

## Model-Based Estimation Methodology for the Annual Food and Poverty Thresholds

*This set of procedures corresponds to the methodology in generating model-based estimates of food and poverty thresholds at the provincial level, without using the actual commodity prices for the time period under consideration. In particular, the following methodology was followed in generating model-based estimates of the 2006 and 2007 thresholds. As more data become available, the interim methodology shall be improved using appropriate statistical tools.*

### I. Summary of the Model-Based Estimation Methodology

The effective annual food threshold growth rate from 2000 to 2004 was modeled with all provinces in the Philippines serving as observations (row entries), while the variables (column entries) are discussed in section II of this Annex. The resulting model is:

$\hat{Y} = 5.612 + 0.02181 * \text{prov'l pov inc 2000} - 0.0467 * \text{proportion of bought (fruit)}$ <div style="display: flex; justify-content: space-around; font-size: small;"> <span>(1.055)</span> <span>(0.09)</span> <span>(0.014)</span> </div>
--

*Note: The figure in parenthesis shown below each coefficient corresponds to the standard error of that particular estimate.*

**Adjusted R- squared = 0.199**

From the list of complex models constructed from stepwise regression procedures, the final model was chosen such that it yields desirable properties such as consistency of statistical significance of the exogenous variables and approximately "stable" parameter estimates across different time periods.

For the purpose of forecasting thresholds, emphasis is given to the accuracy of forecasts rather than the interpretability of the structure of the final model. The set of actual 2005 food thresholds was used to compute for the measures of forecast errors. On the average, the actual 2005 food thresholds computed using the actual price data differ by P548.19 from the model based estimates, at the annual basis. This is approximately equal to P1.51 mean absolute difference of between the two sets of estimates of food threshold per day, for the year 2005.

Final Model	
Mean absolute error	548.19
Root mean squared error	703.05
Mean absolute percentage error	5.40%

Operationally, to compute for the model-based estimate of the food threshold of a province for a future time period t,

$$\text{Estimate of food threshold}_t = \text{food threshold}_{\text{base year}} * (1 + \hat{y})^{t - \text{reference year}}$$

Estimate of poverty threshold<sub>t</sub> = Estimate of food threshold<sub>t</sub> / (FE/TBE)

where:  $\hat{y}$  = estimate of effective annual growth rate generated from the model

FE/TBE = ratio of food expenditure to total basic expenditure from the latest Family Income and Expenditure Survey (FIES)

It has to be noted that in computing the forecast for a time period of interest, careful analysis has to be undertaken in choosing the appropriate reference year into which the forecasted effective annual growth rate will be applied. Ideally, the reference year to be used in the construction of the endogenous variable should be extended in the process of forecasting. However, when "ambiguities" arise from such choice (e.g., decreasing food threshold from one time period to the next), other reference years may be adopted.

The regional and national food thresholds are computed as weighted means of the provincial food thresholds using the magnitude of food poor population from the latest FIES as weights. Similarly, the corresponding regional and national poverty thresholds are computed as weighted averages of the provincial poverty thresholds with the magnitude of poor population from the latest FIES as weights.

## **II. Detailed Procedure Used in Model Building**

This procedure provides a forecast equation in estimating thresholds without the use of actual price data. In general, this is empirical in nature and emphasis is given to the generation of estimates of food and poverty thresholds to address timeliness of estimates. Among numerous preliminary models generated, the final model performs very satisfactorily. However, the theoretical framework of the final model's structure and model interpretability are not weighted as much as the statistical properties of the model.

While the food CPI for the bottom 30% is the more appropriate independent variable, the all income food CPI was used instead since the former is not readily available and would require a great amount of computational effort.

### **A. Independent Variables**

The initial stage of the model building procedure started with the collection of all available possible regressors following the criteria below:

1. Available data that are directly involved in the computation of thresholds and other relevant poverty statistics;
2. For CPI-based variables, the list was only restricted to those which correspond to a significant proportion of the food menu;
3. Available data for the period 2000 to 2005; and
4. Available provincial level data.

Based on these criteria, the following is a complete list of all possible independent variables:

