A Stratified Sample or a Simple Random Sample Approach?

The Case of Mexican Meat Consumption at the Table Cut Level of Disaggregation.

Jose A. Lopez
Jaime E. Malaga

Agricultural & Applied Economics
Texas Tech University

Southern Agricultural Economics Association
Annual Meeting, Orlando, Florida, February 6–9, 2009

Outline

1. Introduction
   - Types of Probability Samples
   - Objectives
   - Empirical Application
   - Previous Studies Vs. This Study
2. Empirical Application - Data
   - ENIGH
   - Data Considerations
3. Empirical Application - Model
   - Censored Demand System
   - Two-Step Estimation
4. Empirical Application - Results
   - Elasticities
   - Regional Differences
5. General Remarks
   - General Remarks

Outline

1. Introduction
   - Types of Probability Samples
   - Objectives
   - Empirical Application
   - Previous Studies Vs. This Study
2. Empirical Application - Data
   - ENIGH
   - Data Considerations
3. Empirical Application - Model
   - Censored Demand System
   - Two-Step Estimation
4. Empirical Application - Results
   - Elasticities
   - Regional Differences
5. General Remarks
   - General Remarks

Stratified Vs. Complex Samples

Three basic types of probability samples (Lohr 1999):

1. Simple random sample (SRS)
   - It is the simplest form of probability sample. In a SRS, every unit in the population has the same chance of being selected.
2. Stratified random sample
   - The population is divided into subgroups called strata. Then an SRS is selected from each stratum, and the SRSs in the strata are selected independently.
3. Cluster sample
   - Observation units in the population are aggregated into larger sampling units, called clusters.

Complex surveys:

- They involve the use of the three different types of probability samples at different stages of the survey.
- They are stratified with several stages of clustering.

Outline

1. Introduction
   - Types of Probability Samples
   - Objectives
   - Empirical Application
   - Previous Studies Vs. This Study
2. Empirical Application - Data
   - ENIGH
   - Data Considerations
3. Empirical Application - Model
   - Censored Demand System
   - Two-Step Estimation
4. Empirical Application - Results
   - Elasticities
   - Regional Differences
5. General Remarks
   - General Remarks

Ojectives

- To highlight the difference between adopting a simple random sample and a stratified sample approach using an empirical aplication.
- To discuss some of the difficulties that emerge in our empirical application not only from using a survey of household incomes and expenditues but also from adopting a stratified sampling approach.

Outline

1. Introduction
   - Types of Probability Samples
   - Objectives
   - Empirical Application
   - Previous Studies Vs. This Study
2. Empirical Application - Data
   - ENIGH
   - Data Considerations
3. Empirical Application - Model
   - Censored Demand System
   - Two-Step Estimation
4. Empirical Application - Results
   - Elasticities
   - Regional Differences
5. General Remarks
   - General Remarks

The Mexican Meat Market

In our empirical application, the Mexican meat market is analyzed. This market is very important for meat exporters because its large size, rapid expansion, meat offal preference, and low per capita meat consumption.

- It is large.
  - Mexico accounts for 8% of the total world meat imports of 13,195,000 MT (1997-2006, USDA 2007).
  - Most Mexican imports come from the United States
    - 79% bovine meat, 84% swine meat, and 92% chicken (2002-2007, Mexican Ministry of Economy 2008)
  - A good share of U.S. meat exports go to Mexico
    - 50% beef and veal, 34% swine meat, and 12% poultry meat (2002-2007, Mexican Ministry of Economy 2008)

The Mexican Meat Market (cont.)

- It is rapidly expanding.
  - According to USDA (2007), from 1997 to 2006, Mexican meat imports
    - More than doubled (increased by 147%),
    - Went from 568,000 MT to 1,405,000 MT,
    - Experienced the fastest growth among the leading importing countries.
    - Swine meat import growth rate = 449%
    - Poultry meat import growth rate = 108%
    - Beef import growth rate = 80%
- It has a relatively high preference for meat offal.
  - A good share of Mexican imports consists of animal remains
  - 22% bovine remains, 36% swine remains, and 16% chicken offal (2002-2007, Mexican Ministry of Economy 2008)

The Mexican Meat Market (cont.)

