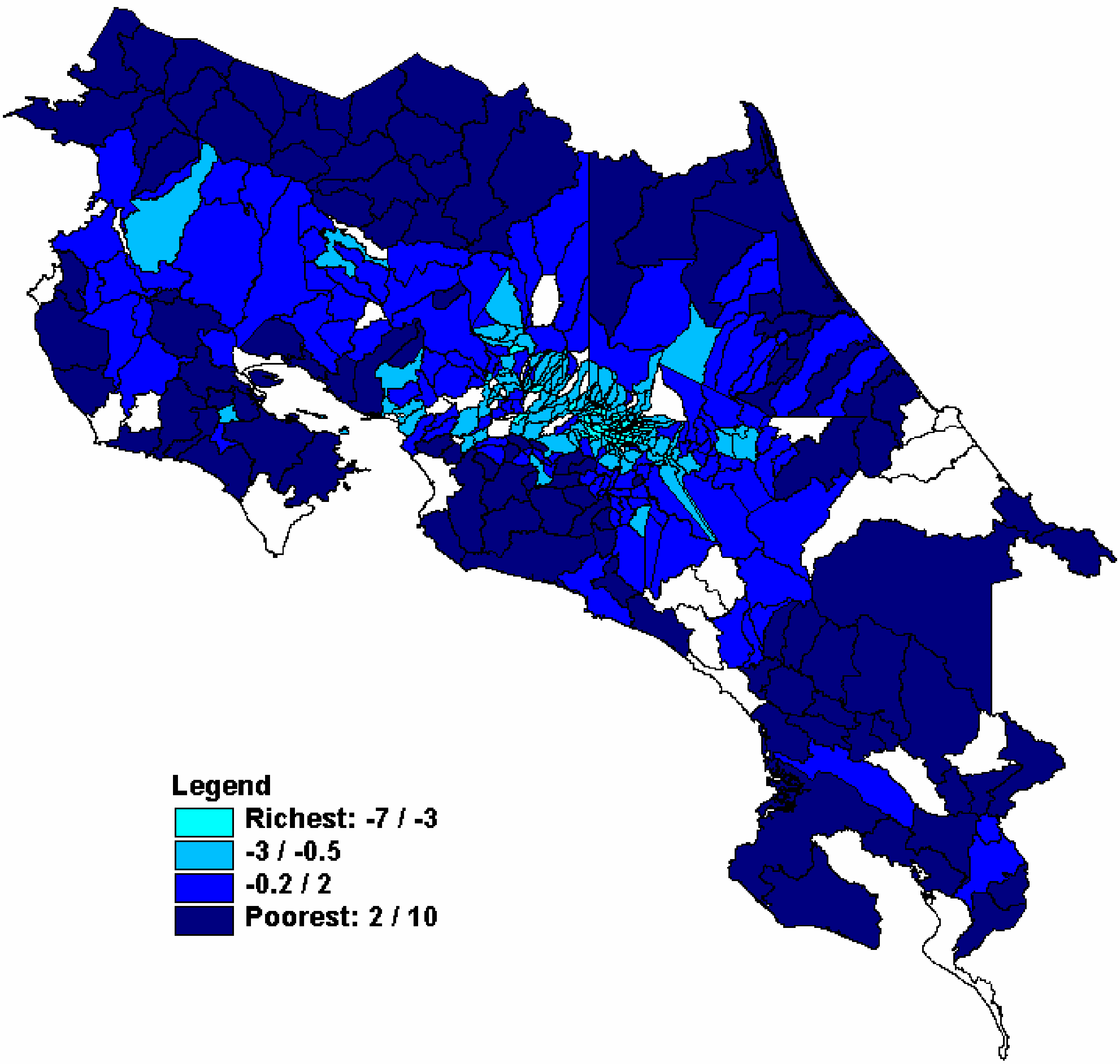
