Construction of poverty map for the HCM city in Vietnam using the 2004 VHLSS and the 2004 HCM Mid-Census

Nguyen Viet, Cuong and Van der Weide, Roy and Le, Hai and Tran, Ngoc Truong

28 September 2007

Construction of poverty map for the HCM city in Vietnam using the 2004 VHLSS and the 2004 HCM Mid-Census

Nguyen Viet Cuong
Roy van der Weide
Le Hai
Tran Ngoc Truong[1]

Abstract

The research estimates the poverty rate for the districts in the Ho Chi Minh city in Vietnam using a method of small area estimation and data from Vietnam Household Living Standard Survey 2004, and the 10 percent mid-census sample of HCM city. It is found that poverty estimates are much higher in suburb districts which have a large proportion of rural area. However, the poverty density is smaller in the poorest districts and higher in the richest districts, since the population density is much lower in the poorest districts than in the richest districts. The standard errors of the poverty estimates are relatively high, which makes the comparison of poverty between districts difficult, especially for districts with poverty rates less than 10%. The Gini estimates at the district level are rather small, around 0.3.

JEL classification: I31, I32, O15

Keywords: Poverty measurement, poverty mapping, agricultural census, household survey, Vietnam.

[1] Nguyen Viet Cuong is a lecturer of National Economic University (contact athour: c_nguyenviet@yahoo.com) ; Roy van der Weide is a consultant at the World Bank; Tran Ngoc Truong is a researcher in Institute of Labor Science and Social Affairs (ILSSA); and Le Hai is researcher from Institute of Economic Research in HCM city.

1. INTRODUCTION

Ho Chi Minh city is the largest and richest city in Vietnam. This report presents a revision of the poverty map for Ho Chi Minh (HCM) city at the district level using the small area estimation method developed by Elbers *et al.* (2003).[2] The old poverty map of HCM city was constructed using Vietnam Household Living Standard Survey (VHLSS) in 2002 and a 10% sample of the HCM city Mid-Census for 2004. One important assumption in the old map construction is that there is no spatial correlation between households within a cluster. This assumption can be very strong. The main objective of the present study is to examine whether there is spatial correlation and how the welfare estimates and standard errors are sensitive to this assumption. In addition, we use VHLSS for 2004 instead of VHLSS for 2002.

The report consists of 7 sections. Section 2 describes the method of small area estimation. Section 3 introduces the data used for the analysis. Section 4 presents the common variables, that are available in both the survey and the census, and verifies their comparability. The income model regressions and the poverty estimates are presented in sections 5 and 6. Finally, section 7 concludes.

2. METHODOLOGY

We will adopt the small-area estimation method developed by Elbers, Lanjouw and Lanjouw (2002, 2003; hereafter referred to as ELL), which is arguably most popular in the context of poverty analysis. In ELL two samples (typically the socio-economic survey with a detailed expenditure module and a population census) are combined through an expenditure model. This combination allows us to obtain small area estimates (SAE) of welfare, and/or of other variables available in the survey but not in the census, for small areas such as districts. Note that by using the survey alone, we would only be able to disaggregate at the region level, or occasionally at the provincial level.

Typical examples of welfare indicators are average expenditure, percentage of poor (with expenditure below poverty line), and poverty density (number of poor per area). The method enables us to determine the point estimates as well as the standard

[2] The work is done by researchers from Institute of Economic Research in HCM city (IER - HCM), General Statistical Office in HCM city (GSO - HCM), Ho Chi Minh Economic Institute, and GSO in Hanoi, and National Economics University in Hanoi.

errors associated with them. The standard errors are important because they make explicit the trade-off between the accuracy of the estimates and the level of disaggregation. While the standard errors for smaller geographic areas tend to be larger, the errors for poverty estimates based on a few thousand households (think of a district) are often found small enough to be acceptable.

The census either enjoys complete coverage or a very large coverage (in comparison to the survey). Due to the size of the census sampling error becomes negligible (and as such may safely be ignored). The basic idea behind the method is to replace a small number of exact observations of expenditure (using households from the survey) with a large number of estimates of expenditure (using households from the census) to obtain accurate estimates of aggregate welfare. This means that we will be replacing sampling error with model error. As model errors cancel out on average, the errors induced by model error tend to be small when the number of households is large.

To date, poverty maps have been produced in around fifty countries across the world. In the South-East Asia region alone, countries with a poverty map other than Vietnam include: Thailand, Lao PDR, Cambodia, Indonesia and the Philippines. Efforts to update the poverty map are under way in both Vietnam and the Philippines.

The ELL framework

Let us provide a brief review of the ELL methodology. In the standard setup, we consider the following model:

$$y_{ch} = x_{ch}^T \beta + \eta_c + \varepsilon_{ch},$$

where y_{ch} denotes the dependent variable (think of logarithmic per capita expenditure), x_{ch} the vector of explanatory variables, β the vector of regression coefficients, η the cluster-specific random effect and ε the household-specific random effect. The subscript *ch* refers to household *h* living in cluster *c*. The explanatory variables x_{ch} must be available in both census and survey.

Once all the parameters of interest have been identified, the dependent variable may be imputed into the census:

$$\hat{y}_{ch} = x_{ch}^T \hat{\beta} + \hat{\eta}_c + \hat{\varepsilon}_{ch},$$

where $\hat{\beta}$, $\hat{\eta}_c$ and $\hat{\varepsilon}_{ch}$ denote the estimates for β, η_c and ε_{ch}. Now suppose that we want to estimate the welfare indicator for a given district. As an illustrative example, let us consider the head-count index, which measures the percentage of poor households in the district:

$$W = \frac{1}{n}\sum_{ch} 1_{(y_{ch}<z)},$$

where $1_{(y<z)}$ denotes the indicator function that equals 1 if $y < z$ and 0 otherwise, and where n denotes the number of households living in the district. An estimate of W can be obtained by replacing y_{ch} with $\hat{y}_{ch}$ for all households ch.

To obtain an accurate estimate of the standard error of W, ELL advocate repeated Monte-Carlo simulations. In each round, a simulated regression coefficient $\tilde{\beta}^{(r)}$ is drawn (from its estimated distribution), where r denotes the r-th round of simulation. Further, $\tilde{\eta}_c^{(r)}$ and $\tilde{\varepsilon}_{ch}^{(r)}$ are drawn from their estimated distributions, which means we will have a simulated cluster error for each cluster and a simulated household error for each household in the census. The imputed dependent variable for household h in cluster c, in the r-th round, is therefore given by:

$$\tilde{y}_{ch}^{(r)} = x_{ch}^T \tilde{\beta}^{(r)} + \tilde{\eta}_c^{(r)} + \tilde{\varepsilon}_{ch}^{(r)}.$$

Each round of simulation yields a new estimate $\tilde{W}^{(r)}$. By taking the average and standard deviation over the R different simulated values of $\tilde{W}^{(r)}$, we obtain the point estimate and the standard error, respectively.

From a practical perspective, the approach is commonly divided into three stages:

Stage 0. *Selection of common and comparable variables*. This pre-stage involves the selection of variables that are available in both census and survey, which may be used as explanatory variables in the model for expenditure. Think of level of education, occupation, age, gender, ownership of (productive) assets, dwelling unit characteristics and village infrastructure. The key task here is to establish comparability of the variables, which involves two parts. First, we screen both questionnaires, searching for common questions and answers. Second, when the candidate variables have been constructed, we compare key statistics between census and survey. Naturally, having accurate survey weights will be of particular importance here. If they are not accurate, comparing statistics between survey and census tend to be unreliable, and as such less of a useful tool when deciding on comparability.

Stage 1. *Building regression models for per capita expenditure.* The objective of this stage is to build regression models that allow us to obtain accurate predictions of (log) expenditure. Naturally, accuracy of the SAE (of poverty) principally depends on the quality of the expenditure model, as well as on the quality of the explanatory variables (accurate measurement and a fair comparability between census and survey). The challenge here is to make sure that no important variables have been omitted. By the same token, the modeler needs to be careful not too overfit the data.

Stage 2. *Obtaining accurate standard errors by means of simulations.* As most welfare indicators are non-linear functions of the per capita expenditures (think of the head-count index), they will also be non-linear functions of the random variables involved (the (random) model parameters, the cluster errors and the household-specific errors), such that it will in general be very difficult to derive the standard errors of the welfare indicators. Note that even when we consider average (log) per capita expenditure, a model for the variance (the heteroskedasticity model) will introduce non-linearity of the welfare indicator with respect to the (random) model parameters. Accordingly, ELL advocate the use of bootstrapping to obtain robust estimates of the standard errors of the SAE, which can readily be implemented regardless of how complex the model is.

With the availability of POVMAP2, a software package developed by the World Bank to develop poverty maps, the user no longer needs to implement any of the procedures him/herself, as they have all been built in. The user can now concentrate all efforts on building the accurate model for expenditure, and on evaluating the results.

