SCHOOL OF STATISTICS
UNIVERSITY OF THE PHILIPPINES DILIMAN

WORKING PAPER SERIES

REGIONAL CLUSTERING OF POVERTY IN THE PHILIPPINES

by

Erniel B. Barrios

Ginalyn B. Suratos

UPSS Working Paper No. 2008-06
November 2008

School of Statistics
Ramon Magsaysay Avenue
U.P. Diliman, Quezon City
Telefax: 928-08-81
Email: updstat@yahoo.com

ABSTRACT

Poverty is analyzed in relation to its location in a space-time continuum. The sociological perspective is aptly tallied in the spatial dimension, while the economic perspective is addressed in the temporal dimension. We postulate a sparse spatio-temporal model that reflects the geographic structure of income and related indicators in the Philippines from 1985 to 2000.

The immense improvement in predictive capability should earn SAR a recommendation in the field of spatio-temporal modeling despite its computational complexity. There is spatial clustering among the provinces in the Philippines in terms of poverty distribution. There is also a regional convergence of poverty level indicative of space-specific dependence of the problem. The significance of temporal autocorrelation also suggests that poverty is still a chronic problem in the Philippines.

Keywords: *Spatio-temporal modeling, Sparse Spatial Autoregression, Poverty*

1. INTRODUCTION

Poverty cripples a nation. It manifests itself in different faces— illiteracy, illnesses, homeless children and families, child labor, maltreated laborers, unemployment and underemployment, prostitution, prisoners deprived of justice, and extreme hunger, among others. All these lead them to a state of incapability to govern themselves and hopelessness to be functional citizens, prompting for the Millennium Development Goals (MDG) to consider poverty reduction as its first goal, to halve the poor in the period 1990-2015.

To mitigate the ill effects of this tragedy, poverty reduction is a task that every developing country has been grappling with. Because all countries are linked, poverty in one country introduces a ripple effect on the rest of the world. This quandary gave leverage to government and non-government institutions, researchers, social scientists, and policy makers around the world, along with multilateral agencies like the United Nations and World Bank to assume complementary roles and exert concerted efforts in profiling poverty, determining the factors associated with it, and sorting out policy directions towards the liberation of the poor from the painful grip of poverty.

In the Philippines, poverty has been both widespread and persistent; thus, poverty eradication has also been one of the major agenda of the different administrations. However, given the scarcity of funds, successful implementation of anti-poverty programs requires creative solutions that will allow the government to allocate project funds efficiently. Aside from good governance, it is essential to objectively locate the people of greatest need and distribute the benefits according to need to reduce the project cost of intervention. Concomitantly, successful formulation of poverty alleviation strategies necessitates clear identification of poverty indicators.

This paper aims to understand poverty in the Philippines in a space-time continuum. We will define the poverty indicators and spatial distance measures that will facilitate the study/ understanding of the geographic distribution of poverty. We will also postulate and empirically characterize a spatio-temporal model for some poverty indicators. We will also compare spatial autoregression (SAR), ordinary least squares(OLS), and mixed models in terms of predictive capability and quality of significant variables.

Poverty is analyzed using the spatio-temporal framework. Poverty, just like other socio-economic and environmental systems, must be analyzed and defined in relation to its location in a space-time continuum since phenomena and systems interact with one another across space and over a range of temporal units (Bennet, 1979). The sociological perspective is aptly tallied in the spatial dimension, while the economic perspective is addressed in the temporal dimension. In addition, spatio-temporal analysis of poverty utilizes information more efficiently than cross-section alone or time series analysis alone. This paper also presents an approach to modeling space-time interactions using regression analysis in a sparse spatial autoregressive framework as a directed tool in establishing the spatio-temporal dimensions of poverty in the Philippines. Sparse spatial autoregression approach achieves better prediction or estimation and correct inference by examining only the relations among nearby observations, (Pace and Barry, 1997). Furthermore, this technique leads to an immense reduction in both computer storage space and execution time. The method was applied to the poverty data in the Philippines.