- Per capita meat consumption still remains low compared to the United States and Canada.
  - According to USDA (2007), from 1997 to 2006,
    - Mexico averaged 61 kg per capita.
    - The Unites States and Canada averaged 122 kg and 98 kg per capita respectively.
  - Mexican per capita meat consumption will continue growing.
  - Mexico will remain an important market for years to come.

Outline

1. Introduction
   - Types of Probability Samples
   - Objectives
   - Empirical Application
   - Previous Studies Vs. This Study
2. Empirical Application - Data
   - ENIGH
   - Data Considerations
3. Empirical Application - Model
   - Censored Demand System
   - Two-Step Estimation
4. Empirical Application - Results
   - Elasticities
   - Regional Differences
5. General Remarks
   - General Remarks

Previous Studies Vs. This Study

- Previous Studies
  - Malaga et al. (2007, 2006), Dong et al. (2004), Gould et al. (2002), Gould and Villarreal (2002), Golan et al. (2001), Sabates et al. (2001), Dong and Gould (2000), Garcia Vega and Garcia (2000), Heien et al. (1989).
  - Use the same data source (ENIGH) used in this study.
  - Have not taken into account the fact that the sample is stratified.
  - This may result in parameter estimates that may not be representative of the population or that may not capture potential differences among the subpopulations (Lohr 1999).
- This Study
  - Incorporates estimation techiniques used in stratified sampling theory to estimates demand elasticities at the table cut level (i.e., beefsteak; ground beef; pork steak; ground pork; chicken legs, thighs and breast; fish, etc.).

Outline

1. Introduction
   - Types of Probability Samples
   - Objectives
   - Empirical Application
   - Previous Studies Vs. This Study
2. Empirical Application - Data
   - ENIGH
   - Data Considerations
3. Empirical Application - Model
   - Censored Demand System
   - Two-Step Estimation
4. Empirical Application - Results
   - Elasticities
   - Regional Differences
5. General Remarks
   - General Remarks

Data

ENIGH

- National Survey of Household Income and Expenditures
  - *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)*
- ENIGH is published by a Mexican Governmental Institution
  - *Instituto Nacional de Estadística Geografía e Informática (INEGI)*
- Extensive Mexican household survey containing information about
  - House infrastructure, appliances and services
  - Household member demographic and sociodemographic characteristics
  - Occupational activities
- Sampling methods are probabilistic, multi-staged, stratified, and conglomerated.
  - Sampling units are selected with a probability from multiple stages, obtained from dividing the population into groups, and made up from the observation units (household members).

Data

ENIGH 2006

- ENIGH's sampling methods are stratified
  - The target population is divided into subgroups (strata) with similar characteristics
    - Stratum 1: hh within a pop. of 100,000 people or more
    - Stratum 2: hh within a pop. between 15,000 and 99,999 people
    - Stratum 3: hh within a pop. between 2,500 and 14,999 people
    - Stratum 4: hh within a pop. of less than 2,500 people
  - A simple random sample is taken from each stratum
  - A weight is assigned to each sampling unit
    - Weight: number of households that the interviewed household represents nationally
- ENIGH performs direct interviews to households during one week
- Data on food, drinks, cigarettes and public transportation is recorded only when the household makes a purchase

Data

ENIGH 2006 (cont.)

Table 1. Number of Observations, Sum of Weights and Average Household Size Per Stratum

Strata	No. of Obs.	Sum of Weights	Avg. hhsize
Str1	7,285	11,473,327	3.99
Str2	3,942	3,241,161	4.13
Str3	1,574	2,837,679	4.52
Str4	4,108	4,554,086	4.28
Total	16,909	22,106,253	4.14

Source: ENIGH 2006 Database, computed by author.

Data

ENIGH 2006 (cont.)

- Not incorporating from ENIGH 2006 the weight variable into the analysis is equivalent to assigning a constant weight of 1,307 (i.e., 22,106,253/16,909) to each observation
  - Malaga et al. (2007, 2006), Dong et al. (2004), Gould and Villarreal (2002), Golan et al. (2001), Sabates et al. (2001), Garcia Vega Garcia (2000), Heien et al. (1989)
- Taking a random sample of 1,000 households and not incorporating the weight variable will only produce a sample that is representative of the 16,909 households assuming a constant weight
  - Golan et al. (2001)
- Restricting the analysis to only strata 1 and 2 (i.e., households that live in cities or towns with a population of 15,000 or more), in ENIGH 2006, is equivalent to excluding 7,391,765 households
  - Malaga et al. (2007, 2006) and Dong et al. (2004)

Data

ENIGH 2006 (cont.)