Two key assumptions

The ELL method is based on two key assumptions:

Model is accurate at each level it is applied: Tarozzi and Deaton (2007) refer to this as the `area homogeneity' assumption. Note that the model is typically estimated at the regional level (thereby often interacting variables with the urban/rural identifier), while the expenditure predictions using the model are aggregated over much smaller areas, think of provinces and districts, which together make up the region. Consistency requires that the model that accurately describes expenditure for each of these smaller areas is the same, and coincides with the model specified for the region (i.e. we assume there is no heterogeneity beyond the variation in the various explanatory variables across the small areas, hence the label `area homogeneity').

Spatial correlation is accurately accounted for: The model error for different households are likely to exhibit a level of correlation, in particular when the households live close to each other such that they are subject to similar geographical effects. An accurate account of this spatial correlation is important for the accuracy of the standard errors of our SAEs, as we will illustrate later.

ELL assumes that the model error can be decomposed into a cluster error (an error that is shared by all households living in the same cluster) and a household specific error. The common error is referred to as location error. The household specific error will also be referred to as idiosyncratic error. Empirical results accumulated to date (in a wide range of countries) indicate that spatial correlation is significant, and that the approach put forward by ELL works quite well.

Any violation of these assumptions will plausibly affect the accuracy of the SAE of welfare. Therefore, each time the methods is used, it is important that the user tests the validity of these assumptions, as this may vary from country to country. Both assumptions, but in particular the assumptions regarding spatial correlation, will be tested extensively in this study.

Accurate standard errors via accurate account of spatial correlation

Let us briefly illustrate the importance of spatial correlation for the standard errors of the SAEs by means of a simple example. We will consider average (log) per capita expenditure as our indicator of aggregate welfare. The model will be:

$$y_{ch} = x_{ch}^T \beta + u_{ch},$$

where the variance of u_{ch} is assumed constant, $\text{var}[u_{ch}] = \sigma_u^2$. Accordingly, assuming we have identified the correct model, the error in our indicator of welfare equals:

$$\frac{1}{n}\sum_{ch} y_{ch} - E[y_{ch} \mid x_{ch}] = \frac{1}{n}\sum_{ch} u_{ch},$$

where n denotes the number of households living in the area of interest.

To appreciate the effect of spatial correlation it may be insightful to distinguish two extreme cases: independently distributed errors versus perfectly correlated errors. When the errors u_{ch} are independent of each other, the variance of the error in aggregate welfare solves: $\text{var}[\sum u_{ch} / n] = \sigma_u^2 / n$. This means that the error will rapidly tend to zero as the number of households n increases. In contrast, when the errors u_{ch} are perfectly correlated, the variance equals: $\text{var}[\sum u_{ch} / n] = \sigma_u^2$. In other words, the precision of our estimate does not increase at all as n becomes larger. Naturally, any realistic scenario is one that lies somewhere in between these two extremes.

Now consider the model assumption made by ELL:

$$u_{ch} = \eta_c + \varepsilon_{ch}.$$

All households living in cluster c share a cluster error η_c with variance σ_η^2. The errors for households from different clusters are assumed uncorrelated. For simplicity, the variance of the household specific error ε_{ch} is also assumed constant, $\text{var}[\varepsilon_{ch}] = \sigma_\varepsilon^2$. Let the number of clusters in our area of interest be denotes by $k \ll n$. It can be verified that the variance of the error in aggregate welfare is now given by:

$$\text{var}\left[\frac{1}{n}\sum_{ch} u_{ch}\right] = \frac{\sigma_\eta^2}{k} + \frac{\sigma_\varepsilon^2}{n}.$$

Note how this indeed falls in between the two extremes: $\sigma_u^2 / n \leq \sigma_\eta^2 / k + \sigma_\varepsilon^2 / n \leq \sigma_u^2$, where $\sigma_u^2 = \sigma_\eta^2 + \sigma_\varepsilon^2$. The error tends to zero if and only if both the number of households and the number of clusters tend to infinity. In practice, the number of households obviously is much larger than the number of clusters, such that the variance of the location error tends to play an important role in the total variance.

Naturally, if one decides to ignore spatial correlation, while it is in fact present, one runs the risk of significantly underestimating the standard errors, and hence overestimating precision. The original poverty map for Vietnam was nevertheless based on the assumption of no spatial correlation. Which of the assumption applies to Vietnam is one of the key empirical questions addressed by this study.

3. DATA SOURCES

The research relies on two data sources to estimate poverty and inequality for the districts of HCM city. The first is the Vietnam Household Living Standard Survey (VHLSS) conducted by the General Statistical Office of Vietnam (GSO) in 2004. The survey collects information on household characteristics including basic demography, employment and labor force participation, education, health, income, expenditure, housing, fixed assets and durable goods, and the participation of households in the most important poverty alleviation programs.

The VHLSS 2004 covers 9000 households. This sample is representative at the regional level, but not at the provincial level. We will consider two sets of income models, one based on the survey sample for HCM city only, and another based on a larger sample that covers the entire (urban) South-East region. As the VHLSS 2004 merely includes 300 households from HCM city, we will use this sample for small models only. The larger sample for the South-East region, with 1188 households, allows us to consider more elaborate model specifications.

The second data source is a 10% sample Population and Housing Mid-Census for HCM city in 2004. The census collects information on basic demography, education of people, unemployment status, and several characteristics on housing and assets. The census sample is designed to be representative at the district level.

4. VARIABLE COMPARISON

The variables used in the income models should meet the following criteria:

- Available in both the survey and the census.
- Comparable between the survey and census, i.e., they are constructed in similar definitions and have similar distribution.
- Correlated with household income.

This section is to present descriptive statistics of the common variables in the VHLSS 2004 and the HCM city Mid-Census 2004. The results are presented in Table 1. Overall, the mean and standard deviation of the variables included are fairly similar between the VHLSS 2004 and the HCM city Mid-Census 2004, which confirms their comparability.

Table 1: Common variable between 2004 Mid-Census and VHLSS 2004 for HCM

Common variable	Type	VHLSS		Mid-Census	
		Mean	Std. Dev.	Mean	Std. Dev.
Number of observations		**300**		**92367**	
% Head ethnic minorities	Binary	0.94		0.40	
% Head male	Binary	52.09		56.06	
Age of head	Continuous	52.5	13.4	50.5	15.2
% Head working	Binary	64.84		65.85	
Education	Categorical				
% head primary school		35.76		31.15	
% Head lower-secondary		23.49		29.80	
% Head upper-secondary		28.65		29.25	
% Head post-secondary		12.11		9.80	
Total		**100**		**100**	
% Households with					
Television	Binary	96.92		91.56	
Radio	Binary	27.66		46.53	
Computer	Binary	34.72		22.14	

Common variable	Type	VHLSS		Mid-Census	
		Mean	Std. Dev.	Mean	Std. Dev.
Internet	Binary	11.37		6.24	
Telephone	Binary	61.72		52.20	
Mobile phone	Binary	35.00		38.52	
Housing types	Categorical				
Permanent house		37.01		28.54	
Semi-Permanent		57.74		64.30	
Temporary		5.25		7.16	
Total		**100**		**100**	
Toilet type	Categorical				
Flush		88.69		89.41	
Others		10.73		9.03	
toilet		0.59		1.56	
Total		**100**		**100**	
Water type	Categorical				
Tap-water		59.96		46.82	
Filtered water		38.74		51.74	
others		1.30		1.43	
Total		**100**		**100**	
Household size	Categorical				
1		5.09		5.83	
2		6.39		10.48	
3		19.75		17.61	
4		27.30		25.10	
5		16.18		15.09	
6		12.52		10.12	
>=7		12.77		**15.76**	
Total		**100**		**100**	
Number of female					
0		2.14		4.99	
1		25.06		25.49	
2		34.05		31.05	
3		21.88		19.66	
>=4		16.87		18.81	
Total		**100**		**100**	
Number of children	Categorical				
0		45.67		42.6	
1		27.99		30	
2		19.72		20.33	
3		4.84		4.88	
>=4		1.78		2.19	
Total		**100**		**100**	
Number of elderly	Categorical				
0		67.08		73.05	
1		21.19		19.68	
2		11.73		7.01	
3		0		0.23	
>=4		0		0.02	
Total		**100**		**100**	
Ratio of female	Continuous	0.531	0.184	0.525	0.199

Common variable	Type	VHLSS		Mid-Census	
		Mean	Std. Dev.	Mean	Std. Dev.
Ratio of children	Continuous	0.200	0.188	0.209	0.183
Ratio of elderly	Continuous	0.100	0.170	0.076	0.141
Number of working members	Categorical				
0		5.21		5.87	
1		21.42		26.13	
2		39.72		32.81	
3		15.24		15.41	
>=4		18.40		19.79	
Total		**100**		**100**	
Number of members with primary school	Categorical				
0		29.00		28.69	
1		28.37		29.11	
2		24.68		20.85	
3		10.61		10.39	
>=4		7.33		10.96	
Total		**100**		**100**	
Number of members with lower-secondary	Categorical				
0		33.02		31.48	
1		29.42		29.67	
2		23.13		20.77	
3		7.94		10.09	
>=4		6.49		8.00	
Total		**100**		**100**	
Number of members with upper-secondary	Categorical				
0		31.05		38.12	
1		31.93		26.90	
2		20.92		19.12	
3		9.24		8.87	
>=4		6.87		6.99	
Total		**100**		**100**	
Number of members with post-secondary					
0		74.51		80.46	
1		14.35		11.13	
2		9.36		5.69	
3		0.71		1.77	
>=4		1.07		0.96	
Total		**100**		**100**	
Ratio of working members	Continuous	0.521	0.205	0.523	0.240
Ratio of primary school members	Continuous	0.327	0.264	0.348	0.272
Ratio of lower-secondary members	Continuous	0.287	0.241	0.307	0.253
Ratio of upper-secondary members	Continuous	0.297	0.258	0.274	0.270
Ratio of post-secondary members	Continuous	0.089	0.178	0.070	0.168