Results of the spatio-temporal analysis of poverty can be useful for a number of reasons: First, the information can help determine and assess regional/provincial disparity in terms of living standards and can help shed light on the possible geographic factors associated with poverty. Second, it will enable the government to effectively direct interventions to the most destitute Filipinos in the country. Such information will permit a more rational area-based approach in targeting the people of greatest need of assistance. Third, (corollary to the second point) using this type of analysis, the poorest in the Philippines can be objectively mapped out, focusing on those who needed most assistance, thus reducing project cost of intervention. Fourth, the geographic distribution will also help in identifying patterns in monitoring impact of interventions to establish progress in poverty status.

2. METHODOLOGY

This paper does not attempt to contribute to the debate on how to define poverty indicators, but rather use the direct indicators of poverty including total household income and total household expenditure. These were averaged for the households in the provinces (spatial unit). Income is the most commonly used indicator since it easily translates into households` purchasing capability. Access to basic necessities that enhances living conditions is facilitated proportionally to the income the households earn. Expenditure will augment the income data since it is an evidence of the actual spending of the household. Expenditure alone may not suffice since it does not reflect the allocative efficiency of the households. Higher expenditure does not necessarily mean it was devoted to amenities that improve the living conditions.

2.1 Determinants of Poverty

Demographic profile, economic profile, living conditions and assets, and geographic attributes are postulated as indirect determinants of poverty. Inclusion of the variables and importance of each are justified by the results of previous researches on poverty.

The demographic profile considered are: household size, proportion of households with elderly members, proportion of households with children between 1 year and 6 years, proportion of male-headed households, proportion of households headed by < 40 years old, proportion of households headed by married persons, and highest level of education of household head.

Large households are strongly associated with lower per capita income-- hence the inclusion of household size. Households with a large number of elderly members and children are likely to have low per capita income. With so many children, poor families can hardly invest in education. Households headed by younger individuals, holding other variables constant, tend to be poorer than those headed by older persons. On another level, higher level of education of the household head can be translated to higher earnings thereby increasing the levels of household welfare.

The economic profile includes: proportion of households with employed head, proportion of households where wife is employed, average number of employed persons in the household, and proportion of households engaged in agriculture will be included. Clearly, the number of employed household members can be equated to increase in total household income. Income of an employed wife in addition to that of the husband will also somehow increase the welfare of the household relative to the household who solely depend on the income of the father. Furthermore, households engaged in high-technology manufacturing and services in the cities are clearly very much better off than households hooked up in economic activity concentrated in low-technology agriculture.

The living conditions and possession of assets indicators includes: proportion of single detached houses, proportion of houses with strong walls and strong roof, proportion of households using water-sealed toilet, proportion of households with electricity, proportion of households with safe water, proportion of households with television, and proportion of households with refrigerator. Undoubtedly, economic status is associated with housing characteristics and possession of assets.

The geographic dependence is defined in terms of neighborhoods represented by membership to a region for simplicity. Other measures of geographic dependence can be defined but was not pursued in this paper.

2.2. The Data

The Philippine Family Income and Expenditure Survey (FIES) data for six time points (1985, 1988, 1991, 1994, 1997, and 2000) were aggregated at the provincial level, for a total of 426 observations from 71 provinces.

2.3. Spatio-temporal models

The study used a simultaneous autoregressive model that is augmented to include the concept of sparse spatial autoregressions. This is the basis in the formulation of spatial association as a component in the model that characterizes the poverty indicators. The foundation for such model is that when regression is applied to spatially distributed data, predictive capabilities can be lost by ignoring the presence of spatial autocorrelation. Ignoring spatial autocorrelation leads to violation of the assumptions underlying ordinary least squares regression (i.e. independence of observations), thereby resulting in erroneous statistical inference. Simultaneous autoregression, conditional autoregressions and kriging can adapt to this problem. These techniques, however, involve examining the explicit relation among observations, which may require order of n^3 operations ($0(n^3)$) (Cressie, 1993) as cited in (Pace and Barry, 1997). Piecewise inclusion of various dependencies coming from spatial and temporal relations is more tedious than simultaneous treatment of the problems.

