MEXICAN MEAT CONSUMPTION:

AN APPLICATION OF SEEMINLGY UNRELATED REGRESSIONS

IN STRATIFIED SAMPLING

By

JOSÉ ANTONIO LÓPEZ, M.S.

A MASTER'S REPORT

In

STATISTICS

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

MASTER OF ATRS

Approved

Dr. James Surles
Chair

Dr. Petros Hadjicostas
Committee Member

Fred Hartmeister
Dean of the Graduate School

August, 2008

ACKNOWLEDGMENTS

I am very thankful to my committee chair, Dr. James Surles, for his guidance and encouragement throughout the development of this Master's report. His option, comments, and ideas were very valuable and insightful. Dr. Surles's guidance and support made this thesis come to a conclusion.

I would also like to say thank you to my other committee member, Dr. Petros Hadjicostas. His assistance in this report was very beneficial. His willingness to help and friendly attitude were also really appreciated.

My gratitude is also extended to all other people that were not in my committee and yet provided valuable support. Dr. Jaime Malaga provided me with plenty of ideas and comments during the elaboration of this report. I am very grateful for his participation as well.

Finally, I express my deep appreciation to my wife for standing by my side throughout my graduate career. She has always provided me with motivation and energy to go through life. I also thank God for this opportunity and for all good things He has given me.

CONTENTS

ABSTRACT

This report provides an understanding of Zellner's (1962) seemingly unrelated regressions (SUR) procedure, a discussion of some of SUR current issues, and an application of SUR in stratified sampling. A survey of Mexican household meat consumption and expenditures was used in the empirical application. One general model and several individual models were estimated by the SUR procedure incorporating sampling weights. Parameter estimates are reported and its standard errors are approximated by using the bootstrap.

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

In 1962, Arnold Zellner presented a method of estimating parameters of a set of regression equations. Zellner (1962) called his method seemingly unrelated regressions (SUR) to reflect the fact that each equation is related to one another, even though they seem to be unrelated. The regression equations seem to be unrelated because unlike a simultaneous equations model where one or more of the independent variables in one or more of the equations is itself a dependent variable, in seemingly unrelated regression none of the variables in the system is simultaneously both independent variable and dependent variable (Srivastava and Giles, 1987, p. 1). Therefore, even though none of the variables in the system is simultaneously both independent variable and dependent variable, "[t]here may still be interaction between the individual equations if the random disturbances associated with at least some of the different equations are correlated with each other" (Srivastava and Giles, 1987, p. 1). In simple words, the terminology arises because estimating each equation separately will ignore possible relationships of the equation errors (Griffiths et al., 1992, p. 549) whereas estimating each equation into one model will consider possible relationships of the equation errors even though the equations seem to be unrelated.

Zellner (1962) found that the regression coefficient estimators are at least asymptotically more efficient than the least squares equation-by-equation estimators. In particular, a quite large gain in efficiency can be obtained when independent variables in different equations are not highly correlated and when error terms in different equations are highly correlated. Additionally, Zellner (1962) explained a test for the equality of all regression equation coefficients (also called test for aggregation bias). His test can be used to determine if aggregated data (macro-data) has an aggregation bias problem or if disaggregated data (micro-data) can be aggregated without suffering from aggregation bias.[1]

Zellner (1962) provided an empirical example to illustrate his SUR method of estimating parameters. This example was later further discussed by Kmenta (1971, pp. 527–528) and Theil (1971, pp. 295–302). Since Zellner (1962) presented his SUR model, substantial literature has emerged. For example, let's consider only the literature on SUR studied by Zellner. Zellner and Hwang (1962) discuss further properties of efficient estimators for SUR equations. Zellner (1963) discusses finite sample properties in estimators for SUR equations. Zellner (1971, pp. 244-246) restudies the error correlations of the regression equations from a Bayesian point of view. Zellner (1969), and Zellner and Montmarquette (1971) revisit the aggregation problem.

Literature dealing with SUR with unequal number of observations has also developed. This topic expands on the main topic discussed in this research. Studies in this topic have illustrated how to handle a set of regression equations when the data is time-series, cross-sectional or panel data. In cross-sectional data, there seems to be more examples of cases when observations from one equation with respect to another equation are missing. For instance, we can encounter the typical nonresponse due to participants refusing to answer or we can encounter missing observations due to the nature of the survey. In the latter case, there might be observations missing because they were censored[2] or because people collecting information were time constrained[3] or some kind of combination of both.[4]

---

[1]In general, aggregated data is a function of disaggregated data. Aggregation is the process of going from disaggregated data to aggregated data. Hence, aggregation bias occurs when noise is gained during the process of aggregation in such a way that any inference made from aggregated data is biased. Section 2.4.1 provides several examples on how data can be aggregated.

[2]For instance, Wooldridge's (2006, p. 610) example presented in Section 3.1 where we know the value of a family's wealth up to a certain threshold.

[3]For example, the design of the ENIGH 2006 survey, which was mentioned in Section 3.3, where during the week of the interview not all possible consumption items will be purchased by the households.

[4]Consider again the case when not all consumption items will be purchased by households. Some items are not purchased because they are too expensive and the households choose not to buy them,

Other studies dealing with SUR with unequal number of observations discuss alternative estimators of the variance-covariance matrix of the error term ($\mathbf{\Sigma}$), the conditions under which one estimator of $\mathbf{\Sigma}$ will perform better than another, and whether it is relevant to use better estimates of $\mathbf{\Sigma}$. In the latter case, it has been found that better estimates of $\mathbf{\Sigma}$ or $\mathbf{\Sigma}^{-1}$ need not imply better estimates of regression coefficients. In addition, as it will be discussed in Section 3.3, a feasible GLS estimator of the regression coefficients that ignores the extra observations in estimating $\mathbf{\Sigma}$ (but not necessarily in estimating $\mathbf{\Sigma}^{-1}$ or $\boldsymbol{\beta}$) compares favorably to a feasible GLS estimator of the regression coefficients that seem to use all extra observations.

Given that SUR have been widely accepted and implemented, and the abundant literature that has emerged, the general objective of this research is to provide an understanding of the SUR procedure and to explain some of its current trends. The specific objectives of this study are:

- explain Zellner's (1962) SUR procedure and why it is preferred over the least squares equation-by-equation,
- explain Zellner's (1962) test for equality of regression coefficients (also called test for aggregation bias),
- provide an empirical application of a SUR model, and
- explain the relevant findings from this empirical application.

The Mexican household meat consumption was selected as the empirical application in this research. In order to familiarize with meat and its world market, Section 1.1 discusses the role meat plays in the agricultural sector, Section 1.2 talks about the meat world market, and Section 1.3 explains the importance of analyzing meat

but some items are not purchased because households did not have the chance to buy them during the week of the interview. In the former case, we do not have a measure of the maximum amount households would have been willing to pay as explained by Pindyck and Rubinfeld (1997, p. 325) in Section 3.1. In the latter case, the item was simply not recorded because of time constraints.

at the table-cut level. Then, Section 1.4 will expand on understanding Mexican meat production and consumption. Finally, Section 1.5 will briefly talk about the data that will be employed in the study.

