Presented at the DLSU Research Congress 2014
De La Salle University, Manila, Philippines
March 6-8, 2014

Zero-Inflated Modelling of Overdispersed Food Shortage Data

Ruth Sarno, Justin Wong and Shirlee Ocampo
De La Salle University, Manila
shirlee.ocampo@dlsu.edu.ph

Abstract: Food shortage and food poverty are major problems of many developing countries which need immediate actions and effective solutions. Food poverty statistics and food shortage models must be generated to implement essential programs to the places where hunger incidence is high. This research focused in Marinduque, one of the poorest provinces in the Philippines, used two hunger indicators, namely: (1) a household was classified as food poor or "hungry" if it has a per capita income (PCI) below the food threshold as set by the National Statistical Coordination Board (NSCB), and (2) whether or not the household had experienced food shortage in the past three months. McNemar test for paired populations showed that there is a significant disagreement between the two hunger status criteria. Multiple linear, Poisson and negative binomial regression models were fitted. Under the PCI criterion, goodness of fit tests showed that multiple linear regression model (MLRM) is the most preferred. The MLRM models suggests households with cell phone, computer, electric fan, and engaged in forestry, mining, crop farming/gardening, fishing, and construction have significant effects to food poverty incidences. However, model fitting of food shortage data encountered difficulties due to excessive zeroes and overdispersion. Hence, zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) models were applied. Under the food shortage criterion, goodness of fit tests showed that ZINB model is the most preferred. The ZINB model suggests that households who received cure or treatment from sickness, with expected deceased, malnourished, and single parent member/s, households with refrigerator, radio and telephone, engaged in transportation and manufacturing, households with roofs made from makeshift materials, with more than 250 meters distance from the water source, and without sanitary toilet/s, and household size have significant effects to food shortage incidences across all barangays in Marinduque.

Key Words: food shortage and poverty; zero-inflated Poisson; zero-inflatednegative binomial; overdispersion

1. INTRODUCTION

Food Poverty is the incapacity of a household to eat or consume adequate, healthy and nutritious meal on a daily basis. The inability of having to eat healthy and nutritious food leads to malnutrition and other health and social concerns of every member of the households (Friel & Conlon, 2004).

Last October 2010, President Benigno Aquino III instructed his cabinet members to make a 5 year plan against poverty. This can be linked to the existing law, Republic Act (R.A.) 8425 or the Social Reform and Poverty Alleviation Act, which President Fidel V. Ramos signed last December 1997 which only took effect during the presidency of Joseph Estrada. The main objective of this law is to eradicate poverty. Under this law, every individual and household should be empowered to be able to meet its minimum basic needs of health, food, income security, shelter, peace and order, education, family care, and the like. The Barangay – Level Total Development and Protection of Children Act (R.A. 6972) was also implemented in order to provide feeding programs and other basic needs such as protection and development to children under 6 years old in every barangay.

Food Poverty is one of the global problems that need urgent and effective solutions. The first Millennium Development Goal is to eradicate poverty and hunger. In 2000, members of the United Nations vouched to participate in achieving the Millennium Development Goals (MDG) by the year 2015. By achieving these goals, poverty and other human deprivations will be reduced. The Philippines, as a member of the United Nations, is part of this project. It is expected that by 2015, there will be a decrease in the percentage of poverty and human deprivation in the Philippines. In a recent report published by Asian Development Bank, the Philippines is expected to miss and unlikely to achieve all the MDG. To be able to eradicate poverty and hunger, the poverty incidence of the Philippines should decrease to about 18% before 2015. The Philippines has the "poorest capability" in achieving the MDG (Remo, 2013).

One of the poorest provinces in the Philippines is Marinduque. The 2008 Community Based Monitoring System (CBMS) poverty maps showed 24,023 out of 49,919 households (48.1%) in this province have PCI below the poverty threshold and are considered poor. Also, 16,908 out of 49,919 households (33.9%) have PCI below food poverty threshold and are classified as food poor. Furthermore, 2,536 out of 49,919 households (5.1%) have suffered food shortage in the past 3 months (PEP-CBMS Network Office, 2011).