Sparse spatial autoregression will not cover unlimited lags of autocorrelations, but the technique may truncate the influence of zero elements thus, reducing the number of relations needed to estimate the spatial regression. Only the relations among nearby observations matter greatly and represented in sparse matrices. The technique avoids performing unnecessary computations on the zero elements, hence the dramatic reduction of storage space and acceleration of execution time (Cressie, 1993).

2.4. Simultaneous Autoregressions (SAR)

(Pace and Barry, 1997) pointed out that when errors exhibit spatial autocorrelations, spatial estimators like simultaneous autoregressions (SAR) can adjust to the problem by correcting the prediction of Y through a weighted average of the preliminary residuals $|Y-X\hat{\beta}|$ on nearby observations in, $Y = X\beta + \alpha D(Y - X\beta) + \varepsilon$ where Y denotes the *n x 1* vector of observations of the dependent variables (total household income, total household expenditure), X stands for the *n x k* matrix of observations on the independent variables of interest, β indicates the *k x1* vector of parameters, and D is the *n x n* spatio-temporal weight matrix with 0's on the diagonal and non-negative off-diagonals. We assume here that each row in D sums to 1. A non-zero entry in the *j*th column of the *i*th row in D (d_{ij}) indicates that the *j*th observation will be used to adjust the prediction of the *i*th observation (i≠j). The value of d_{ij} depends on the distance between units *i* and *j*. After correcting for this interaction, SAR assumes that the residuals are independently and normally distributed. The ε in the foregoing equation represents an *n x 1* vector of normal iid errors.

Assuming the existence of the maximum likelihood estimate, one could predict Y through

$\hat{Y} = X\hat{\beta} + \hat{\alpha} D (Y-X\hat{\beta})$. This leads to the estimated errors .

$\hat{\varepsilon} = Y - \hat{Y} = Y - X\hat{\beta} - \hat{\alpha} D(Y - X\hat{\beta}) = (I - \hat{\alpha} D)(Y - X\hat{\beta})$ The main barrier to the speedy computation of the estimates lies in the *n x n* nature of the spatial weight matrix, *D* (which will be discussed in the next section). Computing $|I - \alpha D|$, a determinant of an *n x n* matrix, requires substantial time for large *n*. Thus, storage requirements increases with the square of *n* while the operation count rises with the cube of *n*. Sparse spatial weight matrix eliminates this immense computational task. The spatial weight matrix *D* contains mainly o`s, thus sparse. It uses the *m* nearest neighbors out of a possible *n* neighbor for each of the n observations. Thus, *mn* non-zero elements exist out of possible n^2 elements. To measure sparsity, we examine the number of non-zero elements of a matrix relative to the total number of elements. Hence, *m* nearest neighbors weighting matrix *D* has proportionally *m/n* non-zero elements and this becomes progressively more sparse for larger empirical applications. . Significance of the parameters can be tested using the likelihood ratio test.

2.5. Spatial weight matrix

For simplicity, let us define proximity of provinces as membership to a region. Provinces from the same regions are statistically postulated as neighbors. To illustrate the weighted averaging of residual, let weight_{ij} represent the distance between a pair of observations. Let $\text{weight}_{ij} = 1$ if a province belongs to a region and zero otherwise. We normalize the initial weights so that $\sum_{j=1, i \neq j}^{n} Dij = 1$, thus making it into a standardized weight matrix

$$D_{ij} = \frac{w_{ij}}{\sum_{j=1, i \neq j}^{n} w_{ij}} \; .$$

We assume here that each row of the weight matrix sums to 1.

For further structure, the observations were stacked one on top of the other according to time with the first set of observations corresponding to the earliest observations (i.e. 1985). This type of time ordering of *D* immensely simplifies the matrix multiplication, equation solutions, and determinant computations necessary to estimate the model (Pace and Barry, 1997). Dummy variables were created for the six time points.