Source: Authors' estimation

Table 2: OLS regression on log of income per capita for South East

Explanatory variables	Sample of South East Region						Sample of HCM city			
	Model 1		Model 2		Model 3		Model 4		Model 5	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
intercept	9.0856	0.0930	9.3091	0.0566	9.0728	0.0483	9.4029	0.0741	9.4928	0.0697
Having computer	0.3091	0.0932					0.2062	0.0670	0.3111	0.0614
HCM city	0.3894	0.0724	0.3056	0.0353	0.3838	0.0372				
Household size	-0.0858	0.0073	-0.0867	0.0072	-0.0797	0.0076	-0.0704	0.0129	-0.0827	0.0129
Permanent house	0.0907	0.0351								
Temporary house	-0.2203	0.0429	-0.2642	0.0427			-0.2449	0.1155		
Using internet connection	0.2086	0.0663	0.3097	0.0621			0.1868	0.0894		
Using mobile phone	0.2226	0.0386	0.2471	0.0380			0.1820	0.0589	0.2049	0.0598
Ratio of primary school members	-0.4250	0.0678	-0.3889	0.0554	-0.3991	0.0488				
Ratio of lower-secondary school members	-0.2230	0.0620	-0.2431	0.0619						
Ratio of post-secondary school members							0.3444	0.1362		
Ratio of female members	-0.1356	0.0632								
Ratio of working members	0.2655	0.0735								
Using desk telephone	0.2260	0.0341	0.2774	0.0335	0.4369	0.0336				
Have no toilet	-0.2472	0.0507	-0.2715	0.0514						
TV_1	0.1475	0.0483								
District variables										
% household without toilet							-1.7461	0.8547	-2.6314	0.8549
Interaction variables										
HCM city * Ratio of elderly members	0.2344	0.0930					0.2268	0.1112		
HCM city * Ratio of working members	-0.2603	0.1135								
Urban areas * Having computer	-0.1957	0.0960								

Explanatory variables	Sample of South East Region						Sample of HCM city			
	Model 1		Model 2		Model 3		Model 4		Model 5	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Urban areas * have garbage treatment	0.0950	0.0441	0.1370	0.0382	0.2211	0.0403				
Urban areas * Ratio of primary school members	0.1828	0.0665								
Urban areas * head with upper-secondary school							0.1717	0.0576		
Urban areas * telephone							0.1990	0.0589	0.2795	0.0586
Number of observations	1188		1188		1188			300		300
Number of regressors	100		100		100			100		100
Number of regressors in model	19		10		5			6		6
Adjusted R squared	0.5988		0.5800		0.5147			0.4518		0.4066
Number of clusters in survey	75		75		75			22		22
Number of clusters in census	24		24		24			24		24
$\hat{\sigma}_{\eta}^{2} / \hat{\sigma}_{u}^{2}$	0.0863		0.0839		0.1065					

Note: *(i) Estimation from VHLSS 2004 – sample of South East region*
(ii) Districts are specified as clusters.
(iii) There is no cluster variable used in the regressions.
(iv) Alpha models of error heterogeneity are kept small with 4 explanatory variables.

5. INCOME MODELS

This section reports the results from the income model regressions using the VHLSS 2004. To examine the sensitivity of the poverty estimates to model specifications, we compared 5 different models, which mostly vary in the number of explanatory variables they included. Models 1, 2 and 3 are estimated using the sample for the Southeast region, and refer to a large, medium, and a relatively small specification. For models 4 and 5 we used the HCM city sample of 300 households. The number of explanatory variables included in model 4 is larger than for model 5.

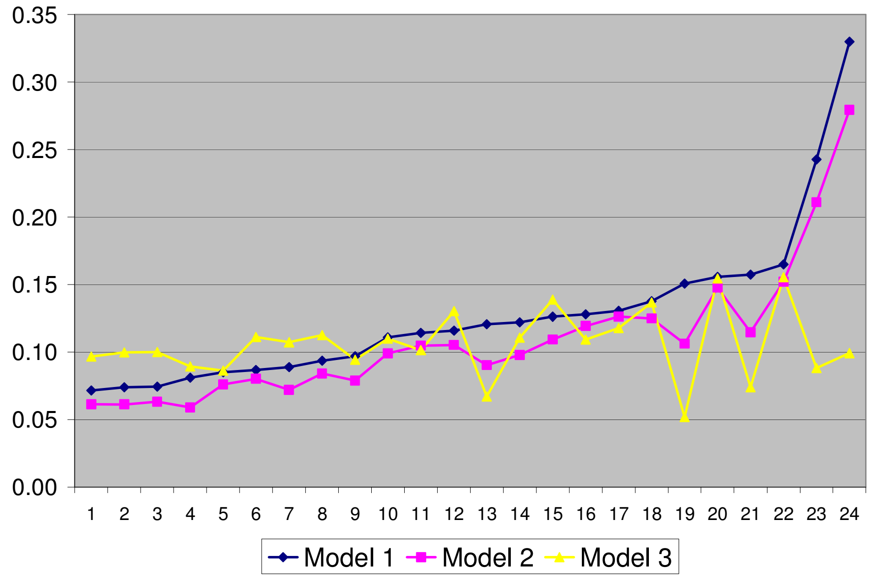
Table 2 shows that variables on housing, household assets, and education are strongly correlated with household income. Models 1 and 2, denoting the large and medium sized specifications, obtain a relatively high R-squared, and manage to account for much of the spatial correlation (the location error as part of the total model error is small, see the bottom row in Table 2). The location error has not been included in the specifications of models 4 and 5, as the HCM sample in the VHLSS 2004 is too small to obtain reliable estimates of the distribution of the location error.

6. WELFARE ESTIMATES

Once the income equations are estimated, they can be applied in the Mid-Census sample to estimate the poverty rate of districts of HCM city. The poverty line used in this study is equal to 6000 thousand VND. This poverty line comes from HCM City People's Com. - Decision No. 145/2004/QĐ-UB on 25/5/2004 on poverty reduction strategy of HCMC. Using these poverty lines allows for comparison of the estimated poverty indexes with poverty estimates reported by other State agencies. The national poverty is not very suitable for HCM city, since the poverty rate of HCM city using this poverty is very low, close to 0%.

Table 3 presents the estimates of poverty incidence (P0) of districts in HCM city for 5 Models. It shows that except for Model 3 which is very small, all the four Models give quite similar ranking of district poverty. Since the data sample of the 2004 VHLSS are not representative for the HCM city, we will not use Models 4 and 5 for final estimation of poverty and inequality. Figure 1 graphs the poverty incidence estimates of Model 1, 2 and 3.

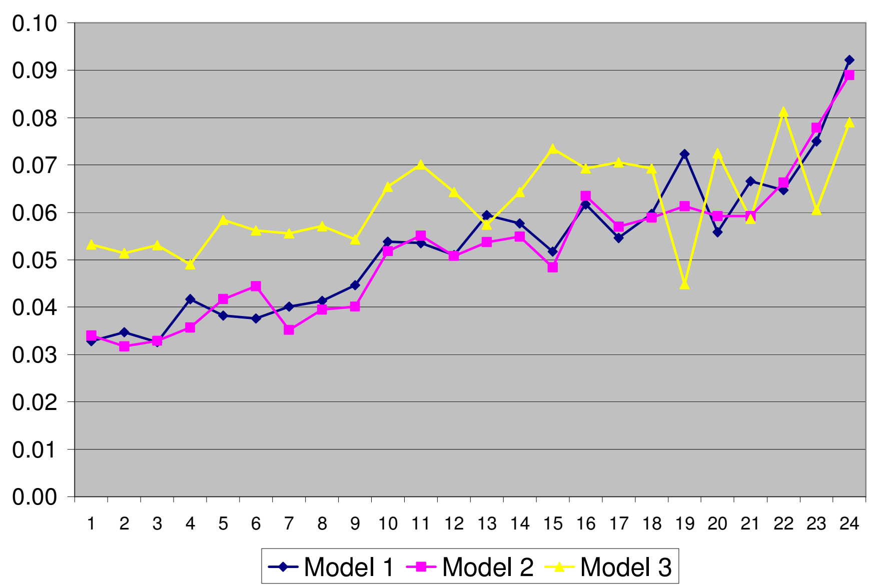
Figure 1: Estimates of poverty headcount index in three models



Source: Authors' estimation

The standard errors of the there models are graphed in Figure 2. Model 1 and 2 result in very close standard errors, while Model 3 produces much higher standard errors.