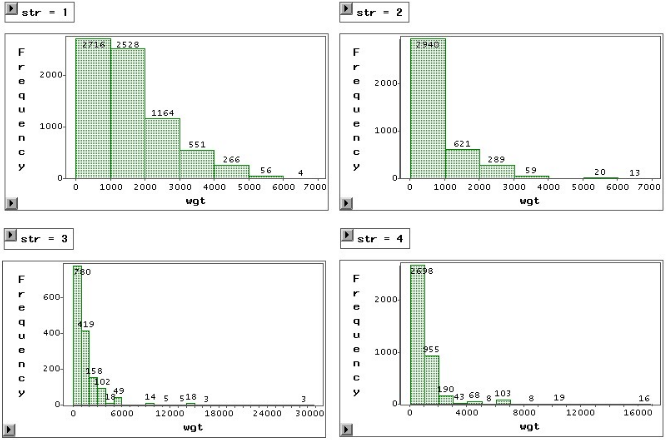


Figure 1. Histogram of the Weight Variable Per Stratum

Source: ENIGH 2006 Database, computed by author.

DuMouchel and Duncan's (1983) Test

To investigate further about the importance of incorporating stratification variables into the analysis, this study implemented the DuMouchel and Duncan's (1983) test

- H_0 favors the use of the unweighted estimator
- H_a favors the use of the weighted estimator
- The test is implemented by performing an F test for $\gamma = 0$ in the following regression model estimated by ordinary least squares

$$\mathbf{Y} = \mathbf{X}\alpha + \mathbf{W}\mathbf{X}\gamma + \epsilon \qquad (1)$$

- Where:
  - $\mathbf{Y}$ is a $(n \times 1)$ vector of observations in the dependent variable
  - $\mathbf{X}$ is a $(n \times p)$ matrix of observations in the independent variables
  - $\mathbf{W}$ is a $(n \times n)$ diagonal matrix whose i^{th} diagonal element is the sample weight w_i
  - α and γ are vector of parameters
  - ϵ is a random error with $E(\epsilon) = 0$ and $\text{var}(\epsilon) = \sigma^2\mathbf{I}_n$ and $\mathbf{Z} = \mathbf{WX}$, where the columns of $\mathbf{Z}$ are further (perhaps unobserved) predictors that should have been included in the regression but were not

Table 2. DuMouchel and Duncan's (1983) Test's Results.

Equation	F	p-value
q_1	1.7907	0.0090
q_2	2.0893	0.0011
q_3	1.7377	0.0126
q_4	1.9422	0.0032
q_5	1.3806	0.0976
q_6	4.3003	$<$0.0001
q_7	3.0603	$<$0.0001
q_8	1.7962	0.0086
q_9	1.7718	0.0101
q_{10}	4.4449	$<$0.0001
q_{11}	1.6708	0.0191
q_{12}	8.3251	$<$0.0001
q_{13}	2.4402	0.0001
q_{14}	9.2035	$<$0.0001
q_{15}	7.3924	$<$0.0001
q_{16}	1.9762	0.0026
q_{17}	1.1127	0.3166
q_{18}	3.7224	$<$0.0001
Critical Values		
$F^*_{25,\ 16884}(0.01) = 1.77$		
$F^*_{25,\ 16884}(0.05) = 1.52$		

Outline

1. Introduction
   - Types of Probability Samples
   - Objectives
   - Empirical Application
   - Previous Studies Vs. This Study
2. Empirical Application - Data
   - ENIGH
   - Data Considerations
3. Empirical Application - Model
   - Censored Demand System
   - Two-Step Estimation
4. Empirical Application - Results
   - Elasticities
   - Regional Differences
5. General Remarks
   - General Remarks

Data Considerations

- Stratified sample
  - Need to incorporate stratification variables (weight and strata)
  - Weighted least squares estimation is consistent (Wooldrigde, 2001; Lohr 1999)
  - However, the standard errors obtained from weighted least squares estimation are incorrect and should be ignored (Lohr 1999)
  - Consequently, standard errors were estimated by bootstrap
- During one week of the interview, not all households reported consumption of all meat cuts (Censoring Problem)
  - Missing observations for some price and quantities
    - To solve the problem of missing prices, a regression imputation approach was adopted for each of the eighteen meat cuts considered in this study
    - Non-missing prices of each meat cut was regressed as function of total income, education level of the household decision maker, regional dummy variables, stratum dummy variables, the number of adult equivalents, a dummy variable for car, and a dummy variable for refrigerator

Data Considerations (cont.)

