Engel Model of Food Consumption and Expenditure

This report estimates Engel curves for a comprehensive, detailed set of food categories using data tabulated from Chinese household surveys. The estimates are intended to capture empirical patterns described in figures 2-5:

1. Rising expenditures for a particular food category can reflect increases in quantity purchased as well as increases in unit value.

2. Engel relationships can be nonlinear.

Income elasticities are estimated for a more detailed breakdown of food categories than has been available from previous studies. The relationship of consumption and expenditure to income is carefully characterized, with no attempt to estimate price effects.[7]

[7]The study uses cross-sectional data from two years of relative price stability, so the data contain relatively little variation in prices.

"Quality" Effects in Engel Relationships

The Engel curve is most commonly expressed as the relationship between household expenditure on item i, e_i, and household income, y. Expenditure is the product of the price, p_i, and quantity purchased, q_i, of item i. In the simplest Engel function, the price is assumed to be independent of y, and the relationship between e and y reflects changes in the quantity purchased in response to a change in y while holding prices fixed:

$$e_i(y) = p_i q_i(y). \qquad (1)$$

The relationship $q_i'(y) > 0$ for a normal good implies $e_i'(y) > 0$. In this simple case, the elasticities of expenditure and quantity with respect to y are equal.

In empirical applications, expenditures on a food category may increase through increases in the "price" as well as quantity purchased. Household survey data used to estimate Engel equations typically consist of household expenditures and quantities for fairly broad categories of food items, such as grain, meats, poultry, fish, vegetables, and fruit. The average "price" for a broad budget category is calculated as the unit value: the ratio of expenditures to quantity purchased. The calculated unit value is actually the average of the prices paid for individual items within the category.

The "quality" component of food expenditures arises from the heterogeneity of food products with varying degrees of quality, processing, marketing services, and safety attributes within a food category. For example, "meat" can include various cuts of meat, processed meat products, organic products, and meats purchased from retail outlets that differ in their convenience or reputation for quality. An increase in expenditures on a particular food category may reflect an increase in quantities purchased (e.g., kilograms of meat) or a shift in the composition of products purchased within that food category (e.g., higher value cuts of meat, branded or processed meat products) or both. The shift in composition toward premium products increases the average unit value (expenditures per kilogram) of products purchased. Thus, the increase in unit value is an indication of food "quality."

Equation 1 can be modified by replacing the price with the unit value of foods in category i, v_i, which may vary with income:

$$e_i(y) = v_i(y)q_i(y). \tag{2}$$

The effect $v'_i(y)$ is the "quality" effect on y. The quality effect is positive if consumers purchase products with higher unit values when their incomes increase.

Taking logs of equation 2 and differentiating with respect to *lny*, the elasticity of expenditures for category i with respect to income has two components:

$$\frac{d\ln e_i}{d\ln y} = \frac{d\ln v_i}{d\ln y} + \frac{d\ln q_i}{d\ln y} \tag{3}$$

The expenditure elasticity, ε_i, is the sum of the "quality" elasticity, θ_i, and the quantity elasticity, η_i:

$$\varepsilon_i = \theta_i + \eta_i \tag{4}$$

These methods are similar to those of Hassan and Johnson (1977), who estimated elasticities of food consumption, expenditure, and quality with respect to household income for Canadian households. This report estimates η_i and ε_i from Chinese household consumption and expenditure statistics. The "quality" elasticity is the difference between the expenditure and quantity elasticies:

$$\theta_i = \varepsilon_i - \eta_i \tag{5}$$

Nonlinear Engel Relationships

Most empirical estimates of Engel curves assume a log-linear relationship between quantity consumed and income, but our exploratory analysis found that the log-linear relationship did not fit the data well. The log-linear relationship assumes a constant income elasticity over all levels of y, but the data indicate that income elasticity falls as income grows, reaching zero at high levels of y for some food items. For example, the consumption of pork and eggs tends to rise with income at low income levels, but plateaus when income reaches a high level.In terms of equation 1, $q_i'(y) > 0, q_i''(y) < 0,$ and $q_i'(y)$ approaches zero at high levels of y.[8] Bai and Wahl (2005) found similar nonlinear patterns in nonparametric Engel curves estimated for urban households in Shandong Province. Banks, Blundell, and Lewbel (1999) have emphasized the importance of nonlinearities in Engel curves.

Nonlinear Engel relationships may reflect physical saturation of demand or nonhomothetic consumer preferences. For example, low-income households may have unsatisfied demand for pork, so more income leads to greater pork purchases. At higher income levels, the demand for pork may top out. Or high-income consumers may prefer to spend additional food dollars on a wider variety of meats or seafood.