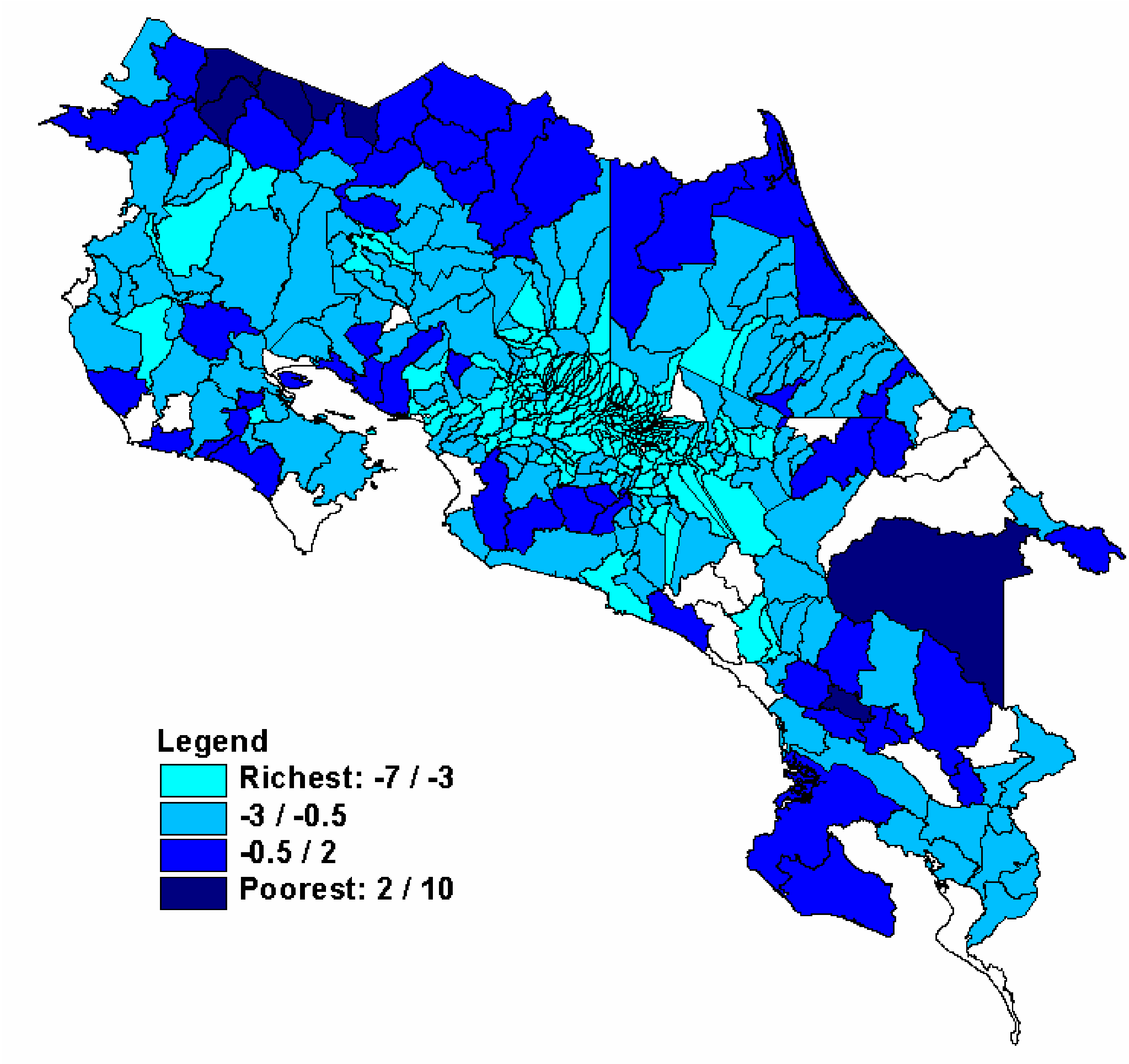