This research focused on Marinduque's hunger incidences based on two criteria (1) food poverty/PCI and (2) food shortage. Multiple linear, Poisson, negative binomial (NB), zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) regression models were fitted and compared. Moreover, estimates of food poverty and food shortage incidences were generated using these models and compared to the actual incidences.

2. DATA AND METHODOLOGY

The data used is the 2008 CBMS complete enumeration for the province of Marinduque. The province of Marinduque is composed of 6 municipalities with 218 barangays. It has a total of 49,919 households (HH) with a population of 210,114. Table 1 shows all the variables considered in the study and their descriptions.

Table 1. Variables and Descriptions

Dependent Variables	
Variable	**Description**
FPOOR	Food poor indicator (Yes = 1 ; No = 0)
FSHORT	Food shortage indicator (Yes = 1 ; No = 0)
Independent Variables	
Variable	**Description**
UPAR	Single parent indicator (Yes = 1 ; No = 0)
DSBLE	Disability indicator (Yes = 1 ; No = 0)
CURE	Received cure for sickness (Yes = 1 ; No = 0)
PREVM	Deceased member (Yes = 1 ; No = 0)
EXPM	Expected family member/s indicator (Yes = 1 ; No = 0)
WELEC	Electricity indicator (Yes = 1 ; No = 0)
RADIO	Own radio (Yes = 1 ; No = 0)
TV	Own television (Yes = 1 ; No = 0)
VPLYR	Own video player (Yes = 1 ; No = 0)
Independent Variables	
Variable	**Description**
STEREO	Own stereo (Yes = 1 ; No = 0)
KRAOKE	Own karaoke (Yes = 1 ; No = 0)
REF	Own refrigerator (Yes = 1 ; No = 0)
EFAN	Own electric fan (Yes = 1 ; No = 0)
IRON	Own iron (Yes = 1 ; No = 0)
STOVE	Own stove (Yes = 1 ; No = 0)
WMCH	Own washing machine (Yes = 1 ; No = 0)
MICRW	Own microwave (Yes = 1 ; No = 0)
COMP	Own computer (Yes = 1 ; No = 0)
CFONE	Own cellular phone (Yes = 1 ; No = 0)
TFONE	Own telephone (Yes = 1 ; No = 0)
AIRC	Own air conditioner (Yes = 1 ; No = 0)
SMCH	Own sewing machine (Yes = 1 ; No = 0)
CAR	Own car (Yes = 1 ; No = 0)

CROP	Engaged in crop farming/gardening (Yes = 1 ; No = 0)
POULT	Engaged in livestock/poultry (Yes =1 ; No = 0)
FISH	Engaged in fishing (Yes = 1 ; No- 0)
FOR	Engaged in forestry (Yes = 1; No = 0)
SAL	Engaged in wholesale/retail (Yes = 1; No = 0)
MAN	Engaged in manufacturing (Yes = 1 ; No = 0)
SERV	Engaged in services (Yes = 1 ; No = 0)
TRN	Engaged in transportation (Yes = 1 ; No = 0)
MIN	Engaged in mining/quarrying (Yes = 1 ; No = 0)
CNS	Engaged in construction (Yes = 1 ; No = 0)
EOTH	Engaged in other activities NEC (Yes = 1 ; No = 0)
OFW	OFW Indicator (Yes = 1 ; No = 0)
CARP	CARP program Indicator (Yes = 1 ; No = 0)
PHEALTH	PHILHEALTH Indicator (Yes = 1 ; No = 0)
NBUS	Engaged in business (Yes = 1 ; No = 0)
MAL05	Malnourished (0-5 years old) children Indicator (Yes = 1 ; No = 0)
JOB	Job/work Indicator (Yes = 1 ; No = 0)
MSCH	Attending School (Yes = 1 ; No = 0)
Independent Variables	
Variable	**Description**
NWATER	With clean/safe water (Yes = 1 ; No = 0)
M250	With more than 250m distance from the water source (Yes = 1 ; No = 0)
NTOIL	Without sanitary toilet (Yes = 1 ; No = 0)
INFSTTLR	Informal settlers (Yes = 1 ; No = 0)
MWALL	Makeshift walls (Yes = 1 ; No = 0)
MROOF	Makeshift roofs (Yes = 1 ; No = 0)
HSIZE	Household size

Hunger statistics were based on two indicators, namely (1) a household (HH) is classified as food poor if its per capita income (PCI) necessary to meet the basic food needs of a HH is below the official 2008 Php9,585 food threshold as set by NSCB for Marinduque (FPOOR), and (2) whether or not a HH had experienced food shortage in the past 3 months (FSHORT).