Figure 2: Standard errors of estimates of poverty headcount index in three models

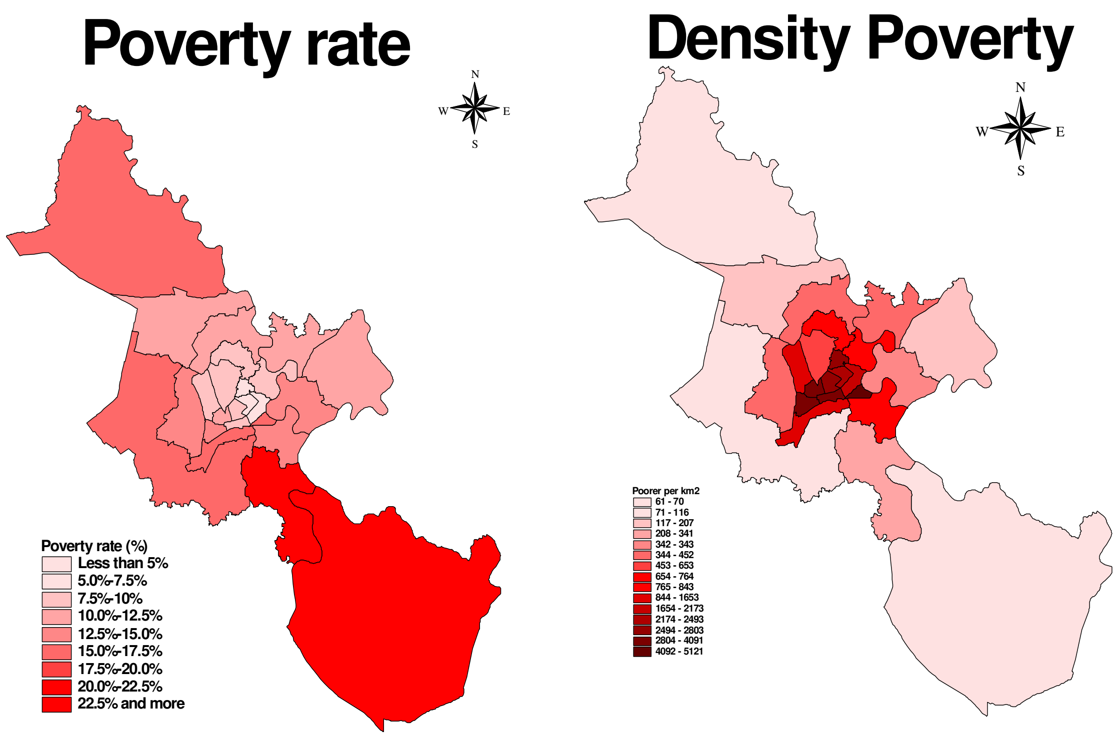


Source: Authors' estimation

According to Models 1 and 2, the poorest district is Can Gio, followed by Nha Be. Many other districts have poverty rates lower than 10%. The District 1 and 3 have lowest poverty estimates. Figure 3 graphs the map of district poverty rates estimated from Model 1. However, the poverty density, which is expressed as the number of poor per kilometer squared, is highest in district 1 and 3 and lowest in Can Gio and Nha Be. The pictures of poverty incidence and poverty density are opposites, since the population density in the rich districts is much higher than in the poor districts.

Tables 4 and 5 present the estimates of poverty-gap and poverty-severity indexes of districts. Again, Can Gio and Nha Be are two districts having highest poverty depth and severity in HCM city. Table 6 presents the estimates of Gini index for the districts.

Figure 3: Poverty estimates of districts of HCM city in 2004



Source: Authors' estimation

Table 3 Estimates of headcount index (P0) at the district and provincial levels

District		No. sampled hhs	Model 1		Model 2		Model 3		Model 4		Model 5	
Code	Name		P0	Std. error.	P0	Std. error.	P0	Std. error.	P0	Std. error.	P0	Std. error.
1	Quận 1	4490	0.0739	0.0347	0.0613	0.0317	0.0998	0.0514	0.0477	0.0087	0.0655	0.0091
3	Quận 2	3106	0.1305	0.0546	0.1264	0.0570	0.1179	0.0706	0.1066	0.0157	0.0987	0.0168
5	Quận 3	3576	0.0745	0.0326	0.0632	0.0329	0.1001	0.0531	0.0499	0.0095	0.0720	0.0107
7	Quận 4	2635	0.1557	0.0558	0.1477	0.0592	0.1545	0.0725	0.1230	0.0178	0.1434	0.0167
9	Quận 5	3455	0.0868	0.0376	0.0802	0.0444	0.1112	0.0562	0.0580	0.0110	0.0825	0.0125
11	Quận 6	3734	0.1262	0.0517	0.1094	0.0484	0.1389	0.0735	0.0871	0.0152	0.1079	0.0146
13	Quận 7	4211	0.1378	0.0597	0.1250	0.0589	0.1367	0.0693	0.1012	0.0150	0.1096	0.0131
15	Quận 8	3902	0.1650	0.0647	0.1521	0.0663	0.1555	0.0813	0.1136	0.0163	0.1249	0.0163
17	Quận 9	4496	0.1108	0.0538	0.0991	0.0518	0.1100	0.0654	0.0751	0.0124	0.0738	0.0169
19	Quận 10	3630	0.0889	0.0401	0.0721	0.0352	0.1073	0.0556	0.0613	0.0112	0.0781	0.0111
21	Quận 11	3703	0.1160	0.0510	0.1053	0.0508	0.1304	0.0643	0.0729	0.0137	0.0850	0.0121
23	Quận 12	4332	0.1142	0.0535	0.1049	0.0551	0.1014	0.0701	0.0752	0.0136	0.1098	0.0157
25	Quận Gò Vấp	4007	0.0851	0.0382	0.0761	0.0417	0.0863	0.0584	0.0579	0.0105	0.0735	0.0106
27	Quận Tân Bình	3820	0.0811	0.0417	0.0590	0.0357	0.0895	0.0490	0.0491	0.0097	0.0638	0.0095
28	Quận Tân Phú	4396	0.0970	0.0446	0.0789	0.0401	0.0944	0.0543	0.0625	0.0118	0.0751	0.0108
29	Quận Bình Thạnh	3844	0.0937	0.0413	0.0841	0.0395	0.1126	0.0571	0.0712	0.0118	0.0931	0.0130
31	Quận Phú Nhuận	4126	0.0715	0.0328	0.0614	0.0340	0.0967	0.0532	0.0476	0.0092	0.0768	0.0119
33	Quận Thủ Đức	4221	0.1220	0.0577	0.0979	0.0549	0.1107	0.0643	0.0709	0.0127	0.0595	0.0195
34	Quận Bình Tân	3750	0.1280	0.0617	0.1194	0.0635	0.1093	0.0693	0.0817	0.0142	0.0897	0.0142
35	Huyện Củ Chi	4254	0.1508	0.0723	0.1062	0.0613	0.0520	0.0448	0.1550	0.0240	0.1461	0.0220
37	Huyện Hóc Môn	4039	0.1206	0.0594	0.0904	0.0537	0.0672	0.0574	0.1321	0.0228	0.1718	0.0217
39	Huyện Bình Chánh	4318	0.1574	0.0666	0.1148	0.0592	0.0739	0.0586	0.1726	0.0256	0.1486	0.0298
41	Huyện Nhà Bè	3054	0.2426	0.0750	0.2110	0.0779	0.0882	0.0605	0.3822	0.1004	0.4329	0.0804
43	Huyện Cần Giờ	3268	0.3300	0.0922	0.2794	0.0890	0.0992	0.0790	0.5103	0.1312	0.5859	0.1372
	All HCM city	**92367**	**0.1172**	**0.0188**	**0.0996**	**0.0181**	**0.1036**	**0.0219**	**0.0930**	**0.0126**	**0.1055**	**0.0117**

Source: Authors' estimation

Table 4 Estimates of poverty gap index (P1) at the district and provincial levels