Manila is left by its lonesome since we excluded the other districts in NCR that are not common to all time points. SAR assumes that all observations belong to one "neighborhood" hence we included Manila as neighbor of Regions 3 and 4. The consumer behavior in NCR, Regions 3 and 4 are assumed to be similar. In fact, some market research studies would collectively call these as the Greater Metro Manila (GMM) area.

As an illustration, Pangasinan, Ilocos Sur, Ilocos Norte, and La Union are adjacent provinces, but Pangasinan is not contiguous to Manila. Pangasinan`s row has the following weights: $D_{1,2:426}$ = [0, 0. 333, 0.333, 0.333, 0, 0,..0]. Note that the first entry is the intersection between Pangasinan and Pangasinan (in the matrix) is zero to prevent it from predicting itself. The second, third and fourth entries are the intersections between Pangasinan and Ilocos Sur; between Pangasinan and Ilocos Norte; and between Pangasinan and La Union, which are equal to 0.333 since they are considered neighbors. The fifth entry which is the intersection between Pangasinan and Manila equals zero since they are not considered neighbors. The row sums to 1.

3. RESULTS AND DISCUSSION

Using the FIES data for six time points and aggregating household-level information at the provincial level, ordinary least squares estimation (OLS), SAR, and the Mixed Model were fitted using the R language and SAS.

3.1. Assessment of Model Performance

The mean absolute percentage error (MAPE) measures the accuracy of prediction or performance of the model in forecasting. A lower MAPE suggests better prediction of the poverty indicator.

With income as dependent variable, SAR (MAPE=15.49) outperforms OLS (MAPE=16.20) and mixed models (MAPE=15.84) slightly. Though the differences are small, any improvement in MAPE for variables that are complicated to model are worthy. The total expenditure model shows that SAR (MAPE=14.24) has lower MAPE compared to OLS (MAPE=14.85) and mixed model (MAPE=14.47). This indicates that the correction for spatial dependence improved slightly the predictive capability of the model. Nonetheless, this is not too bad a predictive capability since other studies have achieved similar amount of reduction in prediction error. (Carter and Haloupek, 2000) obtained a 4% reduction in MAPE in his study on commercial rent when going from OLS to SAR. In one of the examples of (Pace and Gilley, 1998), there was only a negligible difference between OLS and SAR in terms of sample prediction. However, after conducting a cross-validation by deleting each observation and its association comparables from the sample, the results show that SAR dominates OLS's prediction by 40%.

To validate the results, out-of sample prediction through cross-validation was done for OLS. Half of the provinces across the regions were excluded. The parameters were estimated, Y of the left out provinces is predicted from $\hat{Y} = X\hat{\beta}$.

For the household income model, the computed MAPE is 20.31; while for the household expenditure model, the computed MAPE is 18.12.

Out-of sample prediction through cross-validation was also done for SAR. Likewise, half of the provinces across the regions were excluded. The parameters were estimated, Y of the left out provinces is predicted from $\hat{Y} = X\hat{\beta} + \hat{\alpha}\left(Y - X\hat{\beta}\right)$.

Still, the results reveal a lower MAPE compared to OLS. For the household income model, the derived MAPE is 17.21; while for the household expenditure model, the computed MAPE is 17.48. SAR indeed produces a better goodness- of- fit compared to OLS.

3.2. Temporal and Spatial Effects

Temporal autocorrelation is the association of between the values of the same variables at different temporal units. Time dummies and year as random component are used to represent temporal effects in OLS and mixed models respectively. In mixed model, it reveals that the variation in household income is explained primarily by annual variation (70%), while 7% is explained by regional fluctuations. On the other hand, autocorrelation function is obtained to account for temporal autocorrelation in SAR. First order autocorrelation is computed for each of the provinces. The average autocorrelation for total household income and expenditure are estimated at 0.46 and 0.47 respectively.