- To solve the problem of censored quantities, a consistent censored demand system was used
- Variables such as quantity and expenditure are reported at the household level regardless of the household size
  - Households are not comparable
  - Not adjusting is equivalent to assuming a constant household size
  - Compute the number of adult equivalents
    - National Research Council's recommendations of the different food energy allowances for males and/or females during the life cycle as reported by Tedford et al. (1986)
  - Compute per capita meat consumption (i.e., per adult-equivalent consumption)
  - Other studies
    - Use a simple count or proportion of household members (Dong et al., 2004; Golan et al., 2001)
    - Ignore household size

Outline

1. Introduction
   - Types of Probability Samples
   - Objectives
   - Empirical Application
   - Previous Studies Vs. This Study
2. Empirical Application - Data
   - ENIGH
   - Data Considerations
3. Empirical Application - Model
   - Censored Demand System
   - Two-Step Estimation
4. Empirical Application - Results
   - Elasticities
   - Regional Differences
5. General Remarks
   - General Remarks

Censored Demand System

- This study uses the two-step estimation of a censored demand system proposed by Shonkwiler and Yen (1999), but incorporates stratification variables into the estimation procedure
- The censored system of equations can be written as

$$y_i = \Phi(\mathbf{z}_i'\boldsymbol{\alpha}_i)\mathbf{x}_i'\boldsymbol{\beta}_i + \delta_i\phi(\mathbf{z}_i'\boldsymbol{\alpha}_i) + \xi_i, \quad i = 1, \ldots, M$$

- Where:
  - y_i is a (1×1) observed dependent variable
  - $\Phi(\mathbf{z}_i'\boldsymbol{\alpha}_i)$ = standard normal cumulative distribution function (cdf) evaluated at $\mathbf{z}_i'\boldsymbol{\alpha}_i$
  - $\phi(\mathbf{z}_i'\boldsymbol{\alpha}_i)$ = standard normal probability density function (pdf) evaluated at $\mathbf{z}_i'\boldsymbol{\alpha}_i$
  - $\mathbf{z}_i' = \begin{pmatrix} 1 & z_{i2} & \ldots & z_{iK_1} \end{pmatrix}$ is $(1 \times K_1)$ vector of explanatory variables
  - $\mathbf{x}_i' = \begin{pmatrix} 1 & x_{i2} & \ldots & x_{iK_2} \end{pmatrix}$ is $(1 \times K_2)$ vector of explanatory variables
  - $\boldsymbol{\alpha}_i = \begin{pmatrix} \alpha_{i1} & \alpha_{i2} & \ldots & \alpha_{iK_1} \end{pmatrix}'$ is a $(K_1 \times 1)$ vector of parameters
  - $\boldsymbol{\beta}_i = \begin{pmatrix} \beta_{i1} & \beta_{i2} & \ldots & \beta_{iK_2} \end{pmatrix}'$ is a $(K_2 \times 1)$ vector of parameters
  - ξ_i is a (1×1) random error

Outline

1. Introduction
   - Types of Probability Samples
   - Objectives
   - Empirical Application
   - Previous Studies Vs. This Study
2. Empirical Application - Data
   - ENIGH
   - Data Considerations
3. Empirical Application - Model
   - Censored Demand System
   - Two-Step Estimation
4. Empirical Application - Results
   - Elasticities
   - Regional Differences
5. General Remarks
   - General Remarks

Two-Step Estimation

1. Obtain maximum-likelihood probit estimates $\hat{\alpha}_i$ of α_i for $i = 1, 2, \ldots, M$ using the binary dependent variable $d_i = 1$ if $y_i > 0$ and $d_i = 0$ otherwise

$$P(d_i = 1|\mathbf{z}_i) = \Phi(\alpha_{i1} + \alpha_{i2} z_{i2} + \ldots + \alpha_{iK_1} z_{iK_1}) = \Phi(\mathbf{z}_i' \boldsymbol{\alpha}_i)$$