District		No. sampled hhs	**Model 1**		**Model 2**		**Model 3**		**Model 4**		**Model 5**	
Code	Name		P1	Std. error.	P1	Std. error.	P1	Std. error.	P1	Std. error.	P1	Std. error.
1	Quận 1	4490	0.0160	0.0087	0.0119	0.0074	0.0232	0.0140	0.0091	0.0022	0.0147	0.0028
3	Quận 2	3106	0.0284	0.0145	0.0263	0.0148	0.0254	0.0184	0.0205	0.0042	0.0195	0.0046
5	Quận 3	3576	0.0166	0.0083	0.0126	0.0078	0.0236	0.0145	0.0099	0.0026	0.0171	0.0034
7	Quận 4	2635	0.0374	0.0160	0.0322	0.0164	0.0374	0.0212	0.0264	0.0058	0.0362	0.0063
9	Quận 5	3455	0.0186	0.0092	0.0156	0.0107	0.0259	0.0152	0.0111	0.0029	0.0195	0.0041
11	Quận 6	3734	0.0280	0.0136	0.0217	0.0118	0.0321	0.0207	0.0173	0.0042	0.0258	0.0049
13	Quận 7	4211	0.0313	0.0163	0.0258	0.0157	0.0315	0.0194	0.0203	0.0044	0.0245	0.0041
15	Quận 8	3902	0.0390	0.0184	0.0331	0.0185	0.0357	0.0227	0.0232	0.0048	0.0282	0.0056
17	Quận 9	4496	0.0222	0.0130	0.0187	0.0120	0.0227	0.0158	0.0129	0.0028	0.0131	0.0043
19	Quận 10	3630	0.0202	0.0110	0.0142	0.0084	0.0255	0.0157	0.0124	0.0032	0.0185	0.0036
21	Quận 11	3703	0.0248	0.0130	0.0203	0.0125	0.0293	0.0173	0.0138	0.0036	0.0189	0.0036
23	Quận 12	4332	0.0226	0.0125	0.0192	0.0124	0.0190	0.0166	0.0131	0.0032	0.0243	0.0048
25	Quận Gò Vấp	4007	0.0170	0.0088	0.0140	0.0094	0.0172	0.0144	0.0106	0.0026	0.0155	0.0030
27	Quận Tân Bình	3820	0.0170	0.0102	0.0108	0.0078	0.0196	0.0123	0.0090	0.0023	0.0138	0.0028
28	Quận Tân Phú	4396	0.0204	0.0110	0.0146	0.0089	0.0194	0.0129	0.0117	0.0030	0.0163	0.0032
29	Quận Bình Thạnh	3844	0.0215	0.0110	0.0174	0.0100	0.0267	0.0159	0.0150	0.0035	0.0232	0.0046
31	Quận Phú Nhuận	4126	0.0156	0.0082	0.0120	0.0078	0.0225	0.0140	0.0092	0.0024	0.0186	0.0041
33	Quận Thủ Đức	4221	0.0250	0.0141	0.0180	0.0128	0.0231	0.0155	0.0128	0.0031	0.0108	0.0048
34	Quận Bình Tân	3750	0.0263	0.0152	0.0227	0.0155	0.0212	0.0162	0.0150	0.0035	0.0181	0.0039
35	Huyện Củ Chi	4254	0.0299	0.0175	0.0189	0.0133	0.0079	0.0079	0.0278	0.0059	0.0278	0.0060
37	Huyện Hóc Môn	4039	0.0246	0.0142	0.0163	0.0121	0.0126	0.0129	0.0249	0.0057	0.0408	0.0073
39	Huyện Bình Chánh	4318	0.0339	0.0175	0.0226	0.0139	0.0131	0.0126	0.0332	0.0065	0.0279	0.0083
41	Huyện Nhà Bè	3054	0.0618	0.0241	0.0510	0.0247	0.0155	0.0126	0.0937	0.0333	0.1138	0.0300
43	Huyện Cần Giờ	3268	0.0866	0.0333	0.0681	0.0296	0.0163	0.0159	0.1383	0.0534	0.1424	0.0529
	All HCM city	**92367**	**0.0253**	**0.0049**	**0.0196**	**0.0047**	**0.0221**	**0.0057**	**0.0184**	**0.0035**	**0.0035**	**0.0037**

Source: Authors' estimation

Table 5 Estimates of poverty severity index (P2) at the district and provincial levels

District		No. sampled hhs	Model 1		Model 2		Model 3		Model 4		Model 5	
Code	Name		P2	Std. error.	P2	Std. error.	P2	Std. error.	P2	Std. error.	P2	Std. error.
1	Quận 1	4490	0.0055	0.0032	0.0037	0.0026	0.0084	0.0056	0.0028	0.0009	0.0052	0.0013
3	Quận 2	3106	0.0096	0.0055	0.0084	0.0055	0.0086	0.0070	0.0063	0.0016	0.0062	0.0018
5	Quận 3	3576	0.0059	0.0032	0.0040	0.0028	0.0087	0.0059	0.0033	0.0011	0.0064	0.0015
7	Quận 4	2635	0.0136	0.0065	0.0106	0.0063	0.0138	0.0088	0.0088	0.0025	0.0138	0.0032
9	Quận 5	3455	0.0063	0.0034	0.0048	0.0038	0.0093	0.0060	0.0034	0.0011	0.0072	0.0019
11	Quận 6	3734	0.0097	0.0052	0.0067	0.0042	0.0114	0.0083	0.0055	0.0017	0.0095	0.0023
13	Quận 7	4211	0.0110	0.0064	0.0082	0.0060	0.0112	0.0078	0.0064	0.0018	0.0086	0.0019
15	Quận 8	3902	0.0142	0.0075	0.0110	0.0073	0.0127	0.0091	0.0075	0.0021	0.0100	0.0026
17	Quận 9	4496	0.0070	0.0046	0.0055	0.0041	0.0074	0.0057	0.0036	0.0010	0.0038	0.0016
19	Quận 10	3630	0.0072	0.0044	0.0044	0.0030	0.0094	0.0064	0.0040	0.0013	0.0068	0.0017
21	Quận 11	3703	0.0082	0.0048	0.0060	0.0044	0.0102	0.0067	0.0042	0.0014	0.0066	0.0016
23	Quận 12	4332	0.0070	0.0044	0.0055	0.0041	0.0057	0.0059	0.0037	0.0011	0.0084	0.0021
25	Quận Gò Vấp	4007	0.0054	0.0031	0.0041	0.0031	0.0055	0.0053	0.0032	0.0010	0.0052	0.0013
27	Quận Tân Bình	3820	0.0056	0.0037	0.0032	0.0026	0.0069	0.0047	0.0027	0.0009	0.0047	0.0012
28	Quận Tân Phú	4396	0.0067	0.0041	0.0042	0.0030	0.0063	0.0046	0.0036	0.0012	0.0056	0.0014
29	Quận Bình Thạnh	3844	0.0077	0.0043	0.0057	0.0037	0.0098	0.0065	0.0050	0.0015	0.0089	0.0022
31	Quận Phú Nhuận	4126	0.0054	0.0031	0.0037	0.0027	0.0082	0.0055	0.0029	0.0010	0.0070	0.0019
33	Quận Thủ Đức	4221	0.0081	0.0051	0.0053	0.0044	0.0077	0.0056	0.0038	0.0011	0.0033	0.0018
34	Quận Bình Tân	3750	0.0085	0.0055	0.0068	0.0055	0.0066	0.0057	0.0045	0.0013	0.0059	0.0016
35	Huyện Củ Chi	4254	0.0092	0.0061	0.0053	0.0043	0.0020	0.0022	0.0079	0.0021	0.0083	0.0023
37	Huyện Hóc Môn	4039	0.0080	0.0050	0.0048	0.0041	0.0039	0.0044	0.0076	0.0021	0.0151	0.0034
39	Huyện Bình Chánh	4318	0.0112	0.0066	0.0070	0.0048	0.0038	0.0041	0.0100	0.0024	0.0084	0.0032
41	Huyện Nhà Bè	3054	0.0230	0.0104	0.0182	0.0105	0.0044	0.0040	0.0332	0.0140	0.0428	0.0138
43	Huyện Cần Giờ	3268	0.0326	0.0152	0.0241	0.0128	0.0043	0.0048	0.0522	0.0250	0.0488	0.0236
	All HCM city	**92367**	**0.0085**	**0.0019**	**0.0061**	**0.0017**	**0.0075**	**0.0022**	**0.0058**	**0.0014**	**0.0081**	**0.0016**