McNemar's Test for paired population was used in order to determine if there is a significant disagreement between the two hunger indicators, FPOOR and FSHORT. Significant predictor variables for both dependent variables were obtained through the chi-square test of independence using 5% level of significance.

For the dependent variable FPOOR, Multiple Linear Regression Model (MLRM) using arcsine square root transformation and stepwise selection, Poisson, Negative Binomial (p=1) (NB1) and Negative Binomial (p=2) (NB2) were fitted. For the second dependent variable FSHORT, Poisson, NB1, NB2, Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) were fitted.

Poisson regression model is a log-linear model written as $\ln E[Y_i | X_i] = X_i'\beta$ where Y_i follows a Poisson distribution with mean μ_i, X_i is a vector of regressors, and β is a vector of regression coefficients. In this case, Y_i has the same conditional mean and conditional variance (equidispersion). If the variance exceeds the mean, overdispersion occurs. NB model is used to account for overdispersion. It has a mean equal to μ_i and variance equal to $\mu_i + r\mu_i^p$ (Cameron and Travedi, 1986). There are two forms of NB models: NB1 (p =1) is used if the function of the variance is linear and NB2 (p=2) when the function is quadratic.

Tests for overdispersion and zero-inflation were performed. Zero-inflation occurs when count data have excess zeroes in the dependent variable. In the zero-inflated model, two processes may arise. The process is determined by the outcome of the Bernoulli trial. The first process involves only zero counts with probability, φ_i. The second process involves counts from either a Poisson or NB model with probability, $1- \varphi_i$.

The ZIP model is given by

$$P(Y = y_i) = \begin{cases} \varphi_i + (1-\varphi_i)e^{-\mu_i} & , y_i = 0, \\ (1-\varphi_i)\dfrac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} & , y_i > 0, \end{cases}$$

where the mean μ and perfect-zero state probability φ depend on vectors of covariates, X_i and Z_i.

The natural links, $\ln(\mu_i)$ and $\text{logit}(\varphi_i)$ linearize Poisson means, that is, $\ln(\mu_i) = X_i\beta$ and

$$\text{logit}(\varphi_i) = \ln\left(\frac{\varphi_i}{1-\varphi_i}\right) = Z_i\gamma.$$

The conditional mean and variance of Y_i are given by

$E[Y_i \mid X_i, Z_i] = \mu_i(1-\varphi_i)$, and

$V[Y_i \mid X_i, Z_i] = E[Y_i \mid X_i Z_i](1+\mu_i\varphi_i)$.

The ZINB model is of the form

$$P(Y = y_i \mid x_i, z_i) = \begin{cases} \varphi_i + (1-\varphi_i)(1 - r\mu_i)^{-r^{-1}}, & y_i = 0, \\ \left[(1-\varphi_i)\left(\dfrac{\Gamma(y_i + r^{-1})}{y_i!\,\Gamma(r^{-1})}\right)\left(\dfrac{r^{-1}}{r^{-1}+\mu_i}\right)^{-1}\left(\dfrac{\mu_i}{r^{-1}+\mu_i}\right)^{y_i}\right], & y_i > 0. \end{cases}$$

The conditional mean and variance of Y_i are given by $E[y_i \mid x_i, z_i] = \mu_i(1-\varphi_i)$ and

$V[y_i \mid x_i, z_i] = E[y_i \mid x_i, z_i](1+\mu_i(r+\varphi_i))$.

Principal component analysis (PCA) using 95% cumulative variation was applied to the count data to minimize multicollinearity and obtain convergence in the zero-inflated regression models.

Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (SBC) as well as the mean square error (MSE) were used in determining the best model. Generated estimates were compared to the actual FPOOR and FSHORT incidences.

