ARE 251/ECON 270A
Empirical Assignment
Poverty and Inequality in Nicaragua

Due Tuesday September 21, 2004

This assignment will give you some experience using household survey data to characterize the well-being of households. The data are an extract from the 1998 Nicaraguan Living Standards Measurement Survey (LSMS). They include household identifiers, household weights, regional price indices, two poverty lines and a variety of individual and household characteristics. The variable definitions and data (in STATA7 format and Excel) are available at http://are.berkeley.edu/ARE251/assignments. Documentation describing the data - in particular the sampling scheme - is available at http://www.worldbank.org/lsms/ under Documents.

Report all of your results as you would in a journal article, including sample sizes, statistical information, tests and weighting used, etc. Create appropriate tables with titles, column headings and table notes. **Do not** include program code or outputs in your report! When reporting your results do not just give statistical findings, but also comment on the economic meaning and importance of the results as far as possible.

1. The data include a measure of total annual household consumption and an estimate of total household income. Prices vary across the country, thus the consumption and income indicators must be normalized to obtain "real" consumption and income values. After doing this, calculate the P0 and P2 measures of poverty for rural and urban areas separately. Do this for both consumption and income using an equivalence scale of 1 (i.e. per capita values). Do this for both the higher and lower poverty line. Put the results in a table (or 2) that conveys as clearly as possible to the reader what you view as the most interesting comparisons. Comment on what the table shows about the state of poverty in Nicaragua at the time of the survey. Note that the sample is not self-weighting and you need to use the household weights. Explain how you do this weighting.

2. In (1) you used per capita values to measure well-being of household members. But there are strong arguments to suggest that household demographics affect the relationship between consumption (or income) and the average welfare level of household members. To explore this issue, take real household consumption as your indicator of well-being, the higher poverty line, and the poverty gap measure of poverty. Consider the model of equivalence scales: $AE = (A + \alpha K)^\theta$, where A is the number of adults, K is the number of kids (0-14), and θ reflects economies of scale. First setting $\theta=1$ and $\alpha=1$, calculate poverty for households with different demographics (eg, single adult, couples, households with fewer than average kids, more than avg. kids – think about what makes sense as categories). Again, when calculating any statistics – including average number of kids - you need household weights. Now, experiment first with varying α, and see at what levels your conclusions about the relative poverty of the household types begin to differ. Similarly with θ. As in (1) consider how to present your results in an informative manner and suggest what they might mean for targeting.

3. Thus far we have ignored standard errors. However, the data are from a sample and there is certainly sampling error in the estimated poverty rates. Using an equivalence scale of 1 and real income, consider the headcount with the lower poverty line for seven regions. Read

about the sampling scheme used.   What information do you need in the data to correctly estimate the standard errors on the regional estimates? Is that information there – or, if not, can you deduce it from any data that are there? If so, calculate unbiased estimates. If not, calculate the best estimates possible and comment on potential biases.

4. Construct a poverty profile (using the headcount, lower poverty line, real income). Present the results in a table and discuss what they suggest about the relative risk of poverty of different groups.

5. Suppose that you want to organize a transfer program to the poor. Very detailed LSMS data typically are available for just a very small subset of the population. However, simpler data sets often exist with much wider coverage.   If this is the case, one approach to targeting is to construct a model of the probability of being poor (or something similar) using the LSMS data - using only the widely available variables as explanatory variables. Then the model estimates can be used to predict poverty for individuals in the larger population. Estimate such a model for Nicaragua using selected indicators such as quality of dwelling and household and regional characteristics as explanatory variables.   Since you are interested in correlates, not causation, endogeneous variables are fine. Split the sample to estimate the model. Why do we split the data? On the other half of the data, predict headcount poverty rates for each person and give 0/1 based on some threshold. Compare to that person's actual headcount. Find a threshold that will give you less than 10% type I error. Explore the relationship between type I and type II error. Again, think carefully about a clear way to present what has been discovered.

6. Using real consumption and equivalence scale that is 1 for males and 0.75 for females, calculate Gini and Atkinson(2) inequality measures for rural and urban areas separately. Compile a table of your results and comment on what they suggest – in particular note any differences between the two indicators.