1.1 Meat and the Agricultural Sector

The importance of agriculture in an economy varies significantly by country. For example, in 2006, the nominal gross domestic product of the agriculture, forestry, fishing and hunting sector of the United States was $122 billion (Northeast-Midwest Institute, based on data from U.S. Department of Commerce, Bureau of Economic Analysis). However, the total nominal gross domestic product in 2006 was $13.1947 trillion (International Monetary Fund–World Economic Outlook Database). Therefore, during 2006, in the United States, the agriculture, forestry, fishing and hunting sector contributed only about 1% of the total nominal gross domestic product (Figure 1.1). On the other hand, in 2005, the total nominal gross domestic product of Mexico in 2005 was $0.76769 trillion (International Monetary Fund–World Economic Outlook Database). However, during 2005, in Mexico, the agribusiness sector contributed 5% of the total nominal gross domestic produduct (The World Bank and International Monetary Fund, November, 2006) (Figure 1.2).

Meat plays an important role in world trade. According to Dyck and Nelson (2003), global meat trade is over 24 million tons with a value over $43 billion in 2000, which is about 10% of total agricultural trade. Additionally, global meat trade is growing rapidly. From 1990 to 2000, global meat trade grew by aout 6% per year (Dyck and Nelson, 2003).

Additionally, it is important to mention that globalization has lead to dependence of one country's meat consumption on another country's meat production. For instance, East Asia–defined as Japan, South Korea, and Taiwan—is usually the world's largest meat-importing region because the region is densely populated, with mountains and forests that limit the land available for agriculture, making large-scale feed production relatively expensive (Dyck and Nelson, 2003). Furthermore, the region

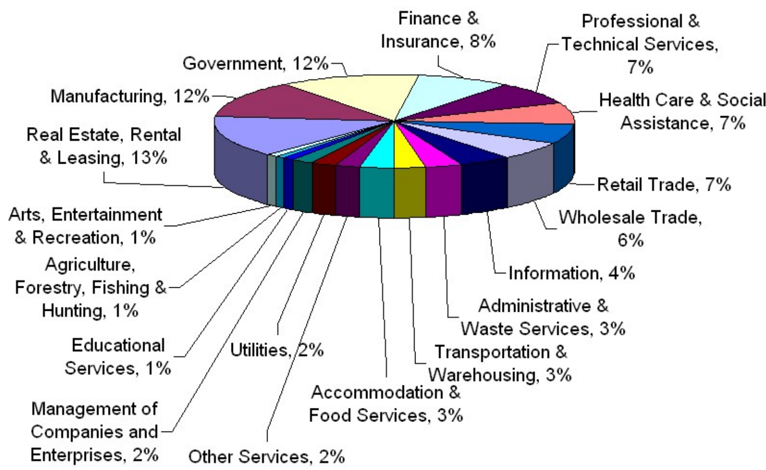

Figure 1.1: U.S. Nominal GDP Contribution by Sector in 2006.

Note: All major sectors are from the North American Industry Classification System. Since data reported by the Northeast-Midwest Institute excluded the mining and construction sector, the U.S. total nominal GDP reported by the World Economic Outlook Database of the International Monetary Fund was used instead. Hence, the nominal GDP share of the mining and construction sectors together is about 7%.

Source: Northeast-Midwest Institute who based its calculations on data from U.S. Department of Commerce, Bureau of Economic Analysis. URL http://www.nemw.org/gdp1.htm (Accessed on May 12, 2008). Pie chart computed by author.

has also relatively high labor costs, and locating large-scale farms and processing plants is sometimes difficult because of pollution concerns and land costs (Dyck and Nelson, 2003). On the other hand, the United States, for example, has abundant grains, meals, grass, forage, a large domestic market, and access to several large foreign markets (Dyck and Nelson, 2003).

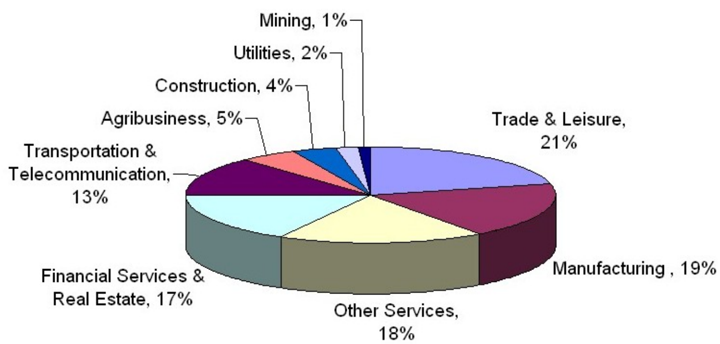

Figure 1.2: Mexican Nominal GDP Contribution by Sector in 2005.

Source: The World Bank and International Monetary Fund—Mexico: Financial Assessment Program Update—Technical Note—Financing of the Private Sector, November 2006, p. 21. URL http://www.imf.org/external/pubs/ft/scr/2007/cr07170.pdf (Accessed on May 12, 2008). Pie chart computed by author.

1.2 Mexico in the World Market

This section starts explaining how a country can have a competitive advantage in producing meat (Section 1.2.1.1). In particular, it briefly discusses key production inputs such as feed input costs, labor costs and capital costs, and the importance of capital investments. Then, Section 1.2.1.2 through Section 1.2.3 explain the world market trends, the relative importance of each meat in the world market and the major players in the world with respect to production, consumption, imports and exports. The discussion of Section 1.2.1.2 through Section 1.2.3 was based on the online data provided by the Production, Supply and Distribution (PSD) of the Economic Research Service (ERS) of the United States Department of Agriculture (USDA). All the charts and tables reported in those sections were computed by the author by using such database. The world total amounts reported by the USDA-ERS-PSD database does not include all countries in the real world but rather a list of countries which represents over 90% of real world total amounts. Furthermore, in order for the USDA-ERS-PSD list of countries to appropriately represent the major players, the list is updated periodically. The list of countries in the USDA-ERS-PSD database is an efficient forecasting basis for identifying world trends. Beef and pork quantities are reported in metric tons (MT) and in carcass weight equivalent (CWE). CWE is the weight of an animal after slaughter and removal of most internal organs, head, and skin. Poultry meat quantities are reported in metric tons (MT) and ready to cook equivalent. In Section 1.2.1.2 through Section 1.2.3, beef includes beef and veal meat while poultry meat only includes broiler meat (it does not include turkey meat).

Additionally, the reader will notice that the author refers to a world region as a country. For instance, the European Union (25 countries) will be referred to as one country. This was done to facilitate the flow of the discussion. It should be noticed that during the period under consideration (1997-2006) not all 25 European Union countries were part of the European Union (EU). For example, according to the Microsoft Encarta Online Encyclopedia (2008), in 1995 Australia, Finland and Sweden joined the European Union bringing the total number of nations to 15.