   - Multiply the contribution of each observation to the likelihood function by the value of the weight variable

2. Calculate $\Phi(\mathbf{z}_i'\hat{\alpha}_i)$ and $\phi(\mathbf{z}_i'\hat{\alpha}_i)$ and estimate $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, …, $\boldsymbol{\beta}_M$, δ_1, δ_2, …, δ_M in the system

$$y_i = \Phi(\mathbf{z}_i'\hat{\alpha}_i)\mathbf{x}_i'\boldsymbol{\beta}_i + \delta_i \phi(\mathbf{z}_i'\hat{\alpha}_i) + \xi_i, \quad i = 1, \ldots, M,$$

   by seemingly unrelated regression (SUR) procedure

   - Weight all observations by the weight variable prior to estimation

Outline

1. Introduction
   - Types of Probability Samples
   - Objectives
   - Empirical Application
   - Previous Studies Vs. This Study
2. Empirical Application - Data
   - ENIGH
   - Data Considerations
3. Empirical Application - Model
   - Censored Demand System
   - Two-Step Estimation
4. Empirical Application - Results
   - Elasticities
   - Regional Differences
5. General Remarks
   - General Remarks

Eighteen Table Cuts of Meats

Beef

1. beefsteak
2. ground beef
3. other beef cuts
4. beef offal

Pork

5. pork steak
6. pork leg & shoulder
7. ground pork
8. other pork

Processed Beef & Pork

9. chorizo
10. ham, bacon & similar products
11. beef & pork sausages
12. other

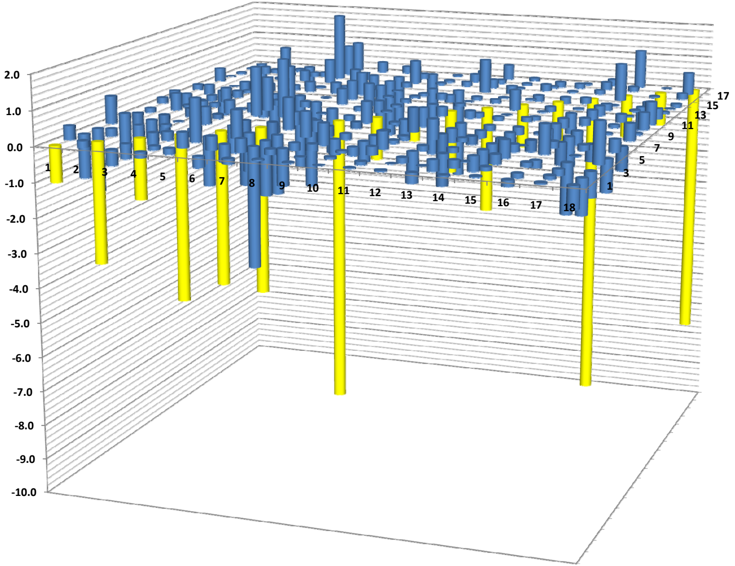
Chicken

13. chicken legs, thighs & breasts
14. whole chicken
15. chicken offal
16. chicken ham & similar