We use the log-log-inverse (LLI) form of the Engel equation, which allows the income elasticity to vary with income:

$$\mathrm{In}q_{ij} = \alpha_i + \beta_i(1/y_j) + \gamma_i \mathit{In}y_j + u_{ij}, \tag{6}$$

[8] If $q_i'(y)$ approaches zero, the income elasticity, $\eta_i = q_i'(y) \times \left(\frac{y}{q_i}\right)$, also approaches zero as y increases.

where the dependent variable q_{ij} represents the per capita quantity of the *i*th food consumed by the *j*th household.[9] The independent variable y_j represents the per capita income of *j*th household, and u_{ij} is a random disturbance term. The parameters α_i, β_i, γ_i are to be estimated. The LLI functional form does not satisfy the adding-up criterion, but this was not a concern since we did not estimate a complete demand system.[10]

The LLI form has the advantage of being a fairly flexible functional form with only three parameters to estimate. It allows the income elasticity to vary with income and change sign. The LLI has two familiar functional forms nested in it. If $\beta_i = 0$, the LLI simplifies to the familiar "double-log" model. If $\gamma_i = 0$, the LLI model simplifies to the log-inverse model.[11]

The quantity elasticity of the *i*th food category, η_i , is calculated:

$$\eta_i = \beta_i / y_j + \gamma_i . \tag{7}$$

The income elasticity varies with income, y_j, if the estimate of $\beta_i \neq 0$. If $\beta_i < 0$ and $\gamma_i > 0$, then η_i decreases as y increases and may reach zero when $\beta_i/y = \gamma_i$ and become negative if $\beta_i/y > \gamma_i$. If $\beta_i = 0$, the income elasticity is independent of the level of income (the double-log model) and equals γ. If $\gamma_i = 0$ (the log-inverse model), the income elasticity equals $-\beta_i/y$ and also varies with income, but it never reaches zero or changes sign.

An expenditure equation is specified in the LLI functional form as:

$$\mathit{ln}\,e_{ij} = \alpha_i * + \beta_i * (1/y_j) + \gamma_i * \mathit{ln}\,y_j + u_{ij} , \tag{8}$$

where the dependent variable e_{ij} represents per capita expenditure on the *i*th food by the *j*th household. The independent variable y_j represents per capita income of the *j*th household, and u_{ij} is a random disturbance term. The estimated expenditure elasticity is calculated from the estimated coefficients and depends on the level of income:

$$\varepsilon_i = -\beta_i * / y + \gamma_i * . \tag{9}$$

Finally, we estimate the quality elasticity as the difference between the estimates of η_i and ε_i.

Data and Estimation

The Engel function is ideally estimated with household-level data, but such data were not available for this study. This study fitted regression equations to group means of per capita quantities, expenditures, and disposable income published by China's National Bureau of Statistics (NBS). NBS annually publishes mean values of income, expenditure, and amount consumed/purchased as calculated from large household surveys (see appendix, "China Household Survey Data"). Rural means for a limited number of food categories are reported for income quintiles. Urban means for detailed food categories are reported for the highest two deciles, the lowest two deciles, and the middle three quintiles. Standard errors are not published, but the means are based on large samples so the standard error is

[9] Hassan and Johnson estimated Engel equations using household totals for consumption and income, but the current study had access only to per capita averages.

[10] The adding up criterion is derived from the consumer's budget constraint that the sum of expenditures on all items equals total expenditure. This condition implies that $\sum_{i=1}^{n} \left(\frac{e_i}{e} \right) \eta_i = 1$.

[11] Hassan and Johnson compare properties of the LLI and several other functional forms used to estimate Engel equations.

likely quite low. Each rural quintile includes 13,600 survey households. Urban deciles include about 4,700 households each, and urban quintiles include about 9,700 households each.

In our regressions, we treated the average per capita values of consumption and income for an income class as an observation of a representative household at the corresponding income level (see box, "Estimating Engel Equations Using Group Means"). The income tabulations gave us seven urban observations and five rural observations for each food category for each year. We pooled the data from these tabulations for 2 years, 2002 and 2003, giving us a total of 14 urban observations and 10 rural observations. Food prices were relatively stable between 2002 and 2003, so the assumption that prices are held constant when we estimated our regressions seemed reasonable.[12]

Engel equations of the form (6) and (8) were estimated for each available food category for urban and rural households using ordinary least squares. A year dummy variable, d_j, equal to 1 for observations from year 2003 and 0 for observations from year 2002, was added to the regression model to capture any shifts in demand between the 2 years due to factors beside income.