Therefore, starting in 1996 the EU was known as EU-15. However, in may 2004, 10 more countries (Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, and Slovenia) were added, bringing the total number of nations to 25. Then, in January 2007 two more countries were added (Romania and Bulgaria), bringing the total number of nation to 27. For the period under consideration in Section 1.2.1.2 through 1.2.3, this study assumed that the EU consisted of 25 countries since 1996. This implies that those countries that were added to the EU in 2004, if they appeared in the USDA-ERS-PSD database, were added to the total EU-15 to compute a new total EU-25.

1.2.1 Production

Before starting any discussion about the leading meat producing countries, it is more appropriate to explain why some countries may produce more meat than other countries based on reasons other than the country's size. Therefore, as explained before, this section will first discuss meat production competitiveness and then continue to explain meat production in the world market context. Even though production competitiveness may be one of the reasons why a country will be a leading exporter, it should be kept in mind that there are other reason such as consumer preferences for meat cuts that will also make a leading exporter be a leading importer.

1.2.1.1 Production Competitiveness

The competitive advantage of producing meat in a country depends on its costs of production. In the world supply chain, the key inputs in producing meat are feed, labor, and capital. These key inputs have an effect on the production of meat. Additionally, they are all negatively related with the production of meat. That is, an increase in any of these costs will negatively impact the production of meat.

For a livestock farmer the cost of feed input depends on the cost of growing, processing, transporting and storing the feeds. The closer the livestock farmer is to the feed input, the lower the transportation costs. Countries that have abundant grassland and feedgrains such as corn, sorghum and oilseed meals such as soymeal will have lower feed input costs and transportation costs. Lower feed input costs and transportation costs will benefit the production of meats and depending on demand conditions they might influence the price of meat to go down and exports to go up.

Lower labor costs either through low wages or economies of scale benefit the production of meats. Labor costs are incurred in different phases: farming, slaughtering, processing and distribution. Labor costs vary across countries depending on the demand of labor and the availability of the workers with the required skills. For example, a country with a relatively small agricultural sector and high unemployment will tend to have very low wages in farming and in the livestock industry. Depending on meat demand conditions, lower labor costs might influence the price of meat to decrease and exports to increase.

Finally, similar to lower input costs and labor costs, lower capital costs benefit the production of meat. Capital costs vary in different stages of the production cycle: livestock farming, meat slaughtering, processing meat, and distributing meat. Capital costs might also vary within a production stage. For instance, livestock farming requires financing in different activities: housing, efficient feeding and cleaning systems, environmental controls, and monitoring systems. Other production stages such as meat slaughter, processing and distribution require even larger capital investments.

As important as having relatively low capital costs, is having access to financing. For example, the United States has low feed input costs, relatively high labor costs, and abundant capital investments. Then, the low U.S. input costs and economies of scales significantly offset the U.S. relatively high labor costs. Mexico, on the other hand, has higher feed input costs, low labor costs, but less capital investments than the United States. As a consequence, the United States is among the leading meat exporting nations while Mexico is among the leading meat importers.

1.2.1.2 World Market

World meat production increased 28% from 1997 to 2006 (see Table 1.1). Swine has the largest world production with an average share of 46%. It is followed by poultry meat with an average share of 27% and beef with 28%. Swine production is experiencing an increasing tendency. It went from 74,361,000 MT in 1997 to 99,776,000 MT in 2006, which is a 34% increase. For the period 1997-1999 beef production was greater than poultry meat; however, since 2000 poultry meat production has been greater than beef. Nonetheless, both meats have an increasing tendency. Poultry meat production went from 43,216,000 MT in 1997 to 60,090,000 MT in 2006, which is a 39% increase. Beef went from 49,237,000 MT in 1996 to 53,511,000 MT in 2006, which is a 9% increase.

The world's largest beef producing countries are the United States, the European Union, Brazil, China, Argentina, Australia, Mexico, India, Russia, and Canada (Figure 1.3). Together these ten countries produce 89% of the total world beef production.

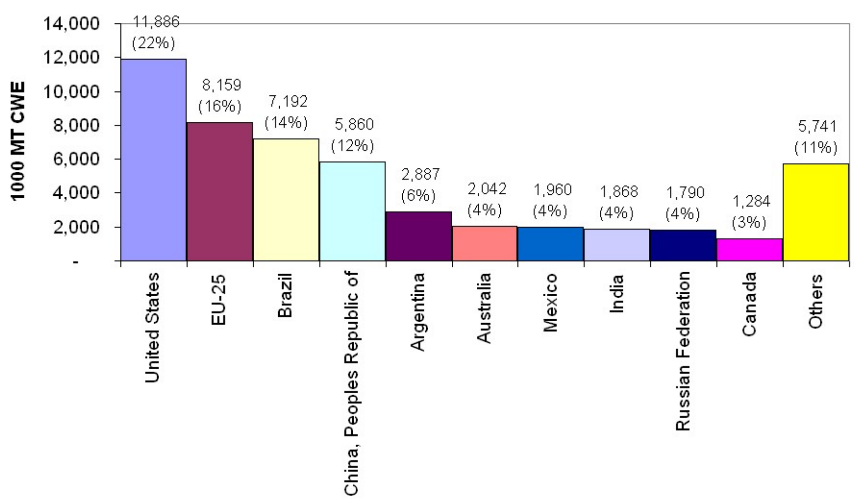

Figure 1.3: World's Largest Beef Producing Countries, Average 1997-2006.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

The world's largest pork producing countries are Republic of China, the European Union, the United States, Brazil, Canada, Russia, Japan, Mexico, Philippines, and

Republic of Korea (Figure 1.4). Together these ten countries account for 95% of the total world pork production.

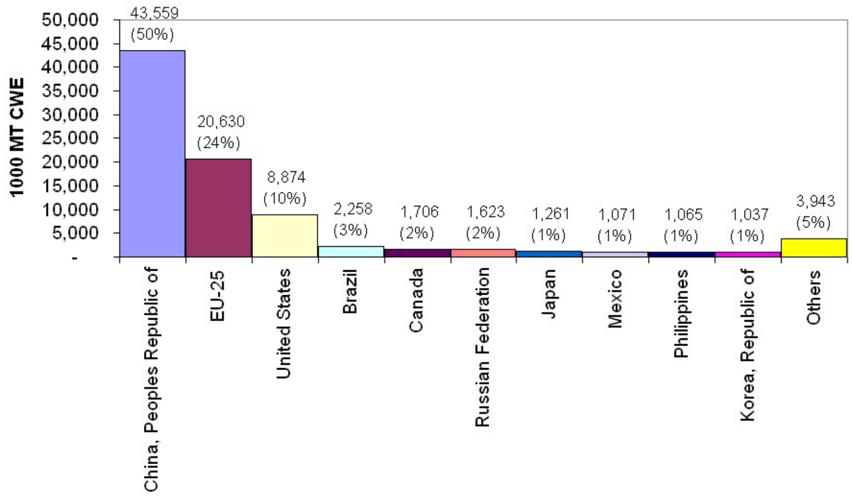

Figure 1.4: World's Largest Pork Producing Countries, Average 1997-2006.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

In the world market of poultry meat, the ten largest producing countries are the United States, Republic of China, the European Union, Brazil, Mexico, Russia, Japan, Thailand, Canada, and Argentina (Figure 1.5). Together these ten countries account for 87% of world total poultry meat production.