On the other hand, to account for spatial effects in the data, we tested the existence of spatial autocorrelation. (Bailey and Gatrel, 1995) defined spatial autocorrelation as the association between the values of the same variables at different spatial locations. If this assumption is violated the generated coefficients in OLS are no longer efficient and the standard errors are biased. SAR adjusts to this problem and can estimate more efficient and correct standard error (Carter and Haloupek, 2000).

The spatial effects in OLS and mixed model are represented by the region dummy variables and region as random component respectively. A fundamental characteristic distinguishing SAR from OLS and mixed models is the spatial weight matrix. The spatial linkages of the provinces are measured by defining a spatial weight matrix with the dimension *nxn*. There is no standard specification of the elements of a spatial weight matrix. For simplicity, the spatial weight matrix in this study is defined by contiguity matrix, that is, it is based on the membership of a province to a region.

There is spatio-temporal clustering of provinces in the Philippines in terms of poverty distribution as indicated by a significant spatial autocorrelation for both income (α = 0.37709, $p<0.000$) and expenditure (α = 0.38144, $p<0.000$). Intuitively, this makes sense since socio-cultural characteristic shared among contiguous communities link them together spatially.

The significance of the spatial autocorrelation found in the analysis conforms to the findings of (Carter and Haloupek, 2000) and (Pace and Gilley, 1998) that OLS estimates are inconsistent and biased.

3. 3. Determinants of Poverty in a Spatio-Temporal Context

We include this criterion for comparison to determine which method produces a set of predictors that are supported by the literature. In table 1, with TOTAL HOUSEHOLD

INCOME as the dependent variable, all significant variables at the 0.05 significance level in the sparse spatial autoregression model remained significant in the OLS and mixed models except for the variables WALLS and ELECTRIC. WALLS is not significant in SAR but significant in both OLS and mixed model. ELECTRIC is significant in both SAR and mixed models but not significant in OLS.

[Table 1 Here]

OLS produces significant estimates of Regions 1, 5, 7, 8,10,11,12, and 13 dummy variables while mixed model produces significant estimates of Regions 3, 10,12,13,14, and 16 at the .05 significance level. OLS and Mixed model do not agree in terms of the direction of effects in the following: Regions 1, 2, 6, 8, and 9. They have the same direction of effects in other regional dummy variables. The agree in terms of significance (at the .05 significance level) and direction of estimates only in Regions 10,12, and 13, though still they vary in the magnitude of estimates. This is to be expected since the spatial autocorrelation is significant hence OLS estimates become inefficient.

With regard to time dummy variables, SAR and OLS produce similar estimates of all dummy variables in terms of significance, direction and magnitude except for dum88. DUM88 is not significant in SAR but significant in OLS. Mixed model on the other hand, produces quite different set of significant time dummy variables: DUM94 and DUM97 fell outside of the .05 significance level. Also, DUM85 and DUM88 are negative estimates in mixed model.

In table 2, with TOTAL HOUSEHOLD EXPENDITURE as the dependent variable, all significant variables at the 0.05 significance level have the same signs across the three models. They have the same set of significant variables except for EMPLOYED, TV**,** and DUM88**.** The variable EMPLOYED is significant in OLS but not in SAR and mixed; TV is significant in SAR but not in OLS though marginally significant in mixed models; DUM88 is likewise significant in SAR and mixed models but not in OLS.

[Table 2 Here]

Also note the treatment by OLS and mixed model of the region and time dummies. Just like in the previous model (household income) OLS produces significant estimates of Regions 1, 5, 7, 8,10,11,12, and 13 while mixed model produces significant estimates of Regions 3, 6, 10,12,13,14, and 16 at the .05 significance level. OLS and mixed model do not agree in terms of the direction of effects in the following dummies: Regions 2, 5, 6, and 9. They have the same direction of effects in the rest. The agree in terms of significance (at the .05 significance level) and direction of estimates only in Regions 10,12, and 13, though still they vary in the magnitude of estimates. Again, this is to be expected since the spatial autocorrelation is significant hence OLS estimates become biased and inefficient.