Figure 4. Time-variant poverty map, 2000 (pooled index, 1973 base year).



Visually the maps indicate spatial clustering by wealth which persists over time In Table 9 the evolution of the ranking of the poorest districts over the three points in time used in the time-variant analysis is shown, along with results from the 1973 static estimation. Table 10 makes the same comparison, but focusing on the 10 wealthiest districts. In both tables, districts are ordered by the rankings from the 1973 pooled estimation. In both cases, very few differences are seen between the 1973 static and time variant results.

The relative ranking of the poorer districts shows remarkable stability over time. Five of the 10 poorest districts in 1973 are still there in 2000. All of the original 10 poorest districts in 1973 remained in the 20 poorest districts in 2000. Looking into a more detail at the district level results in order to understand what drives changes in the relative ranking, we take as an example the district of San Carlos, which managed to improve by 46 spots, from 396th to 350th, between 1973 and 2000. This district is characterized by rapid reductions in illiteracy and growth in educational levels, widespread improvement in household living conditions, as well as dramatic increase in the public provision of sewage, water and telephone services.

In the opposite direction, the district of Tayutic moved from position 343 to 397, thus dropping by 54 spots. Looking into the details behind this movement, we find that a few characteristics worsened, such as percentage of dwellings with dirt floor or without water. Further while nationally there has been considerable progress in the share of dwellings with electricity (from 35% HH without electricity in 1973 to 0.06% in 2000), owing telephone (96% had no telephone in 1973 vs 57% in 2000), and hot water (95% did not have hot water in 1973 as compared to 64% in 2000), in Tayutic the situation with regard to these variables has remained substantially the same or has not changed as much as it did at national level. Similar statement holds true for illiteracy (0.06% at national level versus 22% in Tayutic),average number of occupants per room (1.65 nationally vs 2.12 in Tayutic) and average number of years of education (6.03 vs 3.47).

Table 9. Poorest Districts (evolution from 1973 to 2000)

Province	Canton	District	1973 Static	1973 Pooled	1984 Pooled	2000 pooled
SAN JOSÈ	Acosta	Sabanill	**406**	**406**	**403**	396
PUNTARENAS	Buenos Aires	Colinas	**405**	**405**	387	395
ALAJUELA	Los Chiles	ElAmparo	**404**	**404**	**397**	388
PUNTARENAS	Buenos Aires	PotreroG	**402**	**403**	386	**399**
ALAJUELA	Upala	SnJose	**403**	**402**	**402**	**401**
PUNTARENAS	Buenos Aires	Boruca	**401**	**401**	390	393
PUNTARENAS	Osa	Sierpe	**398**	**400**	396	392
GUANACASTE	La Cruz	StaCecil	**400**	**399**	**406**	**403**
ALAJUELA	Los Chiles	CanyoNeg	**399**	**398**	**405**	**400**
PUNTARENAS	Buenos Aires	Pilas	396	**397**	394	**404**
SAN JOSÈ	Tarrazú	SnCarlos	**397**	396	363	350
ALAJUELA	Guatuso	Buenavis	394	395	**401**	385
ALAJUELA	Upala	Delicias	390	387	**399**	**402**
LIMÓN	Talamanca	Bratsi	389	384	**404**	**406**
ALAJUELA	Upala	DosRios	384	382	**400**	**405**
GUANACASTE	La Cruz	Garita	355	355	**398**	**398**
CARTAGO	Turrialba	Tayutic	347	343	356	**397**