Notice that Argentina, Australia, and India, who are among the top ten leading beef producing countries (Figure 1.3), are not among the top ten leading pork producing countries (Figure 1.4). Similarly, Japan, Philippines, and Korea, who are among the top ten leading pork producing countries, are not among the top ten leading beef producing countries (Figure 1.3). Furthermore, comparing Figure 1.3 with Figure 1.5, it can be seen that Australia and Russia in the top ten leading beef producing countries group are replaced by Japan and Thailand in the top ten leading poultry meat producing countries group. Similarly, comparing Figure 1.4 with 1.5 Russia, Philippines, and Korea in the top ten leading pork producing group are replaced by India, Thailand and Argentina in the top ten leading poultry producing countries

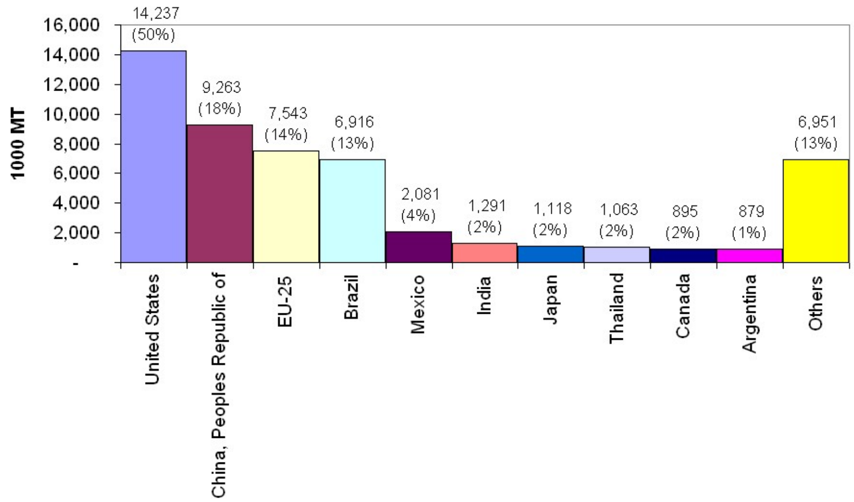

Figure 1.5: World's Largest Poultry Meat Producing Countries, Average 1997-2006.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

group. However, countries such as the United States, the European Union, Brazil, China, Mexico and Canada are leading producing countries in the three types of meats. Therefore, if we consider the combined production of beef, pork, and poultry meat, on average for the period 1997-2006, China produced 58.682 million MT, EU-25 36.332 million MT, United States 34.997 million MT, Brazil 16.366 million MT, Mexico 5.112 million MT, and Canada 3.885 million MT.

1.2.2 Consumption

World meat consumption increased 27% from 1997 to 2006 (Table I.1). Pork has the largest world consumption. Beef, swine and poultry meat has a world consumption share of 27%, 46% and 27% respectively. From 1997 to 2006, swine consumption experienced the largest increase, it went from 74,097,000 MT in 1997 to 98,914,000 MT in 2006 (46% increase). It is followed by poultry meat consumption, which increased from 42,785,000 MT in 1997 to 58,888,000 MT in 2006 (27% increase). Beef, with the smallest increase (26%), went from 48,275,000 MT in 1997 to 51,509,000 MT in 2006.

The world's largest beef consuming countries are the United States, the European Union, Brazil, China, Russia, Argentina, Mexico, India, Japan and Canada (Figure 1.6). Together these countries account for 88% of total world beef consumption. Compared to Figure 1.3, Australia which was the sixth largest beef producing country is not within the ten largest consuming countries; instead, Japan joined the group. However, the United States, EU-25, Brazil, and China has kept their leading top four positions with all of them except the United States producing more than consuming. Similarly, Rusia and Mexico consume more than what they produce; and Argentina, India, and Canada produce more than what they consume.

With respect to pork consumption, the world largest consuming countries are China, the European Union, the United States, Japan, Russia, Brazil, Mexico, Korea, Philippines, and Canada (Figure 1.7). Comparing the largest producing countries (Figure 1.4) with the largest consuming countries (Figure 1.7), we observe the countries are the same but only China, EU-25, and the United States have kept their leading top three positions with all of them producing more than what they consume. Similarly, Japan, Russia, Mexico, Korea, and Philippines consume more than what they produce; and Brazil and Canada produce more than what they consume.

Finally, the world's largest consuming countries of poultry meat are the United States, China, European Union, Brazil, Mexico, Japan, Russia, India, Canada, and Argentina (Figure 1.8). Compared to Figure 1.5, Thailand which was the eighth

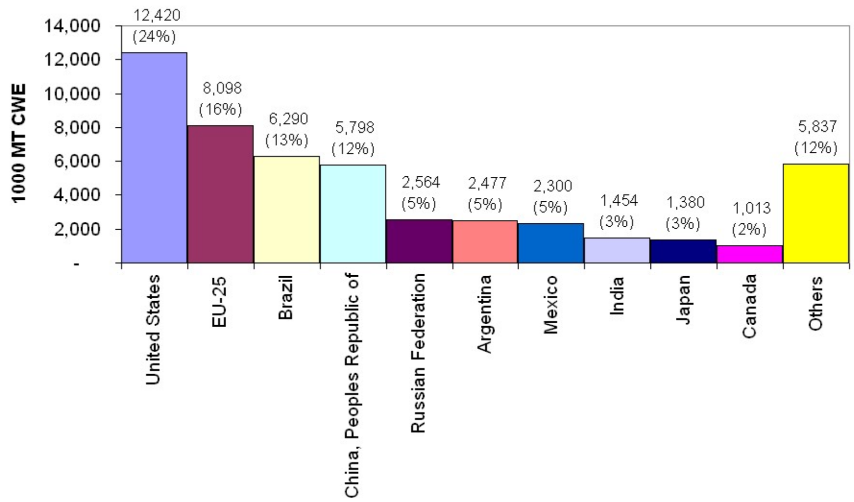

Figure 1.6: World's Largest Beef Consuming Countries, Average 1997-2006.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

largest poultry meat producing country is not within the ten largest consuming countries; instead, Russia joined the group. However, the United States, China, EU-25, Brazil, and Mexico kept their top five leading positions with the United States, EU-25, and Brazil producing more poultry meat than what they consume; and China and Mexico consuming more than what they produce. Similarly, Japan and Canada consume more than what they produce; and India and Argentina slightly producing more than what they consume.