With regard to time dummies, SAR and OLS produce similar estimates of all dummies in terms of significance, direction and magnitude except for dum88. Dum88 is not significant in SAR but significant in OLS. Mixed model on the other hand produces quite different set of significant time dummies: dum94 and 94 fell outside of the .05 significance level. Also, dum85 and dum88 are negative estimates in mixed model.

Noticeably, the three models produced different sets of significant predictors. Though there was only a slight difference in the sample prediction among the three models, their estimates vary. Given that SAR has lower MAPE, its results are still more convincing than those of the other two models.

The parameter estimates in the SAR model are discussed. The variation in household income is explained by family size, proportion of employed family members, and proportion of households with electricity, refrigerator, and have a water-sealed toilet. With every addition of a household member (increasing chance of employment), ensures an increase in household income by PhP17,651.00. Economic status is clearly attributable to housing characteristics such as proportion of households with electricity, refrigerator and water-sealed toilet. Thus, improving accessibility to electricity and water-sealed toilet leads to improvements in household living conditions.

Total household expenditure is attributed to time dummy variables, except for DUM88**,** FAMILY SIZE, and proportion of households with TV, REF, and ELECTRICITY. This is reasonably consistent with the results in the model with total income as the dependent variable since expenditure is directly correlated with income. Similar interpretation applies here as well. Poverty status of provinces is clearly related to housing characteristics like proportion of households with TV, REF and ELECTRICITY.

The foregoing results go with a caveat that the absence of the other selected variables that fell outside the .05 significance level does not imply no effect at all. Variables that turned out significant are those whose effects can be statistically demonstrated.

3.4. Spatial Effects

The construction of spatial weight matrix in SAR sets the venue at which neighborhood effects make themselves felt. α represents the spatial pattern observed in the spatial weight matrix. It denotes the strength of the interaction between and among contiguous locations. The high significance level of the spatial autocorrelation for both models (total household income (α = 0.37709, p value= 2.331e-09) and total household expenditure (α = 0.38144, p value= 8.9392e-10) offers an explanation why the same economic situation happens nearby. This re-affirms that there is clustering among the provinces in the Philippines in terms of poverty distribution. Intuitively, this makes sense since sharing of socio-cultural situation links contiguous communities. This result is complementary to the findings of Barrios and Landagan (2004) in their study on poverty.

4. CONCLUDING NOTES

There is enough reason to conclude that spatial clustering among the provinces in the Philippines in terms of poverty distribution exists, and poverty is contaminating. The poverty incidence in one province is related to the poverty incidence in other contiguous provinces. This is explained by the interdependence among the economic and socio-cultural factors among neighboring provinces. The kind of industries present in an area, the physical accessibility of a community, the kind of welfare received by the stakeholders, the productivity of land, weather conditions, cultural beliefs, and many other factors can explain the spatial interdependence of poverty status. Also, the intervention menu applied in one province may be similar to those applied in the neighboring provinces.

SAR ably represented poverty in the Philippines in the space-time continuum by examining only the relations among nearby observations. SAR outperforms OLS and mixed model in terms of predictive capability, though not as much reduction in MAPE as in similar studies. MAPE reduction might be highlighted in other variables that are simpler to model.

The improvement in predictive capability, dimension reduction, spatio-temporal effects, and the quality of predictors relative to OLS and mixed models should earn SAR a recommendation in spatio-temporal modeling. There is a trade-off between computational complexity and other benefits such as lower MAPE, location specific indices, immense reduction in both computer storage space and execution time.

Geographic targeting is more effective when smaller geographic units, such as municipalities or districts are targeted.