Source: Authors' estimation

Table 6 Estimates of Gini at the district and provincial levels

District		No. sampled hhs	Model 1		Model 2		Model 3		Model 4		Model 5	
Code	Name		Gini	Std. error.	Gini	Std. error.	Gini	Std. error.	Gini	Std. error.	Gini	Std. error.
1	Quận 1	4490	0.3198	0.0168	0.3204	0.0195	0.3234	0.0188	0.3134	0.0202	0.3130	0.0167
3	Quận 2	3106	0.3041	0.0119	0.3054	0.0133	0.2882	0.0128	0.2865	0.0137	0.2697	0.0156
5	Quận 3	3576	0.3137	0.0159	0.3131	0.0184	0.3183	0.0192	0.3077	0.0186	0.3158	0.0169
7	Quận 4	2635	0.3206	0.0129	0.3209	0.0141	0.3188	0.0151	0.2978	0.0139	0.3247	0.0148
9	Quận 5	3455	0.3082	0.0138	0.3102	0.0162	0.3135	0.0165	0.2987	0.0163	0.3213	0.0174
11	Quận 6	3734	0.3017	0.0116	0.3021	0.0130	0.3048	0.0137	0.2826	0.0134	0.3164	0.0150
13	Quận 7	4211	0.3091	0.0123	0.3121	0.0143	0.3122	0.0155	0.2901	0.0135	0.3034	0.0129
15	Quận 8	3902	0.3064	0.0118	0.3077	0.0137	0.2978	0.0123	0.2809	0.0131	0.2864	0.0145
17	Quận 9	4496	0.2867	0.0113	0.2875	0.0133	0.2828	0.0128	0.2634	0.0114	0.2534	0.0180
19	Quận 10	3630	0.3156	0.0155	0.3130	0.0173	0.3196	0.0180	0.3078	0.0176	0.3151	0.0153
21	Quận 11	3703	0.3092	0.0132	0.3117	0.0146	0.3078	0.0137	0.2958	0.0159	0.3066	0.0139
23	Quận 12	4332	0.2806	0.0105	0.2825	0.0122	0.2554	0.0148	0.2592	0.0113	0.3027	0.0155
25	Quận Gò Vấp	4007	0.2924	0.0130	0.2954	0.0152	0.2707	0.0167	0.2843	0.0142	0.2927	0.0125
27	Quận Tân Bình	3820	0.2990	0.0141	0.2959	0.0164	0.2962	0.0159	0.2896	0.0149	0.2901	0.0130
28	Quận Tân Phú	4396	0.2967	0.0125	0.2961	0.0145	0.2742	0.0155	0.2801	0.0136	0.2913	0.0126
29	Quận Bình Thạnh	3844	0.3165	0.0149	0.3156	0.0164	0.3173	0.0170	0.3133	0.0170	0.3231	0.0173
31	Quận Phú Nhuận	4126	0.3161	0.0179	0.3177	0.0201	0.3163	0.0182	0.3127	0.0206	0.3280	0.0216
33	Quận Thủ Đức	4221	0.2948	0.0120	0.2948	0.0141	0.2830	0.0131	0.2737	0.0129	0.2462	0.023
34	Quận Bình Tân	3750	0.2746	0.0101	0.2764	0.0122	0.2549	0.0147	0.2477	0.0112	0.2661	0.0134
35	Huyện Củ Chi	4254	0.2679	0.0094	0.2605	0.0107	0.2360	0.0142	0.2242	0.0111	0.2374	0.0152
37	Huyện Hóc Môn	4039	0.2930	0.0106	0.2763	0.0114	0.2595	0.0143	0.2341	0.0113	0.2884	0.0170
39	Huyện Bình Chánh	4318	0.2935	0.0098	0.2770	0.0105	0.2576	0.0118	0.2298	0.0114	0.2211	0.0219
41	Huyện Nhà Bè	3054	0.3147	0.0097	0.3035	0.0111	0.2562	0.0121	0.2519	0.0126	0.2585	0.0139
43	Huyện Cần Giờ	3268	0.3008	0.0101	0.2937	0.0106	0.2317	0.0146	0.2359	0.0138	0.1887	0.0295
	All HCM city	**92367**	**0.3138**	**0.0119**	**0.3113**	**0.0136**	**0.3011**	**0.0133**	**0.2952**	**0.0131**	**0.3053**	**0.0119**

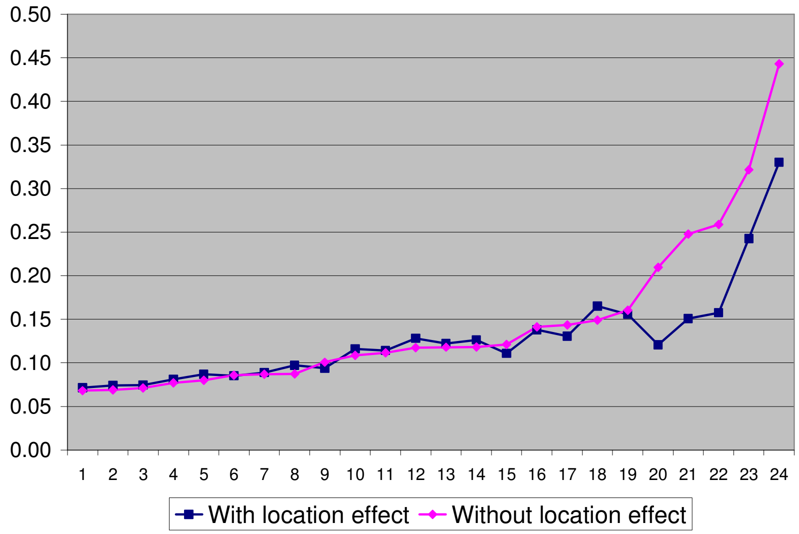
Source: Authors' estimation

Table 7 Welfare estimates in models without location effect

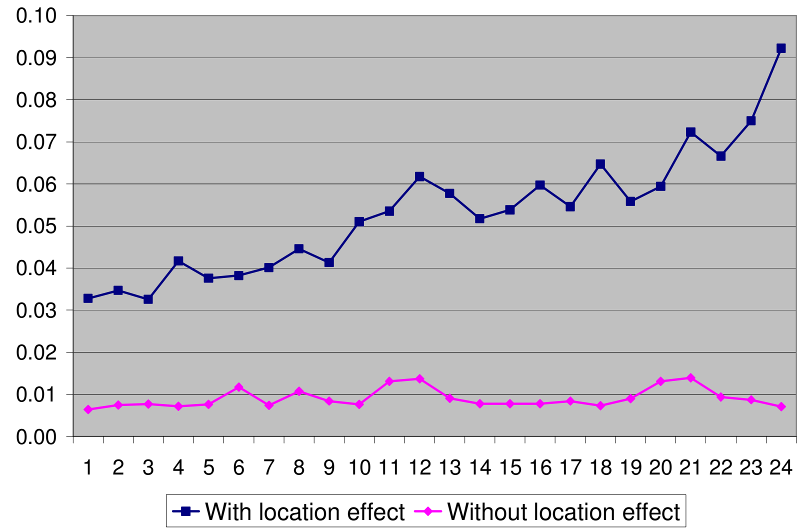
District		No. sampled hhs	P0		P1		P2		Gini	
Code	Name		Est.	Std. error.	Est.	Std. error.	Est.	Std. error.	Est.	Std. error.
1	Quận 1	4490	0.0686	0.0092	0.0155	0.0027	0.0055	0.0012	0.3294	0.0179
3	Quận 2	3106	0.1410	0.0151	0.0325	0.0045	0.0115	0.0019	0.3153	0.0123
5	Quận 3	3576	0.0710	0.0100	0.0170	0.0030	0.0064	0.0014	0.3247	0.0180
7	Quận 4	2635	0.1570	0.0155	0.0402	0.0054	0.0155	0.0027	0.3340	0.0127
9	Quận 5	3455	0.0788	0.0102	0.0180	0.0031	0.0064	0.0014	0.3191	0.0155
11	Quận 6	3734	0.1172	0.0129	0.0278	0.0041	0.0102	0.0019	0.3164	0.0129
13	Quận 7	4211	0.1340	0.0137	0.0318	0.0044	0.0116	0.0020	0.3231	0.0124
15	Quận 8	3902	0.1472	0.0153	0.0360	0.0050	0.0136	0.0023	0.3205	0.0126
17	Quận 9	4496	0.1179	0.0137	0.0249	0.0039	0.0082	0.0016	0.3005	0.0119
19	Quận 10	3630	0.0851	0.0108	0.0204	0.0035	0.0076	0.0016	0.3269	0.0174
21	Quận 11	3703	0.1075	0.0134	0.0244	0.0039	0.0086	0.0017	0.3229	0.0141
23	Quận 12	4332	0.1126	0.0140	0.0236	0.0037	0.0077	0.0014	0.2937	0.0104
25	Quận Gò Vấp	4007	0.0844	0.0113	0.0181	0.0030	0.0061	0.0012	0.3039	0.0143
27	Quận Tân Bình	3820	0.0739	0.0106	0.0160	0.0031	0.0055	0.0014	0.3103	0.0151
28	Quận Tân Phú	4396	0.0872	0.0109	0.0195	0.0032	0.0068	0.0014	0.3070	0.0130
29	Quận Bình Thạnh	3844	0.0966	0.0119	0.0237	0.0038	0.0090	0.0018	0.3294	0.0168
31	Quận Phú Nhuận	4126	0.0700	0.0097	0.0161	0.0029	0.0058	0.0013	0.3276	0.0202
33	Quận Thủ Đức	4221	0.1153	0.0150	0.0248	0.0042	0.0084	0.0017	0.3046	0.0121
34	Quận Bình Tân	3750	0.1179	0.0145	0.0257	0.0041	0.0088	0.0017	0.2896	0.0105
35	Huyện Củ Chi	4254	0.2597	0.0241	0.0603	0.0078	0.0209	0.0034	0.2846	0.0094
37	Huyện Hóc Môn	4039	0.2182	0.0217	0.0527	0.0070	0.0194	0.0031	0.3068	0.0104
39	Huyện Bình Chánh	4318	0.2597	0.0224	0.0641	0.0074	0.0234	0.0033	0.3032	0.0095
41	Huyện Nhà Bè	3054	0.3237	0.0219	0.0909	0.0092	0.0365	0.0048	0.3324	0.0106
43	Huyện Cần Giờ	3268	0.4430	0.0257	0.1298	0.0127	0.0528	0.0071	0.3148	0.0106
	All HCM city	**92367**	0.1304	0.0115	0.0306	0.0036	0.0110	0.0016	0.3250	0.0120