Additionally, notice that Argentina and India, who are among the top ten leading beef consuming countries (Figure 1.6), are not among the top ten leading pork consuming countries (Figure 1.7). Similarly, Korea and Philippines, who are among the top ten leading pork consuming countries (Figure 1.7), are not among the top ten leading beef consuming countries (Figure 1.6). Furthermore, all countries who are the top ten leading beef consuming countries group (Figure 1.6) are also the top ten leading poultry meat consuming countries group (Figure 1.8). Finally, countries such as the United States, EU-25, Brazil, China, Russia, Mexico, Japan, and Canada are leading consuming countries in the three types of meat. Therefore, if we consider

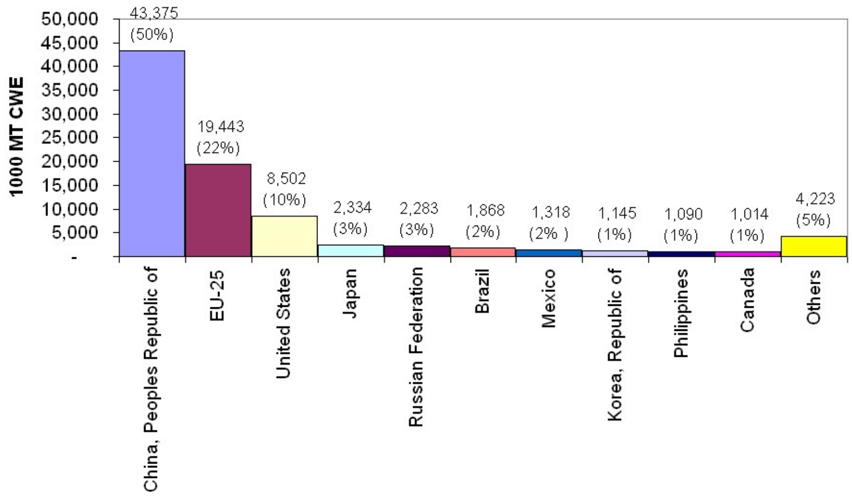

Figure 1.7: World's Largest Pork Consuming Countries, Average 1997-2006.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

the combined consumption of beef, pork, and poultry meat, on average for the period 1997-2006, China consumed 58.464 million MT, EU-25 34.668 million MT, United States 32.943 million MT, Brazil 13.558 million MT, Russia 6.494 million MT, Mexico 5.971 million MT, Japan 5.504 million MT, and Canada 2.929 million MT.

Finally, comparing the combined production with the combined consumption of beef, pork, and poultry meat, on average for the period 1997-2006, Mexico is a net meat consumer with excess consumption of 0.859 million MT while Brazil, United States, EU-25, Canada, and China are net meat producers with excess production of 2.808, 2.054, 1.664, 0.956, and 0.218 million MT respectively. Therefore, Mexico is a very important market for all net meat producers.

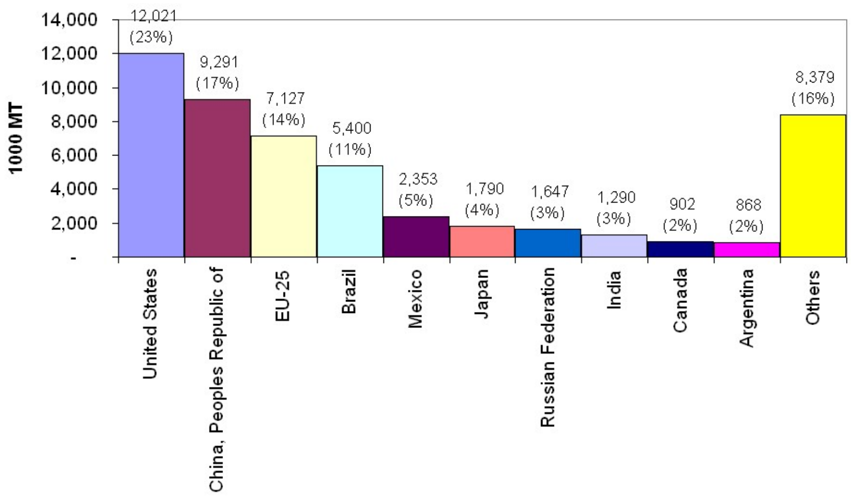

Figure 1.8: World's Largest Poultry Meat Consuming Countries, Average 1997-2006.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

Table 1.1: World Production, Imports, Exports and Consumption by Meat Type.

	1997	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**Average 1997-06**
World Production (1000 MT CWE)[a]											
Beef[b]	49,237	48,958	49,977	50,311	49,646	51,241	50,095	51,327	52,374	53,511	50,668
Pork[c]	74,361	79,574	82,266	81,819	84,940	88,011	90,488	92,801	96,227	99,776	87,026
Poultry[d]	43,216	44,903	47,904	50,474	52,303	54,155	54,282	55,952	59,092	60,090	52,237
Total Meat	166,814	173,435	180,147	182,604	186,889	193,407	194,865	200,080	207,693	213,377	189,931
World Imports (1000 MT CWE)											
Beef[b]	5,016	4,771	5,065	4,935	4,978	5,242	5,074	4,891	5,423	5,007	5,040
Pork[c]	2,587	2,658	2,891	2,960	3,195	3,871	3,962	4,172	4,342	4,232	3,487
Poultry[d]	3,597	3,440	3,844	3,823	4,149	4,443	4,625	4,384	5,063	5,168	4,254
Total Meat	11,200	10,869	11,800	11,718	12,322	13,556	13,661	13,447	14,828	14,407	12,781
World Exports (1000 MT CWE)											
Beef[b]	5,795	5,439	5,724	5,746	5,670	6,274	6,339	6,496	7,092	6,996	6,157
Pork[c]	1,620	1,697	1,636	1,735	2,267	2,757	3,051	3,418	3,875	3,800	2,586
Poultry[d]	4,059	4,196	4,415	4,808	5,526	5,702	6,023	6,055	6,791	6,470	5,405
Total Meat	11,474	11,332	11,775	12,289	13,463	14,733	15,413	15,969	17,758	17,266	14,147
World Domestic Consumption (1000 MT CWE)											
Beef	48,275	48,496	49,818	49,536	48,716	50,277	49,049	49,874	50,770	51,509	49,632
Pork	74,097	79,345	81,908	81,461	84,727	87,829	90,297	92,139	95,236	98,914	86,595
Poultry	42,785	44,120	47,306	49,454	50,855	52,846	52,903	54,172	57,339	58,888	51,067
Total Meat	165,157	171,961	179,032	180,451	184,298	190,952	192,249	196,185	203,345	209,311	187,294

a. MT = metric tons and CWE = Carcass Weight Equivalent. CWE is the weight of an animal after slaughter and removal of most internal organs, head, and skin. CWE applies only to beef and pork, poultry meat is reported by the USDA-ERS-PSD database in ready to cook equivalent.
b. Beef includes beef and veal meat.
c. Pork is also called swine meat.
d. Poultry includes broiler meat and excludes turkey meat.
Note: The amounts reported in this table reflect only those countries that make up the USDA-ERS-PSD database and not all countries in the world. Any production, import, export or consumption amount represent the most important players in the world meat PSD situation, which represents over 90% of the world's situation. In addition, the list of countries that make up the USDA-ERS-PSD database changes periodically.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

1.2.3 Imports and Exports

World meat exports increased 50% from 1997 to 2006 (Table 1.1). Beef has the largest amount exported. Beef, swine and poultry meat has a world export share of 44%, 18%, and 38% respectively. From 1997 to 2006, swine exports experienced the largest increase, going from 1,620,000 MT in 1997 to 3,800,000 MT in 2006 (135% increase). It is followed by poultry meat exports, which increased from 4,059,000 MT in 1997 to 6,470,000 MT in 2006 (59% increase). Beef, with the smallest increase (21%), went from 5,795,000 MT in 1997 to 6,996,000 MT in 2006.