In Table 10, which explores the evolution of the 10 wealthiest districts over time, there is also considerable stability in the provincial rankings, with somewhat more variation at the canton and district level. The ten wealthiest districts are consistently located in the provinces of San Jose and Heredia over all time periods. However, only 4 of the 10 wealthiest districts in 1973 remain so in 2000, and three others are ranked 11th, 12th and 13th. One district, however, Merced in the Province of San Jose, moved from 5th in 1973 to 17th in 1984, to 55th in 2000. The rate of living condition improvement in this district is lower than in other areas, although generally in the same direction. Increases in the number of occupants per room (from 0.86 to 1.48 in Meced and from 2.60 to 1.65 at national level) and no major changes in percentage of households without electricity or connections to sewage lines (as compared to a dramatic change at national level as described above), indicates the possible growth of urban and peri-urban slums. On the other hand, the district of Sánchez, also in the province of San Jose, moved from 44th in 1973 to 1st place in 2000. This district is characterized by the highest educational levels (11.43 years in 2000 as compared to a national average of 6.03), high rates of wage labor remuneration and very good performance in the provision of public services (telephone, sewage, water) during the time period considered[6]. and.

[6] Percentage of households without refrigerator dropped to 0.02% in 2000 from 70% in 1973 as compared to a national average of 20% in 2000. Similarly households without washing machine were 79% in 1973 and 2% in 2000 versus a national average of 23%. Same trend holds true for other assets such as telephone, hot water etc.

Table 10. Wealthiest districts (evolution from 1973 to 2000)

Province	Canton	District	1973 Static	1973 Pooled	1984 Pooled	2000 Pooled
SAN JOSE	San Jose	Carmen	**1**	**1**	**1**	**2**
SAN JOSE	San Jose	MataRedo	**2**	**2**	**2**	**6**
SAN JOSE	San Jose	Catedral	**3**	**3**	13	27
SAN JOSE	Montes de Oca	SnPedro	**4**	**4**	**3**	**5**
SAN JOSE	San Jose	Merced	**5**	**5**	17	55
SAN JOSE	San Jose	SnFranDR	**6**	**6**	11	**8**
SAN JOSE	Goicoechea	Guadalup	**7**	**7**	15	22
SAN JOSE	San Jose	Zapote	**9**	**8**	**10**	12
HEREDIA	Heredia	Heredia	**10**	**9**	14	13
SAN JOSE	Tibas	SnJuan	**8**	**10**	**7**	11
SAN JOSE	Curridabat	Sánchez	40	44	**4**	**1**
SAN JOSE	Montes de Oca	Mercedes (Betania)	13	13	**5**	**3**
SAN JOSE	Moravia	San Vicente	11	11	**6**	**9**
HEREDIA	Santo Domingo	Santo Domingo	14	12	**8**	**10**
SAN JOSE	Tibas	Anselmo Llorente	17	14	**9**	**4**
SAN JOSE	Montes de Oca	Sabanilla	34	32	12	**7**

To examine the evolution over time of all districts in our study and to validate static versus dynamic method of analysis, we also calculated the Spearman ranking correlation coefficient. This confirmed that the poverty index at one point in time for 1973 provides a very similar ranking of districts as compared to the pooled poverty index for 1973. This holds true for the entire sample as well as for the 30 best off and worst off districts (Spearman rho is always > 0.9). Similarly in comparing the ranking provided by the three pooled indexes over time we can confirm the validity of the methodology at least for the better off districts (1973 vs 1984 and 1984 vs 2000) whilst for the worst off districts the index does show a weak correlation and this might be due to missing data for districts that changed over time as explained before.

IV. Conclusions

In this paper we have constructed a variable measuring poverty for use as an explanatory variable in a district-level multivariate regression analysis. The method we chose, principal components, requires fairly limited amounts of data as compared with other methods, which was a major consideration in our choice of principal components. The method also allowed estimating a poverty index comparable over different points in time, although based upon the somewhat stringent assumption of no changes in the relative importance of each component of the index over the time period considered. A comparison of the results from estimations based on pooled versus static data indicate that for this dataset, the assumption of stability in poverty components over time is valid. This is one of the first examples in the literature of a time-variant poverty index.

We also addressed the issue of changes in administrative units over time, which is likely to be a common problem in any time variant spatial analysis. We argue for the use of a re-aggregation procedure, where the number and boundaries of districts at the earliest point in time for the analysis is used as the base, to which all subsequent units are re-aggregated. In this analysis, the

the gain in error reduction was more important than the loss of information it entailed, although in other situations the opposite may hold true.