3. RESULTS AND DISCUSSION

3.1 Food Poverty and Shortage Incidences

Results showed that 33.9% of the households in Marinduque have PCI below food poverty threshold and are classified as food poor, and 5.1% of these HH have experienced food shortage in the past three months. Actual food poverty and food shortage incidences are shown in Tables 2 and 3. Barangay (Brgy) 6 from Municipality (Mun) 2 is the food poorest barangay with an incidence of 77.7%. Also, four more barangays from Mun 2 are included Table 2. These are 12, 2, 1 and 18 with food poverty incidences of 77.4%, 73.8%, 73.3% and 72.4%, respectively. The barangay with the highest food shortage incidence is Brgy 8 from Mun 4 (65.5%). This is followed by Brgy 20 from Mun 3 with food shortage incidence of 49.65%.

Table 2. Actual Food Poverty Incidences

Mun	Brgy	Barangay Name	Food Poor Incidences
2	6	Barangay IV(Pob.)	0.7770
2	12	Sihi	0.7749
1	55	Tagwak	0.7748
4	33	Sayao	0.7527
2	2	Bagtingon	0.7382
2	1	Bagacay	0.7327
1	15	Boi	0.7273
2	18	Yook	0.7244
3	21	Matandang Gasan	0.7176
1	48	Puting Buhangin	0.6897

Table 3. Actual Food Shortage Incidences

Mun.	Brgy.	Barangay Name	Food Shortage Incidences
4	8	Candahon	0.6552
3	20	Masiga	0.4965
2	18	Yook	0.4160
4	13	Guisian	0.3430
4	35	Silangan	0.3333
3	8	Banot	0.3216
1	53	Tabi	0.3175
3	13	Bognuyan	0.3079
3	15	Dawis	0.2869
2	3	Barangay 1(Pob.)	0.2841

3.2 McNemar's Test

Mcnemar's test indicated that there is a significant disagreement between, FPOOR and FSHORT critera ($p<0.001$). Figure 1 shows that the hunger incidences based on FPOOR and FSHORT hunger indicators are indeed significantly different across all barangays.

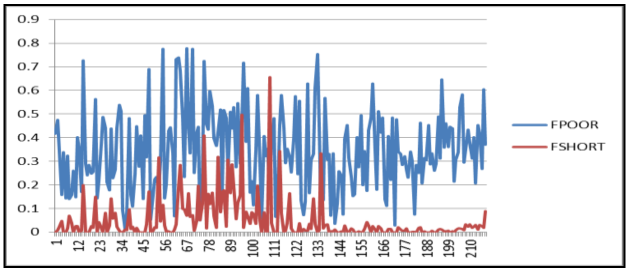

Fig. 1. Comparison of FPOOR and FSHORT incidences

3.3 Food Poverty Models

Poisson, NB1 and NB2 food poverty models are generated with and without PCA. Since the overdispersion parameter *_Alpha* is significantly different from 0 ($p<0.0001$) as shown in Table 4, FPOOR data are overdispersed. This indicates that Poisson regression model is not fitted for the data. Thus, NB1 and NB2 regression models are preferred.

Table 4. Test for Overdispersion

With PCA			
		t-value	*p*-value
NB1	*_Alpha*	9.38	<.0001
NB2	*_Alpha*	9.43	<.0001
Without PCA			
		t-value	*p*-value
NB1	*_Alpha*	9.26	<.0001
NB2	*_Alpha*	9.88	<.0001

Goodness of fit tests show that NB1 model without PCA has the lowest AIC and SBC values as shown in Table 5 which indicates it has a better fit compared to the to the other two count regression models.

Table 5. Summary of AIC and SBC values

Model	With PCA		Without PCA	
	AIC	SBC	AIC	SBC
POISSON	3848	3889	2982	3090
NB1	2103	2116	**2044**	**2067**
NB2	2088	2108	2158	2178

The NB1 model without PCA for FPOOR barangay level count $\hat{Y}$ is given by

$$\ln(\hat{Y}) = 1.699602 + 0.007232\ \text{WELEC} - 0.007546\ \text{TV} - 0.023342\ \text{AIRC} + 0.002869\ \text{NTOIL} + 0.00822\text{MWALL}.$$

In addition, the MLRM parameter estimates were obtained with an adjusted R-square of 59.25%. The MLRM for FPOOR barangay level incidence $\hat{Y}$ is given by

$$\arcsin(\sqrt{\hat{Y}}) = 1.06142 - 0.46193\text{EFAN} - 0.70463\text{COMP} - 0.22391\text{CFONE} - 0.13876\ \text{CROP} - 0.17348\ \text{FISH} - 0.09398\ \text{FOR} - 0.68892\ \text{MIN} + 0.46416\ \text{CNS}.$$

The lower MSE of MLRM (0.01215) compared to that of NB1 without PCA (0.0272) indicated that MLRM is the best fit food poverty (FPOOR) model.