In general, world meat imports experience the same trend as exports. World meat imports increased 29% from 1997 to 2006 (Table I.1). Beef still has the largest amount exported. Beef, swine and poultry meat has a world export share of 39%, 27%, and 33% respectively. From 1997 to 2006, swine imports experienced the largest increase, as they went from 2,587,000 MT in 1997 to 3,487,000 MT in 2006 (64% increase). It is followed by poultry meat imports, which increased from 3,597,000 MT in 1997 to 5,168,000 MT in 2006 (44% increase). Beef imports in 2006 remained at almost the same level that in 1997, 5,007,000 MT.

According to the USDA-ERS-PSD online database, the ten largest importers of beef are the United States (28%), Japan (17%), Russia (15%), European Union (9%), Mexico (7%), Korea (5%), Canada (5%), Egypt (4%), Philippines (2%), and Taiwan (2%) (Figure 1.9). The largest exporters of beef are Australia (22%), Brazil (16%), United States (14%), the European Union (9%), New Zealand (8%), Canada (8%), Argentina (7%), India (7%), Uruguay (5%), and Ukraine (2%) (Figure 1.10). Comparing these last two figures, notice that countries such as the United States, EU-25, and Canada are both among the top ten leading beef exporters and beef importers. However, the United States is a net beef importer while EU-25 and Canada are net beef exporters.

Analyzing pork, the ten largest importing countries are Japan (31%), Russia (19%), United States (12%), Mexico (9%), Hong Kong (8%), Korea (5%), China (4%), Romania (3%), Canada (3%), Australia (1%), and the European Union (1%) (Figure

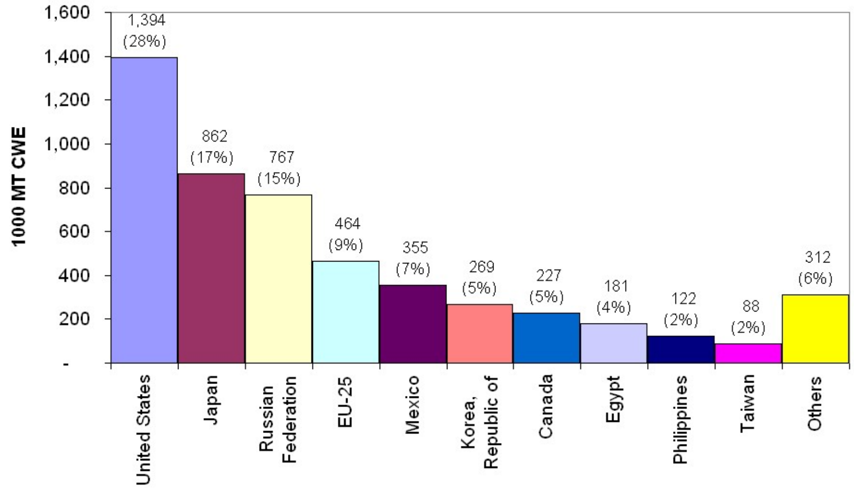

Figure 1.9: World Largest Beef Importing Countries, Average 1997-2006.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

1.11). The ten largest pork exporting countries are the United States (50%), Canada (18%), Brazil (15%), China (12%), Mexico (2%), Chile (2%), Australia (2%), EU (2%), Korea (2%), and Russia (1%) (Figure 1.12). In these last two figures, countries such as Russia, United States, Mexico, Korea, China, Canada, Australia and EU-25 are both among the top ten leading pork importers and pork exporters. However, the United States, Canda, and China are net pork exporters while Russia, Mexico, Korea, Australia, and EU-25 are net pork importers.

Finally, analyzing chicken, the ten largest importing countries are Russia (27%), Japan (16%), China (9%), Saudi Arabia (9%), European Union (9%), Mexico (6%), Hong Kong (5%), United Arab Emirates (3%), South Africa (3%), and Ukraine (2%) (Figure 1.13). In the exports side, these countries are Uited States (41.1%), Brazil, (28.1%) European Union (14.5%), China (6.9%), Thailand (5.7%), Canada (1.3%), Argentina (0.6%), United Arab Emirates (0.4%), Saudi Arabia (0.3%), and Kuwait (0.2%) (Figure 1.14). Comparing these last two figures, countries such as China, Saudi Arabia, EU-25, and the United Arab Emirates are both among the top ten leading poultry meat importers and poultry meat exporters. However, EU-25 is a net

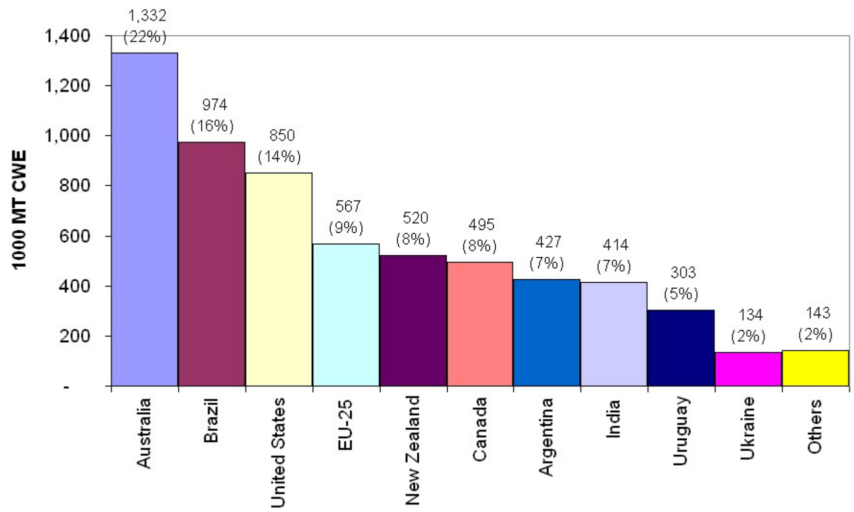

Figure 1.10: World's Largest Beef Exporting Countries, Average 1997-2006.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

poultry meat exporter while China, Saudi Arabia, and the United Arab Emirates are net poultry meat importers.