We estimated two sets of poverty indices over time at the district level for Costa Rica; one based upon a set of variables and district configurations from 1963, and the second based upon variables and district configuration from 1973. We found that there is considerable similarity in the results between the two base year estimations in terms of poverty rankings. In both cases the principal components analysis yielded a first principal component which explained approximately 64 percent of the variation in the data. We also found the direction of the signs of the factors derived from the eigenvectors the same for the set of variables included in each of the base year estimations and that the relative impact of the included variables remained fairly consistent between the two estimations. The results of the poverty map are broadly consistent with other studies of poverty in Costa Rica.

Finally, our results indicate the importance of scale and location in the analysis of poverty. We found that over the time period from 1973 to 2000, the wealthiest and poorest districts were consistently located in the same provinces, although there was variation in which districts these were. Thus, at a provincial level, poverty and wealth are spatially clustered in Costa Rica over time, and the effect is quite strong. At a district level of analysis however, although we still do find patterns of spatial clustering, particularly at the relative extremes of wealth and poverty, we are also able to see much more heterogeneity in the distribution of wealth over time. These results do indicate that location is an important factor, which needs to be carefully considered in statistical analyses of the determinants of wealth and poverty.

Bibliography

Bigman, D., Dercon, S., Guillaume, D. & Lambotte, M. 2000. Community targeting for poverty reduction in Burkina Faso. *The World Bank Economic Review*, 14(1): 167–194.

Bixy, L.R. and Palloni, A. 1996. Population and Deforestation in Costa Rica, Paper presented at the Annual Meeting of the Population Association of America. New Orleans, 199: http://ccp.ucr.ac.cr/proyecto/poyam/pdf/pyd7383i.pdf...

Davis, B. 2003. *Choosing a Poverty Mapping Method*, Rome, Food and Agricultural Organization.

Elbers, C., J. Lanjouw, and P. Lanjouw. 2001. "Welfare in villages and towns: micro-level estimation of poverty and inequality," mimeo, Vrije Universiteit, Yale University and The World Bank.

Filmer D. and Pritchett L., 1998. *Estimating wealth effects without expenditure data -- or tears: An application to educational enrolments in States of India,* World Bank Policy Research Working Paper No. 1994

Henninger, N. 1998. *Mapping and geographic analysis of poverty and human welfare – Review and assessment*, Report prepared for the UNEP/CGIAR Initiative on GIS, World Resources Institute, Washington, D.C.

Hentschel, J., Lanjouw, J., Lanjouw, P. & Poggi, J. 2000. Combining census and survey data to trace spatial dimensions of poverty: a case study of Ecuador. *The World Bank Economic Review*, 14(1): 147–165.

Minot, N. 2000. Generating disaggregated poverty maps: an application to Viet Nam. *World Development*, 28(2): 319–331.

Minot, N. & Baulch, B. 2002. *Poverty mapping with aggregate census data: what is the loss in precision?* (Paper for Understanding Poverty and Growth in Sub-Saharan Africa conference, St Catherine's College, Oxford, UK, 18–19 March.)

Skoufias, E., Davis, B. & de la Vega, S. 2001. Targeting the poor in Mexico: an evaluation of the selection of households into PROGRESA. *World Development*, 29(10): 1 769–1 784.

Snel, M. & Henninger, N. 2002. Where are the poor? Experiences with the development and use of poverty maps. Washington DC, World Resources Institute.

World Bank. 1997. "Costa Rica: Identifying the Social Needs of the Poor, An Update" Report, No: 15449-CR

World Bank. 2000. "The geography of poverty: estimation and analysis of small area welfare indicators," proposal for funding.

World Bank. 2000. "Costa Rica Social Spending and the Poor"Report No. 24300.