3.4 Food Shortage Models

Poisson, NB1 and NB2 food poverty incidence models are generated with and without PCA. MLRM did not fit the FSHORT data using various transformations. However, test for zero-inflation showed that the number of excessive zeroes in the FSHORT response variable is significant, and hence, zero-inflated regression models should be considered. Table 6 shows that the zero-inflation parameters for ZIP ($p= 0.003$) and ZINB ($p < 0.0001$), and hence, these zero-inflated models are more appropriate.

Table 6. Test for Zero-Inflation in the data

Model		*t*-value	*p*-value
ZIP	*Inf_parameter*	2.97	0.003
ZINB	*Inf_parameter*	19.73	<.0001

PCA was applied in order to meet convergence in generating zero-inflated models. The parameter estimates of the ZIP and ZINB food shortage models are shown in Table 7.

Table 7. Zero-Inflated Regression Estimates

Variable	**Parameter Estimates (β)**	
	ZIP	**ZINB**
INTERCEPT	0.2338	-0.2745
REF	-0.0041	-

RADIO	-0.0023	-0.0067
NWATER	-0.0018	-
TFONE	0.0125	-
MROOF	-0.0164	-
MAL05	-0.0077	-
UPAR	0.0151	-
MIN	-0.0094	-
INFSTTLR	-0.0102	-
HSIZE	-0.0005	-
PREVM	0.0555	0.0705
CURE	0.0038	-
TRN	-0.0057	-
FOR	-0.0024	-
SERV	-0.0145	-
M250	-0.0075	-0.0113
SAL	0.0087	-
OFW	-0.0138	-
NTOIL	0.0074	0.0175

Variable	Parameter Estimates (β)	
	ZIP	ZINB
INF_INTERCEPT	1.0649	18.7147
INF_REF	0.0118	-2.4173
INF_RADIO	-	6.5465
INF_EXPM	-	-11.584
INF_TFONE	-	-18.2984
INF_MROOF	-0.1302	-5.7268
INF_MAL05	-0.0786	4.7062
INF_FISH	0.0201	-
INF_UPAR	-	5.781
INF_HSIZE	-	0.169
INF_PREVM	-0.1803	10.0959
INF_CURE	-	-2.56
INF_TRN	-	-6.1279
INF_MAN	-0.2003	-4.1773
INF_M250	-	-3.0569
INF_NTOIL	-0.019	-7.7723

Test for overdispersion in Table 8 indicates that ZINB is preferred than ZIP model.

Table 8. Test for Overdispersion for ZIP and ZINB

		t-value	*p*-value
ZINB	*_Alpha*	9.88	<.0001

This is further supported by the goodness of fit tests in Table 9 indicating that ZINB has the lowest AIC and SBC of 1313 and 1381, respectively.

Table 9. Goodness of Fit Tests for ZIP and ZINB

Model	AIC	SBC
ZIP	3152	3246
ZINB	**1313**	**1381**

The ZINB Regression Model for FSHORT barangay level count is given by,

$$\ln(\hat{Y}) = -0.2745 - 0.0067RADIO + 0.0705PREVM - 0.0113M250 + 0.0175NTOIL,$$

along with the regression model for zero-inflation given by

$$\mathrm{logit}(\varphi_i) = 18.7147 - 2.4173\ Inf\_REF + 6.5465\ Inf\_RADIO - 11.584\ Inf\_EXPM - 18.2984\ Inf\_TFONE - 5.7268\ Inf\_MROOF + 4.7062\ Inf\_MAL05 + 5.781\ Inf\_UPAR + 0.169\ Inf\_HSIZE + 10.0959\ Inf\_PREVM - 2.56\ Inf\_CURE - 6.1279\ Inf\_TRN - 4.1773\ Inf\_MAN - 3.0569\ Inf\_M250 - 7.7723\ Inf\_NTOIL.$$