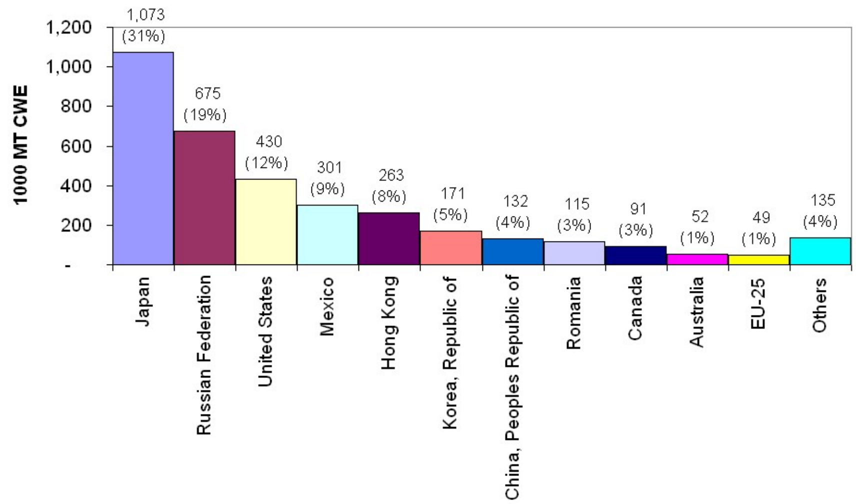

Figure 1.11: World's Largest Pork Importing Countries, Average 1997-2006.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

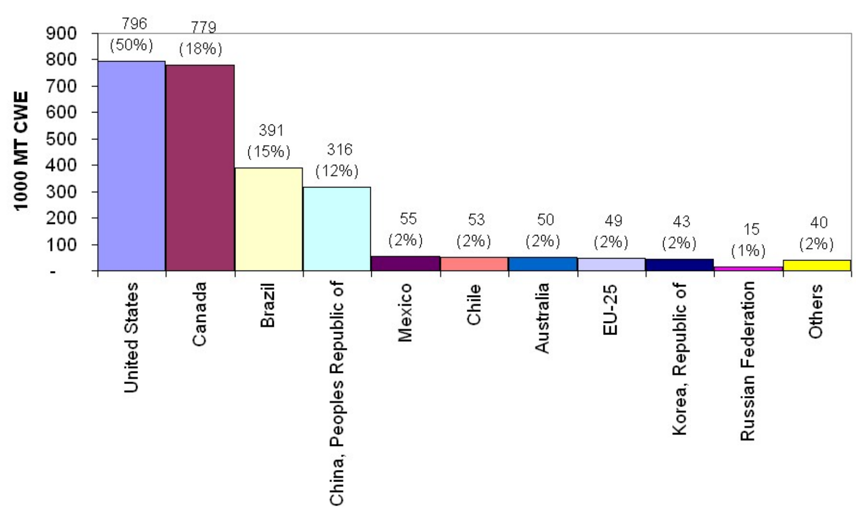

Figure 1.12: World's Largest Pork Exporting Countries, Average 1997-2006.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

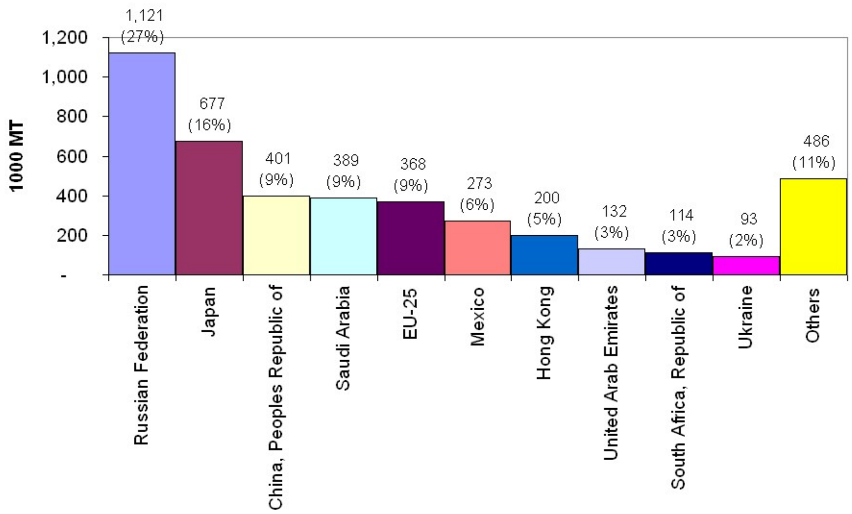

Figure 1.13: World's Largest Poultry Meat Importing Countries, Average 1997-2006.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

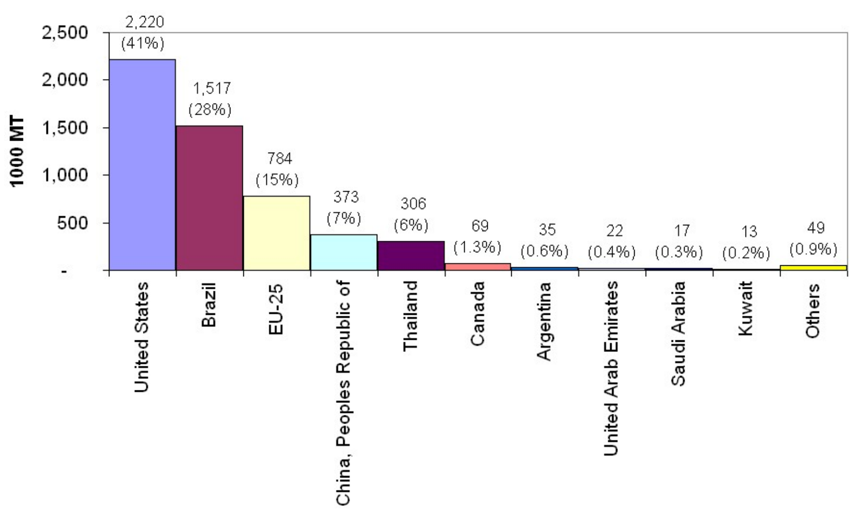

Figure 1.14: World's Largest Poultry Meat Exporting Countries, Average 1997-2006.

Source: USDA-ERS-PSD Online Database, computed by author. Accessed on March 17, 2007.

1.3 Analyzing Meat at the Table-Cut Level

Growing populations, rising incomes, and increased urbanization have contributed to an increased in global demand for meat. Countries with good resources in the production of meat such as abundant feed grains and grassland will tend to export meat and countries with good resources in the production of other goods but scare resources in the production of meat will import meat. However, in Section 1.2.3 we have seen that many countries are both meat importers are exporters, this finding points to the importance of trade in the form of cuts.

As a matter of fact, most meat trade is in the form of cuts (Dyck and Nelson, 2003). Demand for the parts varies considerably across countries, depending on consumer tastes and preferences, whether cuts can be substituted for one another, and other factors. The largest meat producing companies will look across international markets for the consumers with the highest willingness to pay. The ability to match meat cuts with the highest paying markets will allow firms to increase the aggregate value of each animal. Therefore, any study on meat demand should attempt to analyze meat at the table-cut level to better understand the demand for meat.