Pfaff et at. 2003

http://censos.ccp.ucr.ac.cr/

Appendix

In this appendix we present the estimation of an index based on a reduced set of 12 variables which were available for the 1963 census year as well as the other 3 years. We begin by pooling the 1963, 1973, 1984 and 2000 data, and estimating the principal components over the pooled data. Table A1 contains the eigenvalues of the correlation matrix ordered from largest to smallest. As in previous estimations, the first factor explains approximately 63 percent of the variance. Table A2 contains the factor scores obtained the first principal component. Again, the signs on these factor scores are as expected.

Table A1. Principal components, pooled 1963, 1973, 1984, and 2000 district-level census data.

Component	Eigenvalue	Difference	Proportion	Cumulative
1	7.6405	6.36979	0.6367	0.6367
2	1.27072	0.3429	0.1059	0.7426
3	0.92782	0.29736	0.0773	0.8199
4	0.63046	0.17711	0.0525	0.8725
5	0.45334	0.16111	0.0378	0.9102
6	0.29224	0.06394	0.0244	0.9346
7	0.2283	0.03806	0.019	0.9536
8	0.19024	0.04378	0.0159	0.9695
9	0.14646	0.03163	0.0122	0.9817
10	0.11483	0.03045	0.0096	0.9912
11	0.08438	0.06367	0.007	0.9983
12	0.02071	.	0.0017	1

Table A2. Eigenvectors, principal components estimated over time, including 1963

Variables	**Eigenvector**
male	0.17041
no bathroom	0.33097
no hot water	0.27368
use coal or wood	0.32105
dirt floor	0.28975
house in bad conditions	0.2539
no washing machine	0.31518
no electricity	0.32382
no refrigerator	0.32826
no water	0.20363
no sewage	0.29855
occupants per room	0.30412

Applying factor scores from Table A2 to the 1963, 1973, 1984 and 2000 data, following equation (1), we get poverty indexes for these years at the district level. In Table A3 the evolution of the ranking of the poorest districts over the four points in time used in the time-variant analysis are shown, and in Table A4 evolution of the ranking of the wealthiest districts over the four points in time. The districts are ordered by the 1963 ranking.

Table A3. Poorest Districts (evolution from 1963 to 2000)

Province	Canton	District	1963 ranking	1973 ranking	1984 ranking	2000 ranking
PUNTARENAS	Buenos Aires	Boruca	**333**	**333**	**327**	**329**
SAN JOSE	Acosta	Sabanillas	**332**	**331**	**333**	**331**
GUANACASTE	Nandayure	Bejuco	**331**	**328**	**330**	323
SAN JOSE	Tarrazu	San Carlos	**330**	319	290	270
PUNTARENAS	Puntarenas	Lepanto	**329**	**325**	314	308
SAN JOSÉ	Leon Cortes	Llano Bonito	**328**	310	**279**	261
SAN JOSÉ	Mora	Guayabo	**327**	300	278	216
GUANACASTE	santa Cruz	Veintisiete de Abril	**326**	**329**	317	313
SAN JOSE	Puriscal	Mercedes Sur	**325**	312	281	259
SAN JOSE	Turrubara	San Juan de Mata	**324**	318	**329**	317
SAN JOSE	Acosta	Cangrejal	323	**326**	315	321
PUNTARENAS	Buenos Aires	Potrero Grande	321	**332**	323	**333**
PUNTARENAS	Puntarenas	Manzanillo	319	322	312	**328**
PUNTARENAS	Aguirre	Savegre	317	317	**326**	310
GUANACASTE	Abangares	Colorado	315	**324**	288	314
GUANACASTE	Nandayure	Santa Rita	311	306	**324**	311
GUANACASTE	Liberia	la cruz	308	267	301	**326**
SAN JOSE	Mora	Piedras Negras	303	293	**325**	281
ALAJUELA	Upala	Upala	298	**327**	322	319
PUNTARENAS	Golfito	Jiménez	279	313	**331**	**324**
PUNTARENAS	Puntarenas	Pitahaya	275	301	316	**327**
Alajuela	Los Chiles	Los Chiles	256	316	**328**	**332**
PUNTARENAS	Puntarenas	Chomes	244	288	291	**325**
PUNTARENAS	Osa	Sierpe	153	**330**	**332**	**330**