Results of this study are certainly not sufficient and exhaustive to back up a national poverty reduction strategy. Nonetheless, this paper, at least, supports the findings of other authors of poverty studies. We should, thus, treat the SAR results from a panoramic point of view as indicative of meaningful trends and patterns, rather than mere statistical figures. The more important thing is how to use these empirical findings as supplementary information in an effort to relieve the poor from the painful grip of poverty. Implementing effective anti-poverty programs based on a valid model requires a strong and not necessarily explicit commitment to represent the interests of the poor.

R E F E R E N C E S

Barrios, E. and Landagan, O., 2004, Geographic Distribution of the Poor: Is Poverty Contaminating? Proceedings of the 9th Convention of Statistics, Manila.

Bennet, 1979, Spatial Time Series: Analysis-Forecasting- Control. Great Britain: Page Bros (Norwich) Limited.

Carter, C and Haloupek, W., 2000, Predicting Commercial Rent Using Spatial Autoregression Techniques, Paper presented at the 12th Annual PAPCAP/FMA Finance Conference and Pacific Basin Financial Management Society Annual Meeting, July 6-8, 2000. Melbourne, Australia.

Cressie, N., 1993, Statistics for Spatial Data, 2nd ed., New York:John Wiley and Sons.

Montgomery, D., 2001, Design and Analysis of Experiments, New York:John Wiley and Sons.

Pace, R. and Barry, R., 1997, Sparse Spatial Autoregressions**.** Statistics & Probability Letters **33:**291-287.

Pace, K. and Gilley, O., 1998, Generalizing the OLS and Grid Estimators. Available online at http://www.finance.lsu.edu/academics/finance/re/pace=manuscripts/sarols/htm.

Table 1. SAR, OLS, and Mixed model estimates for Total Household Income

Total Income	β_{SAR}	p-value	β_{OLS}	p-value	β_{MIXED}	p-value
Intercept	13598.44	.63581	-18852	.5476	7453.25	.8335
Fsize	8462.15	.01248	9326.90	.0100	9310.71	.0093
Age7	-3810.74	.65907	260.08	.9774	-258.27	.9775
Employed	12478.17	.05368	17651	.0109	17050	.0126
Sex	-116.29	.70015	181.68	.5684	136.15	.6673
MS	-344.05	.26989	-385.93	.2543	-422.05	.2070
HHHjob	49.84	.53070	-5.22	.9489	4.6	.9547
Occup	-134.34	.12566	-165.04	.083	-157.52	.0933
Hhtype	-205.59	.08541	-130.52	.3184	-143.43	2604
Wife_emp	63.065	.64828	-6.12	.9671	-4.61	.9748
Bld_type	-229.64	.16467	-146.45	.4071	-153.26	.3812
Roof	22.68	.81113	-172.84	.1330	-111.21	.3125
Walls	164.82	.08887	259.20	.0196	228.41	0.0344
Electric	-183.30	.02944	-169.05	.0594	-180.19	.0434
Water	92.64	.17005	44.03	.5500	58.02	.4256
TV	131.78	.25623	-99.47	.4573	-6194	.6362
Ref	915.09	9.71e-08	1189.08	<.0001	1154.93	<.0001
Toilet	104.43	.04943	283.27	<.0001	235.08	.0002
Region 1			-10409	0.0147	256.14	0.9529
Region2			-6172.98	0.1524	4927.77	0.2264
Region3			Base	Base	9894.26	0.0028
Region5			-14249	0.0002	-2596.92	.4874
Region 6			-6161.41	0.0974	3938.57	0.3215
Region7			-17054	<.0001	-5506.60	0.1628

Region8			-11012	0.0053	34.42	0.9928
Region9			-7903.45	0.0882	2727.40	0.5178
Region 10			-21146	<.0001	-9168.31	0.0179
Region 11			-16052	0.0001	-4269.18	0.2762
Region 12			-21928	<.0001	-9380.79	0.0260
Region 13			-30898	<.0001	-16295	0.0002
Region 14			7511.99	0.1386	15005	0.0010
Regn16			2727.47	0.5676	10433	0.0202
Dum85			Base	base	-30203	0.0060
Dum88	5319.36	.20668	6839.17	.0211	-23613	0.0305
Dum91	22569.94	1.407e11	23532	<.0001	-7009.53	0.5178
Dum94	28901.49	4.741e11	30425	<.0001	160.75	0.9882
Dum97	55280.91	<2.2e-16	57027	<.0001	26267	0.0161
Dum00	63132.39	<2.2e-16	65803	<.0001	34719	0.0016
MAPE	15.49		16.20		15.84	
α	.37709	2.331e-09				