Source: Authors' estimation

Figure 4: Estimates of headcount index and standard error at the district level

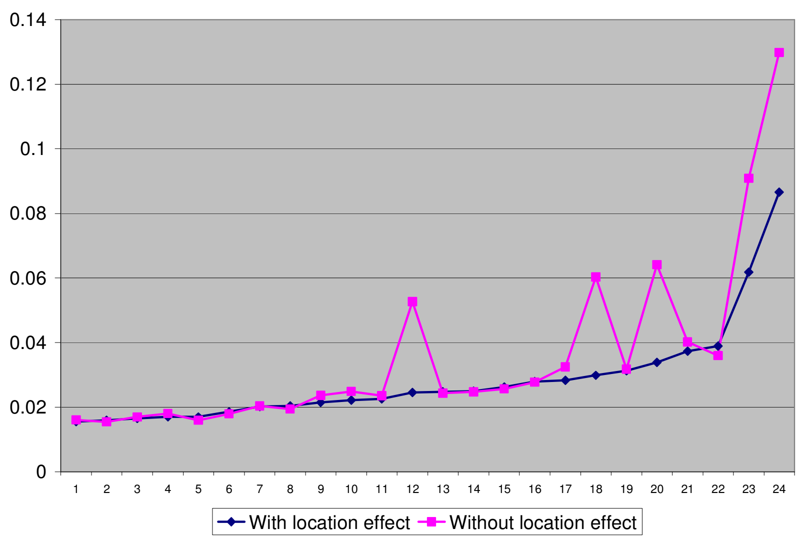


Estimates of poverty headcount index

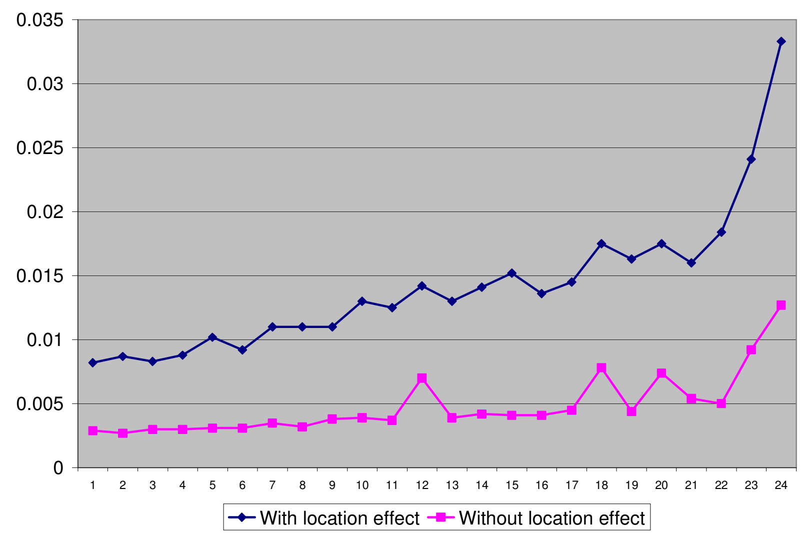


Standard error

Figure 5: Estimates of P1 index and standard error at the district level

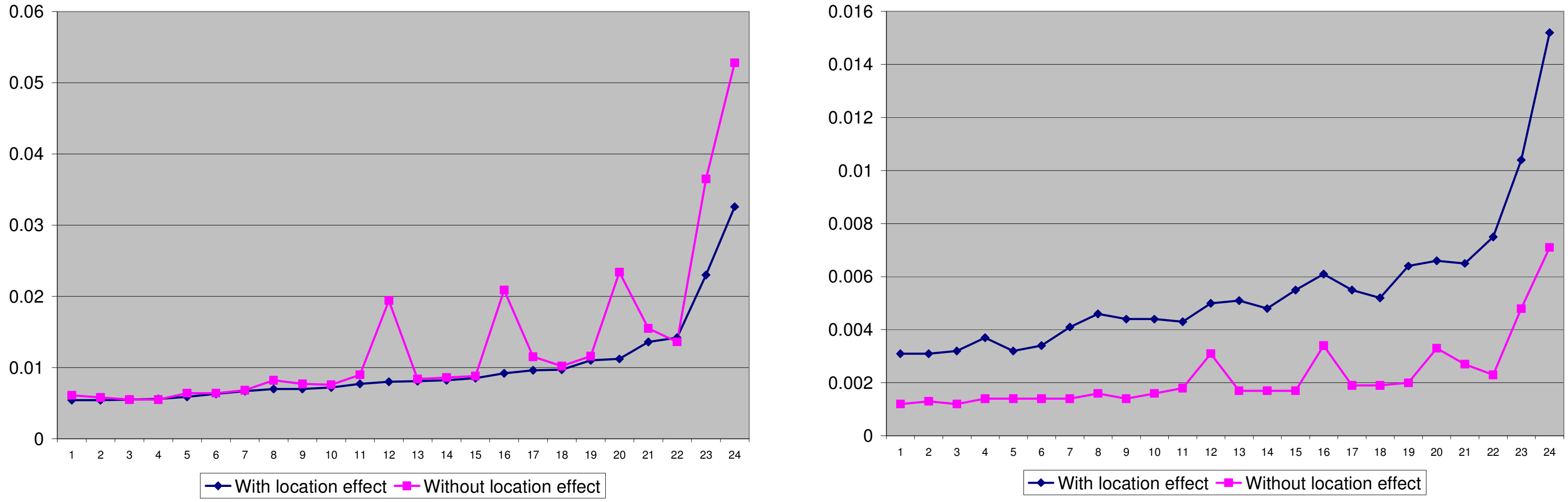


Estimates of poverty depth index



Standard error

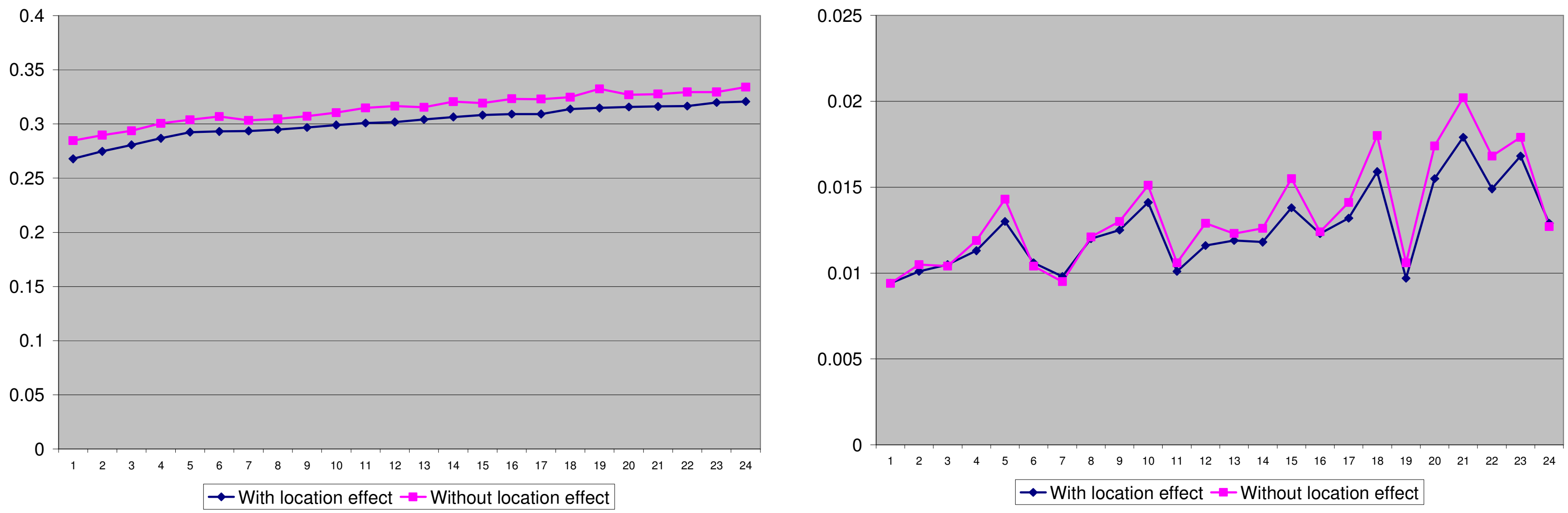
Figure 6: Estimates of P2 index and standard error at the district level



Estimates of poverty severity index

Standard error

Figure 7: Estimates of Gini index and standard error at the district level



Estimates of Gini index

Standard error

One important objective of the study is to examine whether the poverty and inequality estimates are sensitive to location error effect. Table 7 presents the estimates of poverty and inequality indexes under assumption that there is no special correlation between households within a cluster. This assumption is imposed by the previous study of HCM map. The model specification is the same as Model 1 (Table 2).

Figure 4 compares the estimates of poverty incidences in models with and without location error effect. For districts of low poverty incidences, the two models give very close estimates. For districts of higher poverty incidences, the no-location-effect model results in higher estimates of poverty. Regarding to standard errors, as expected, the no-location-effect model results in much lower estimates than the location-effect model. It indicates that the model under no spatial correlation assumption tends to underestimates the standard errors.

Figure 5 and 6 graphs the estimates of P1 and P2 of two models. Again, the model without location effect gives higher estimates of poverty indexes than the model with location effect for some districts. The standard errors are always smaller than in the model without location effect.

However, for Gini estimates, the standard errors estimated from the two models are very close. The estimates of Gini index are still higher in the no-location-effect model than the location-effect model.

Finally Figures in Appendix graph different household characteristics at the district level and compare them with poverty rate.