Barangay level food shortage counts were estimated using the ZINB model and the corresponding incidences were computed and compared with the actual FSHORT incidences. Figure 2 shows the comparison of the actual FSHORT incidences and the ZINB estimates

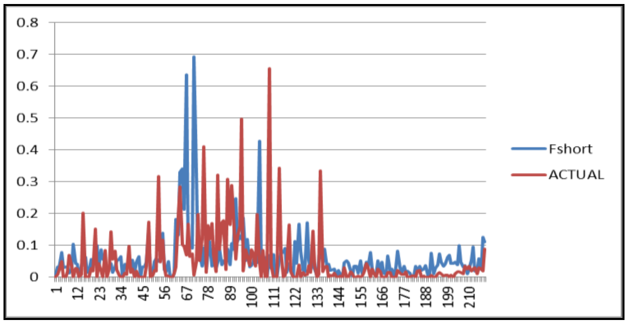

Fig. 2. FSHORT Incidences and ZINB Estimates

4. CONCLUSIONS

MLRM, Poisson, NB1 and NB2 regression models were obtained for FPOOR data. Results indicated that the data are overdispersed, and hence, NB models are preferred. Goodness of fit tests showed that NB1 is of better fit than the NB2 model. However, the MSE of MLRM is lower than that of NB2 indicating that MLRM with arcsine square root transformation is the best food poverty model. The MLRM suggests that the more households in a barangay with cellular phones (CFONE), computers (COMP), electric fans (EFAN), and are engaged in forestry (FOR), fishing (FISH), mining (MIN), and crop farming/gardening (CROP), the lower is its food poverty incidence.

For food shortage (FSHORT) count data, overdispersion and zero-inflation were indicated. Poisson, NB1, NB2, ZIP and ZINB models were obtained. Goodness of fit tests showed that ZINB is the most appropriate model. The ZINB model suggests that households who received cure or treatment from sickness, with expected, deceased, malnourished, and single parent member/s, households with refrigerator, radio and telephone, engaged in transportation and manufacturing, households with roofs made from makeshift materials, with more than 250m distance from the water source and without sanitary toilet/s and household size have significant effects to the food shortage barangay level counts and incidences.

For further studies, Bayesian and multivariate approaches may be considered. Panel estimation may be integrated with Poisson, NB, ZIP and ZINB models.

5. ACKNOWLEDGEMENTS

The researchers would like to express their sincere gratitude to Community Based Monitoring System (CBMS) and local government unit of Marinduque province for sharing the needed data.

6. REFERENCES

Andan, J., Cortez, A., & Ocampo, S. (2013). Clustering and model-fitting of overdispersed food poverty data, Proceedings of Research Congress 2013. DLSU.

Arcilla, R. Co, F., & Ocampo, S.. (2011). Correlates of poverty: Evidence from the Community-Based Monitoring System (CBMS) data. DLSU Business & Economics Review, 20(2), 33-43.

Cameron, C. & Trivedi, P. (1986). Econometric Models based on Count Data; Comparisons and Applications of some Estimators and Tests, Department of Econometrics, Ohio State University.

dela Paz, A., Valera, G., & Ocampo, S.R. (2012). Regression-based estimation methods of barangay level poverty incidences using CBMS Marinduque data. Proceedings of Science and Technology Science and Technology Congress 2012, De La Salle University.

Erdman, D., Jacksin, L., & Sinko, A. (2008). Zero - Inflated Poisson and Zero-Inflated Negative Binomial Models using the COUNTREG Procedure, SAS Global Forum 2008, SAS Institute, Inc,

Friel, S. & Conlon, C. (2004). Food Poverty and Policy. Retrieved on November 5, 2012 from http://www.nuigalway.ie/healthpromotion/documents/General_Staff_Publications/2004_rep_friel_food_poverty_and_policy.pdf

Marinduque: The Heart of the Philippines (2012). Retrieved on November 5, 2012 from http://www.marinduque.gov.ph.

Remo, M . (2013). PH seen missing Millennium Development Goals, Philippine Daily Inquirer.

PEP-CBMS Network Office (2011). Poverty maps of the province of Marinduque. Community-Based Monitoring System.

PEP-CBMS Network Office (2011). The many faces of poverty: Vol. 2. Community-Based Monitoring System, 95-138.