<b>Variable</b>	<b>Period covered</b>	<b>Number of variables</b>	<b>Source of Basic Data</b>
1. Growth rate of food threshold	2002-2003, 2001-2002, 2000-2001	3	NSCB menu-based estimates
2. Effective annual growth rate of food threshold	2000-2003, 2000-2002	2	NSCB menu-based estimates
3. Reciprocal of growth rate of food threshold	2002-2003, 2001-2002, 2000-2001	3	NSCB menu-based estimates
4. Interaction / product of growth rates of food threshold	2000-2001 and 2001-2002, 2000-2001 and 2002-2003, 2001-2002 and 2002-2003	3	NSCB menu-based estimates
5. Growth rate of food CPI	2002-2003, 2001-2002, 2000-2001	3	NSO
6. Growth rate of rice index	2002-2003, 2001-2002, 2000-2001	3	NSO
7. Growth rate of meat index	2002-2003, 2001-2002, 2000-2001	3	NSO
8. Growth rate of dairy products index	2002-2003, 2001-2002, 2000-2001	3	NSO
9. Growth rate of fish index	2002-2003, 2001-2002, 2000-2001	3	NSO
10. Proportion of <i>barangays</i> (in the province) with at least one market		1	<i>Barangay</i> profile from the 2000 Census of Population and Housing (CPH), NSO
11. Proportion of <i>barangays</i> with access to major highways		1	<i>Barangay</i> profile from the 2000 CPH, NSO
12. Proportion of rice consumption bought by households, at the regional level		1	Food Consumption Survey (FCS), Food and Nutrition Research Institute (FNRI)
13. Proportion of fruit consumption bought by households, at the regional level		1	2003 FCS, FNRI
14. Proportion of vegetable consumption bought by households, at the regional level		1	2003 FCS, FNRI

<b>Variable</b>	<b>Period covered</b>	<b>Number of variables</b>	<b>Source of Basic Data</b>
15. Official provincial poverty incidence	2000, 2003	2	NSCB menu-based estimates

*Note that in the discussion in section I of this Annex, the independent variables finally included in the model were: 1) 2000 provincial poverty incidence; and 2) proportion of bought component for fruit consumed by households/individuals. These two variables were found to be persistently statistically significant in explaining the variation of the dependent variable. Further, in varying model specifications wherein reference years of these variables were changed, the resulting coefficients for the two variables showed no significant differences.*

Since there are many possible independent variables with only 81 observations, parameter estimation of a full model may be computationally intensive. Hence, factor analysis, a data reduction technique, was utilized. In addition, this technique was also used to generate a list of more comprehensive possible independent variables. This is to account for simultaneous linear interactions among the initial list of exogenous variables with respect to modeling the annual food threshold growth rate. For the purpose of this study, it must be noted that any model involving “factors” as independent variables was not considered for forecasting as this required more computational and analytical burden. Hence, such models were only used to support the relationships established from the other models.

## **B. Dependent Variables**

Two link functions were used for the dependent variable: (i) the *identity function of the food threshold growth rate from 2003 to 2004* and (ii) the *effective annual food threshold growth rates from 2000 to 2004*. Note that the effective annual provincial food threshold growth rate incorporates information about each year-on-year inflation.

## **C. Data Cleaning**

Data cleaning was undertaken by estimating missing values and making adjustments for aberrant observations. Missing values for the CPI-related variables were estimated using the formula:

$$\text{Imputed for missing} = \{ (\text{Actual food CPI}\%) + (\text{Regional food CPI}\%_{t-1}) \} / 2$$

For each variable, boxplots were constructed. Outliers were identified as values falling outside the boxplots’ fences. Such aberrant observations were adjusted using the formula:

**“Cleaned data”**

=  $(\text{Original value} + 2 * (\text{Mean} + 1.5 * \text{IQR})) / 3$ , if original  
value is aberrant  
= Original value, otherwise

where: IQR is the interquartile range

#### **D. Generation of Preliminary Models**

Stepwise regression procedures were conducted to generate competing models. Significance levels of 0.11 and 0.15 for point of entry and removal, respectively were used in the regression procedures. Note that these levels are relatively higher than the usual significance level of 0.05 or 0.10. This is to accommodate “intuitive” exogenous variables for modeling the food threshold inflation to be included in the models.

#### **E. Improvement of the Preliminary Models**

The use of backcast models and extreme bound analysis facilitated the identification of exogenous variables that are persistently statistically significant in explaining the variation in the dependent variable. This was done to compensate, to some extent, the lack of formal economic theories that support the relationships established from the models.

#### **F. Model Adequacy Checking**

Model assessment was undertaken through comparison of statistics on goodness of fit, measures of parsimony, and forecast errors generated by competing models and practical implications of the established relationships. The validity of the assumptions of the regression analysis was also assessed. This set of information was evaluated to select the final model.

### **III. Additional Notes**

Overall, this methodology of model-based estimation of food thresholds is inclined to produce long-run estimate(s) of threshold inflation (since the variable being modeled is the effective annual food threshold growth rate). Consequently, when two time periods are being compared, there is higher probability of minimal downward bias with respect to the realized value, when the realized value is brought about by erratic price increases among basic food commodities.

### **IV. Statistical Advocacy**

Towards appreciation of model-based estimation of food and poverty thresholds among data users, the NSCB will conduct statistical advocacy and public information activities that present the limitations of the methodology, in partnership with other government agencies, such as the *Bangko Sentral ng Pilipinas*, among others.