The relative ranking of the poor districts is less stable, with only 2 of the top 10 poorest districts in 1963 still there in 2000. Still, many of the same trends in ranking changes are found in both the 1963 and 1973 pooled estimations. For example, the district of San Carlos, which with the 1973 pooled estimate increased 46 spots from 1973 to 2000, with the 1963 pooled results increased 49 spots over the same period. The district which improves most spectacularly over time, Guayabo, moving from 327th to 216th position, also shows dramatic improvement using the 1973 pooled estimates.

Similar trends for the relative ranking of wealthy districts are evident. The two districts which showed greatest improvement using the 1973 pooled estimates, Sánchez and Sabanilla, have similar results using the 1963 pooled estimates. Further, the district which fares worst over time, Hospital, dropping from 8th wealthiest to 157th in 2000, suffers a similar trend using the 1973 pooled data.

Table A4. Wealthiest Districts (evolution from 1963 to 2000)

Province	Canton	District	1963 ranking	1973 ranking	1984 ranking	2000 ranking
SAN JOSE	San Jose	Carmen	**1**	**1**	**1**	**1**
SAN JOSE	San Jose	Catedral	**2**	**3**	13	30
SAN JOSE	San Jose	Merced	**3**	**5**	15	13
SAN JOSE	Montes de Oca	San Pedro	**4**	**4**	**4**	**9**
SAN JOSE	Desampar	Desamparados	**5**	22	27	28
SAN JOSE	Goicoechea	Guadalupe	**6**	**6**	11	15
SAN JOSE	Tibas	San Juan	**7**	**8**	**6**	12
SAN JOSE	San Jose	Hospital	**8**	18	55	157
SAN JOSE	San Jose	San Francisco de Dos Ríos	**9**	**7**	**7**	**5**
HEREDIA	Heredia	Heredia	**10**	**9**	14	18
SAN JOSE	San Jose	Mata Redonda	11	**2**	**2**	**7**
SAN JOSE	Moravia	San Vicente	14	**10**	**5**	**8**
SAN JOSE	San Jose	Zapote	15	11	**10**	11
CARTAGO	Cartago	Occidental	17	16	29	**10**
HEREDIA	Santo Domingo	Santo Domingo	19	12	**9**	**4**
SAN JOSE	Tibas	Anselmo Llorente	26	15	**3**	**3**
SAN JOSE	Montes de Oca	Sabanilla	44	34	12	**6**
SAN JOSE	Curridabat	Sánchez	80	53	**8**	**2**

ESA Working Papers

WORKING PAPERS

The ESA Working Papers are produced by the Agricultural and Development Economics Division (ESA) of the Economic and Social Department of the United Nations Food and Agriculture Organization (FAO). The series presents ESA's ongoing research. Working papers are circulated to stimulate discussion and comments. They are made available to the public through the Division's website. The analysis and conclusions are those of the authors and do not indicate concurrence by FAO.

ESA

The Agricultural and Development Economics Division (ESA) is FAO's focal point for economic research and policy analysis on issues relating to world food security and sustainable development. ESA contributes to the generation of knowledge and evolution of scientific thought on hunger and poverty alleviation through its economic studies publications which include this working paper series as well as periodic and occasional publications.

Agricultural and Development Economics Division (ESA)
The Food and Agriculture Organization
Viale delle Terme di Caracalla
00100 Rome
Italy

Contact:
Office of the Director
Telephone: +39 06 57054358
Facsimile: + 39 06 57055522
Website: www.fao.org/es/esa
e-mail: ESA@fao.org