7. CONCLUSION

The research estimates the poverty rate for the districts in HCM city using method of small area estimation and verifies the assumption on spatial correlation between households within a cluster. The old poverty map of HCM city assumes that there is no spatial correlation. There are two data sources used for this estimation. The first is VHLSS 2004, which is used to run regression of expenditure equation for HCM city. The second is the 10% mid-census sample of HCM city.

It is found that there is spatial correlation between households at the district level, albeit at the low magnitude. Without taking into account this location effect, the standard errors of welfare estimates are underestimated. The model without location effect results in much lower standard errors of estimates of all three poverty indexes and Gini coefficient. Poverty estimates are also a bit different between the location-effect model and no-location-effect model, especially for the poor districts.

Poverty estimates are much higher in suburb districts which have a large proportion of rural area. However, the poverty density is smaller in the poorest districts and higher in the richest districts, since the population density is much lower in the poorest districts than in the richest districts. The standard errors of the poverty estimates are relatively high, which makes the comparison of poverty between districts difficult, especially for districts with poverty rates less than 10%. The Gini estimates at the district level are rather small, around 0.3.

REFERENCES

Elbers, C., Lanjouw, J. and Lanjouw, P., 2003. "Micro-level estimation of poverty and inequality." *Econometrica* 71 (1): 355-364.

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", *World Bank Economic Review*, Vol. 14, No. 1: 147-65

APPENDIX: MAPS OF POVERTY AND HOUSEHOLD CHARACTERISTICS

Figure 4: Poverty, employment, and education of household heads

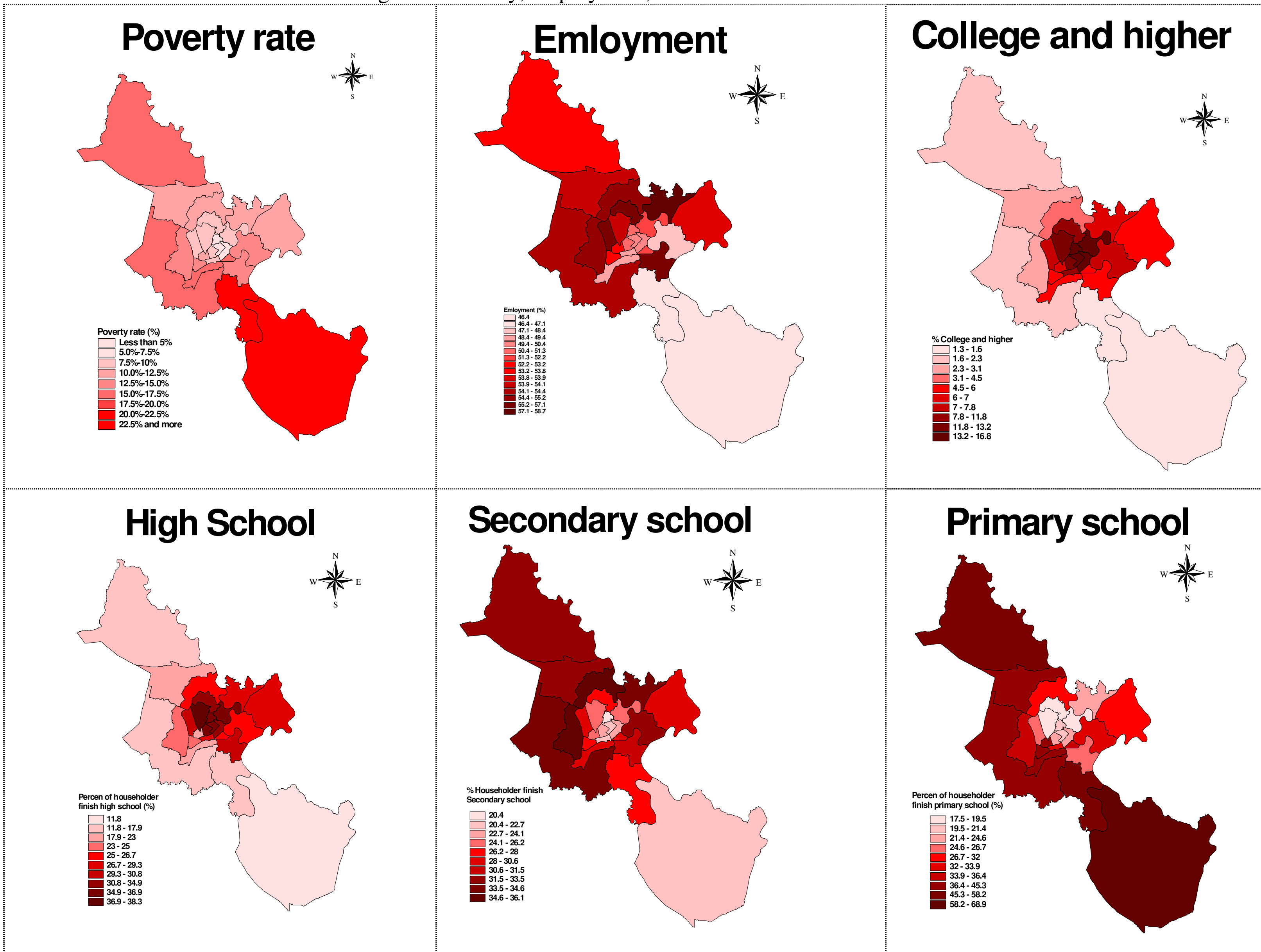


Figure 10: Poverty, housing and computer

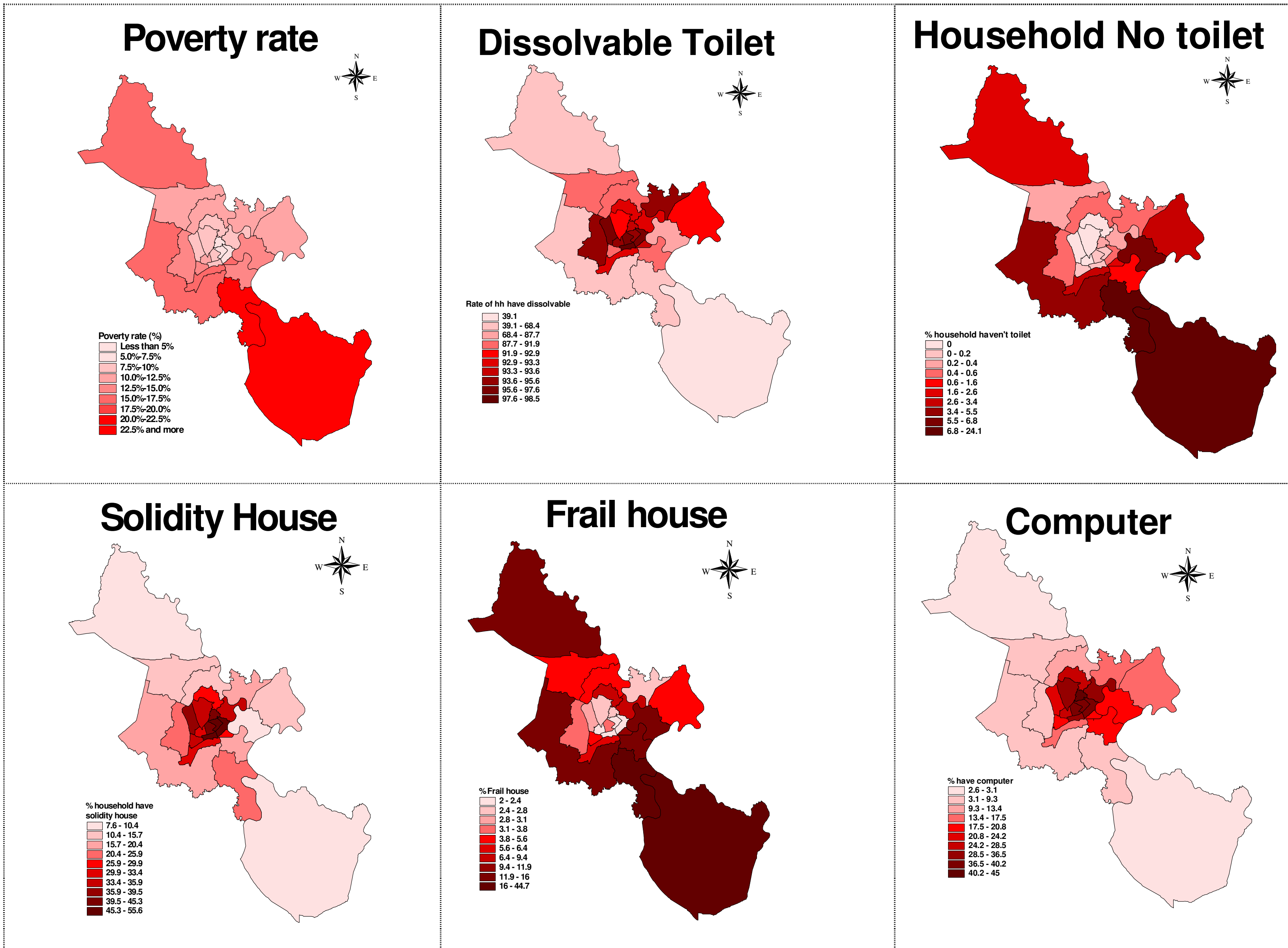


Figure 11: Poverty and durables