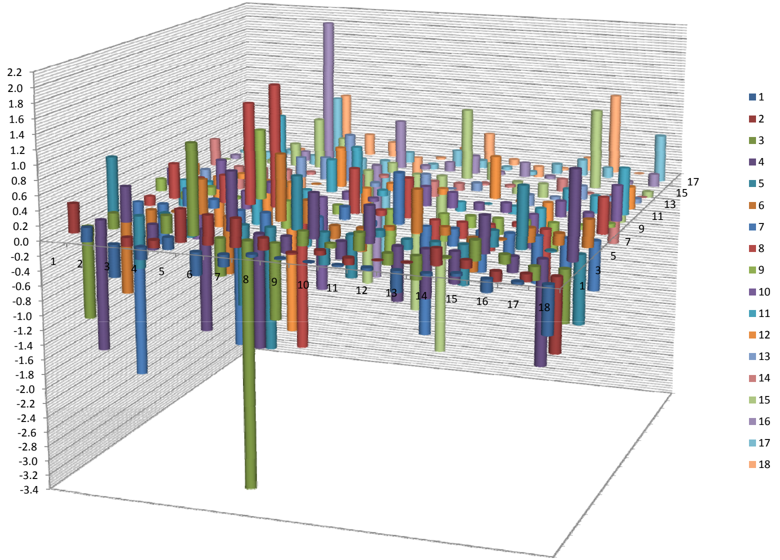
Seafood

17. fish
18. shellfish

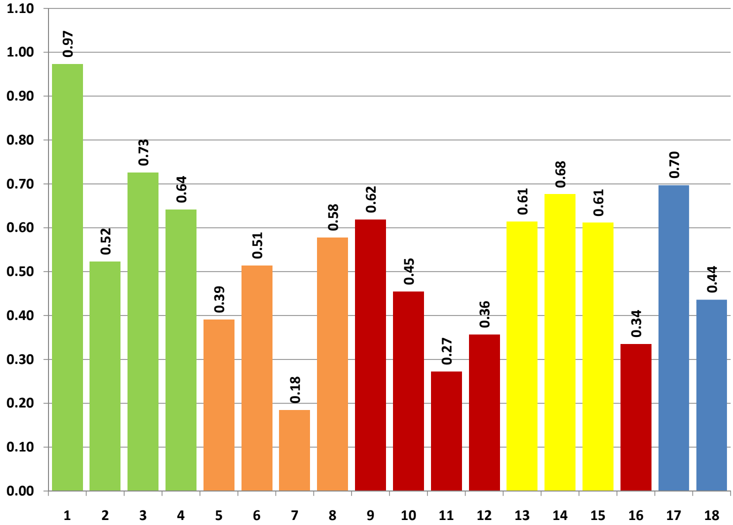
Marshallian Price Elasticities



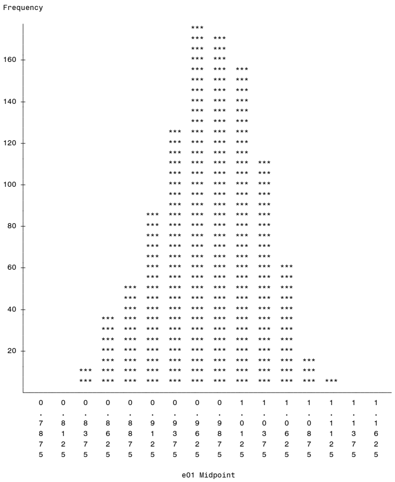
Marshallian Cross-Price Elasticities

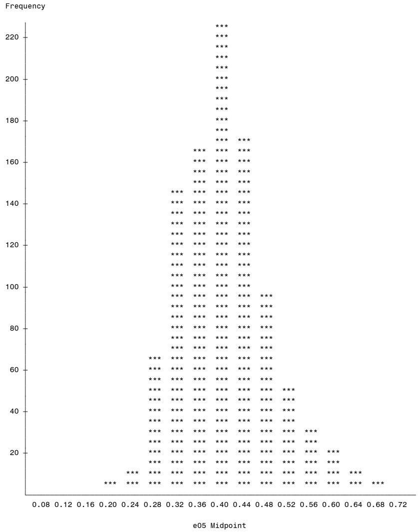


Expenditure Elasticities

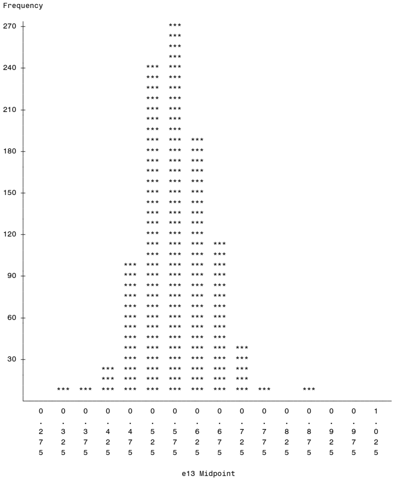


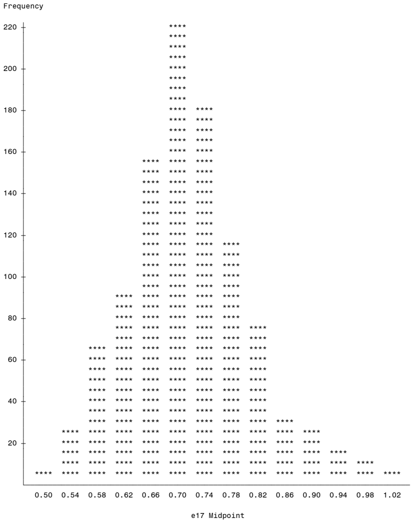
Expenditure Elasticity Distributions





Expenditure Elasticity Distributions (Cont.)