The Chinese household survey data are compiled from diaries of income and expenditures kept by sample households. The diaries are kept by a household member year round with assistance from NBS enumerators who visit the household periodically. For urban households, records of food purchases for consumption at home are the primary measure of consumption. This may overstate consumption to the extent that some food purchased is wasted, given away, consumed by guests, or not consumed for other reasons. Expenditures are also recorded, which allows the computation of unit values that approximate the average price paid. (Expenditures on food away from home are recorded, but there is no breakdown on what foods are purchased or consumed away from home.) The quantity for rural households is per capita quantity consumed, and includes both purchased and self-produced food.[13] Quantity data do not include consumption in restaurants, cafeterias, or other foodservice establishments.

The income measure used for urban households was per capita disposable income. This includes income from wages, business earnings, interest, and transfer payments, less tax and social insurance contributions. It excludes proceeds from loans or sale of personal items. Per capita net income for rural households includes net income from farming and other businesses plus interest, transfer payments and remittances. Farming income includes the imputed value of products grown and consumed or used on the household's own farm. We did not deflate income since there was little inflation between the 2 years.

[12]In 2003, China's change in consumer prices for most food commodities ranged from -1 percent to 3 percent. We excluded data from 2004 because large changes in prices occurred between 2003 and 2004. We did not include data from years prior to 2002 because the sampling method of the urban household survey was changed between 2001 and 2002 to include a larger number of households from small cities and towns. Tabulations of rural household data by quintile were not available for years prior to 2002.

[13]About 40 percent of food "expenditures" for rural households are the imputed value of self-produced food (Gale et al., 2005). The tabulations of rural household data used for this study do not report expenditures, so the rural analysis only estimates Engel equations for quantity consumed.

Estimating Engel Equations Using Group Means

Econometric models must often be estimated from aggregated data when values for individual observations are unavailable or too costly to obtain or analyze. Many analyses use grouped data such as means for states, countries, or demographic groups.

Engel models are ideally estimated using household data, but the current analysis fits regression models to group means. This approach provides an accurate estimate of the relationship between household consumption/expenditures and income as long as other factors (besides income) that affect household consumption are not correlated with income.

The group means may be viewed as representative households for each income group. For a quintile A, consisting of N_A households, the mean values for income and quantity consumed of food item i are:

$$\bar{y}_A = \frac{1}{N_A} \sum_{j \in A} y_j \text{ and } \bar{q}_{iA} = \frac{1}{N_A} \sum_{j \in A} q_{ij}.$$

The disturbance for the means for quintile A is the average of disturbances for the households in quintile A:

$$\bar{u}_{iA} = \frac{1}{N_A} \sum_{j \in A} u_{ij}.$$

The means of q and y will have the same relationship as the household level q and y and the model for quintile A is therefore:

$$\text{In } \bar{q}_{iA} = \alpha_i + \beta_i (1/\bar{y}_A) + \gamma_i \text{ In } \bar{y}_A + \bar{u}_{iA},$$

Kmenta (1971) showed that ordinary least squares estimates obtained from group means are unbiased estimates of the parameters, but the variance of the error for each mean is proportional to the number of individual observations in each group. Consequently, errors are heteroskedastic and parameter estimates are inefficient if the groups are of different sizes. In this case, weighted least squares can obtain efficient estimates.

The rural household models are estimated with means for quintiles that contain equal numbers of households, so no corrective action is needed. The urban models are estimated using three quintiles (containing over 9,000 households each) and four deciles (containing over 4,500 households each). Urban models are estimated with weighted least squares using the corresponding number of households in each group as weights.

Compared with other groupings (e.g., provincial means used in other studies), income quantiles are a particularly useful grouping for estimating Engel relationships. Kmenta showed that the difference between the variance of the group-mean estimator and the variance of the estimator obtained from individual observations (the variance of the group-mean estimator is always larger) depends on the ratio of within-group to between-group variation in the explanatory variable. Using a grouping sorted by the explanatory variable (e.g., income quantiles) minimizes the within-group variation in the explanatory variable and maximizes the between-group variation as compared with other possible groupings. The relatively large degree of variation in the explanatory variable improves the efficiency of the group-mean estimator.

Models estimated from grouped means also tend to have high R^2 values (Cramer, 1964). When averaged over N households in quintile A, positive and negative random errors will cancel one another out, and variation in the group means will be much lower (and the R^2 much higher) than the variation in individual household values. We obtain very high R^2 values in our models because most of the variation in q_i due to measurement error, individual effects, and other factors besides y is removed by using group means.