Table 2. SAR, OLS, and Mixed model estimates for Total Household Expenditure

Total Expenditures	β_{SAR}	p-value	β_{OLS}	p-value	β_{MIXED}	p-value
Intercept	10066.1	.645623	-15879	.5098	6482.47	.8121
Fsize	7068.94	.006169	7543.04	.0067	7498.29	.0064
Age7	-1427.3	.82240	652.77	.9264	525.75	.9405
Employed	7520.95	.127092	10520	.0479	10166	.0521
Sex	-182.11	.428655	64.94	.7907	14.63	.9520
MS	-190.171	.423545	-232.24	.3718	-250.09	.3296
HHHjob	31.92	.598244	16.23	.7955	21.59	.7284
Occup	-54.71	.413037	-68.24	.3499	-66.59	.3545
Hhtype	-97.10	.286363	-67.17	.5039	-71.98	.4676
Wife_emp	107.78	.306336	63.40	.5782	62.53	.5766
Bld_type	-200.80	.110793	-145.03	.2855	-147.91	.2710
Roof	17.95	.804172	-76.56	.3860	-42.18	.6162

Walls	86.20	.243163	129.90	.1274	113.80	.1680
Electric	-123.52	.054080	-126.12	.0672	-131.87	.0542
Water	58.66	.254209	23.15	.6825	33.56	.5480
TV	174.187	.049064	11.44	.9114	39.52	.6939
Ref	685.41	1.601e07	894.34	<.0001	866.14	<.00.1
Toilet	78.57	.052615	217.96	<.0001	176.57	.0002
Region 1			-9774.84	0.0029	-1348.16	0.6769
Region2			-5591.94	0.0917	3051.76	0.3151
Region3			base	Base	7408.80	0.0025
Region5			-7654.23	0.0093	694.90	0.8028
Region 6			-1298.46	0.6488	5840.92	0.0491
Region7			-13487	<.0001	-4650.56	0.1140
Region8			-9741.65	0.0014	-1281.06	0.6526
Region9			-6645.92	0.0622	1354.19	0.6671
Region 10			-17306	<.0001	-7987.77	0.0057
Region 11			-13309	<.0001	-4137.85	0.1571
Region 12			-16801	<.0001	-7075.82	0.0245
Region 13			-19830	<.0001	-9134.68	0.0047
Region 14			4952.23	0.2037	10441	0.0022
Regn16			914.62	0.8030	6824.53	0.0419
Dum85			base	Base	-23308	0.0071
Dum88	4085.8	.205815	4652.49	.0410	-18775	0.0288
Dum91	17246.5	1.534e07	17428	<.0001	-6053.91	0.4778
Dum94	23177.6	5.767e12	23691	<.0001	193.90	0.9819
Dum97	42871.4	<2.2e16	43364	<.0001	19777	0.0213
Dum00	50961.3	<2.2e-16	51959	<.0001	28167	0.0011
MAPE	14.24		14.86		14.47	
α	0.38144	8.932e-10				

Table 3. Covariance parameter estimate for household income

Covariance parameter	Estimate
YEAR	6.8596E8
REGION	85907508
RESIDUAL	2.1677E8
% variability explained by regn	9%
% variability explained by year	70%

Table 4. Covariance parameter estimate for expenditure

Covariance parameter	Estimate
YEAR	4.2535E8
REGION	44575007
RESIDUAL	1.2798E8
% variability explained by regn	7%
% variability explained by year	71%