Outline

1 Introduction
- Types of Probability Samples
- Objectives
- Empirical Application
- Previous Studies Vs. This Study

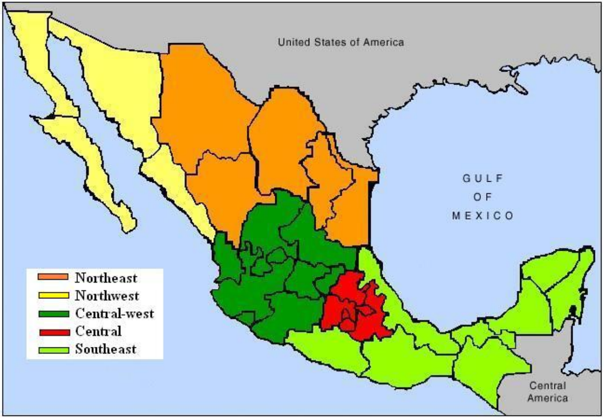
2 Empirical Application - Data
- ENIGH
- Data Considerations

3 Empirical Application - Model
- Censored Demand System
- Two-Step Estimation

4 Empirical Application - Results
- Elasticities
- Regional Differences

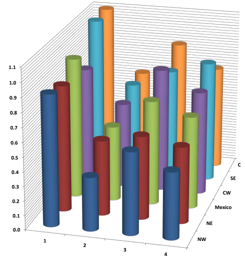
5 General Remarks
- General Remarks

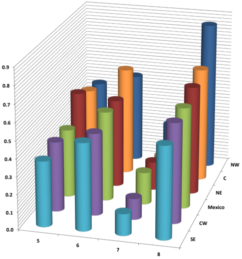
Map of Mexican Geographical Regions

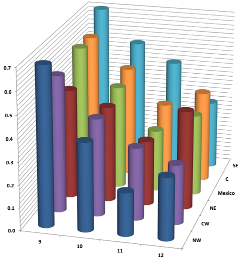


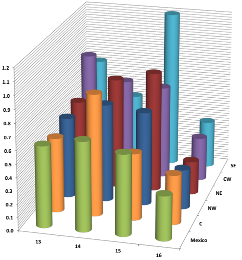
Northeast = Chihuahua, Cohahuila de Zaragoza, Durango, Nuevo León, and Tamaulipas.
Northwest = Baja California, Sonora, Baja California Sur, and Sinaloa.
Central-West = Zacatecas, Mayarit, Aguascalientes, San Luis Potosí, Jalisco, Guanajuato, Querétaro Arteaga, Colima, and Michoacán de Ocampo.
Central = Hidalgo, Estado de México, Distrito Federal, Tlaxcala, Morelos, and Puebla.
Southeast = Veracruz de Ignacio de la Llave, Yucatán, Quintana Roo, Campeche, Tabasco, Guerrero, Oaxaca, and Chiapas.

Expenditure Elasticities by Region









Outline

1 Introduction
- Types of Probability Samples
- Objectives
- Empirical Application
- Previous Studies Vs. This Study

2 Empirical Application - Data
- ENIGH
- Data Considerations

3 Empirical Application - Model
- Censored Demand System
- Two-Step Estimation

4 Empirical Application - Results
- Elasticities
- Regional Differences

5 General Remarks
- General Remarks

General Remarks

- In our empirical application, the Mexican meat market was analyzed because it is an important market for meat exporters not only because its large size, rapid growth, meat offal preference, and low per capita meat consumption.
- The parameter and elasticity estimates obtained are not biased not only because stratification variables are incorporated in the estimation procedure but also because a censored regression model is employed.
- The parameter and elasticities estimates can also be interpreted as population estimates (or viewed as census estimates) because the study uses a stratified sampling methodology and the entire target population.
- Standard errors of parameter estimates are correctly approximated using the bootstrap method because data was obtained from a complex survey.

Thank You.