1.4 Mexican Meat Production and Consumption

Meat is produced in Mexico in all its national territory despite the different environmental and climatic regions of the country (Figure 1.15). In 1999, the ten largest meat producing states in Mexico in decreasing order were Jalisco, Veracruz, Guanajuato, Puebla, Sonora, México, Yucatán, Querétaro, Durango, Chiapas and Michoacán (Table 1.2). However, in 2002 the ten largest producing states became Jalisco, Veracruz, Sonora, Puebla, Guanajuato, Querétaro, Durango, México, Yucatán, Nuevo León and Coahuila (Table 1.2).

Meat is produced with different technologies. The technology employed ranges from high technology and integrated industries to very basic techniques used by lower class farmers. For firms with high technology, meat production represents a form of wealth accumulation for its owners. On the other hand, for people with very basic

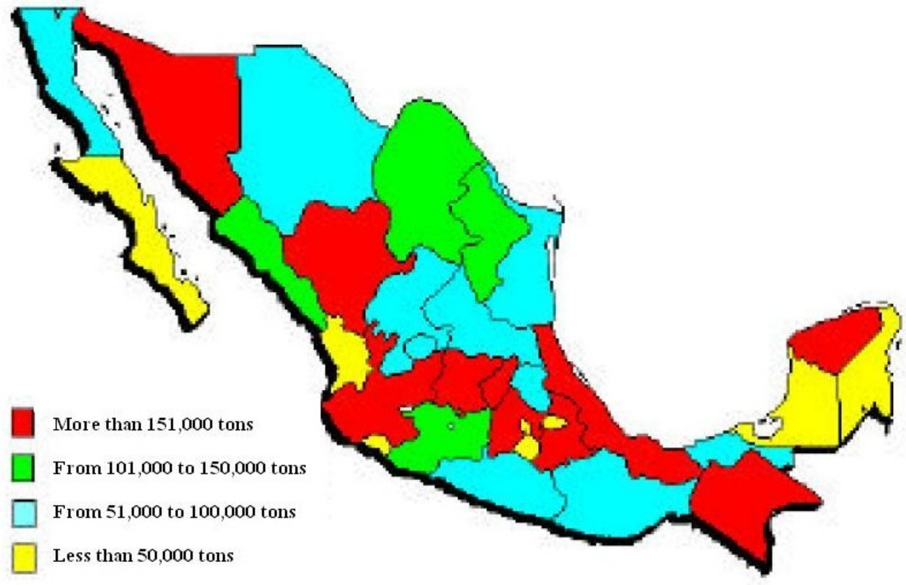

Figure 1.15: Geographical Distribution of Mexican Meat Production in 1999.
Source: Marín et al. (2000, p. 17).

techniques, meat production represents an activity that allows farmers for subsistence.

According to Marín et al. (2000), Mexican infrastructure of animal slaughter facilities is classified in three groups according to the technology being implemented. The first group is formed by those facilities that have an up-to-date technology. In this group the quality and sanitary standards are inspected by a government agency, a Federal Type Inspection ("Tipo Inspección Federal, TIP"). The second group is formed by an old-dated technology, which is the most traditional in Mexico. This group has several types of sanitary controls and the quality and sanitary inspections are performed by the Health Department ("Secretaría de Salud"). Finally, the last group is composed by the few facilities that perform an ancestral type of slaughter corresponding the ancestral period. In 1999, there were a total of 87 facilities of the Federal Type Inspection group (first group): 43 corresponding to the slaughter of cows, 31 for the slaughter of pigs, and 13 for chicken. Figure 1.16 shows the geographical distribution of the Federal Type Inspection Group by meat.

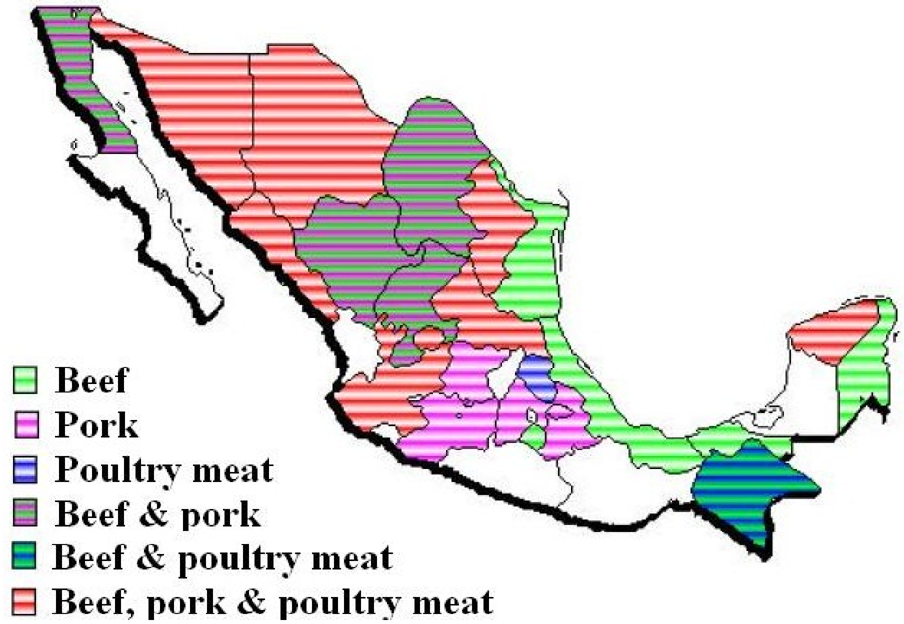

Figure 1.16: Geographical Distribution of the Federal Type Inspection Slaughter Facilities in Mexico in 1999.
Source: Marín et al. (2000, p. 20).

Marín et al. (2000) explain that in the 1990s Mexican meat production was affected by different factors. Climatic changes in the first half of the 1990s combined with droughts lead to a deficit of forage crops. Poor feeding lead to lower quality of slaughter cattle and affected the next generation of feeder cattle. A macroeconomic crisis led to high interest rates, exchange rate depreciation, and high prices of inputs—grains and forage crops. At the microeconomic level, the consumers' purchasing power decreased during this period.

According to Marín et al. (2000), negative factors such as natural phenomenon (climatic changes, droughts, etc.) and economic conditions (peso devaluations, interest rate changes, etc.) do not immediately affect the production of beef. This is because of the planning process in the production of beef and the biological cycles of the different breeds of cattle. According to Marín et al. (2000), negative factors affecting the production of beef in Mexico will have an effect on beef production up to three to four years later. In the case of pork, these negative factors will have an effect on pork production approximately one year later. In the case of chicken the impact of negative factors is observed immediately, it could take only from two to four months to observe the negative effect.

On the other hand, the modification of the agricultural legislation in 1992, the beginning of NAFTA in 1994, and the implementation of the "Alianza para el Campo" program were all oriented to motivate meat production.

The main meats produced in Mexico are beef, pork, and chicken. According to Marín et al. (2000) during the 1970s, beef had the greatest production in Mexico. During the first half of the 1980s, pork had the greatest production, but it was surpass by the beef production in the second half. However, since 1997 chicken has experienced the greatest production (Figure 1.17).

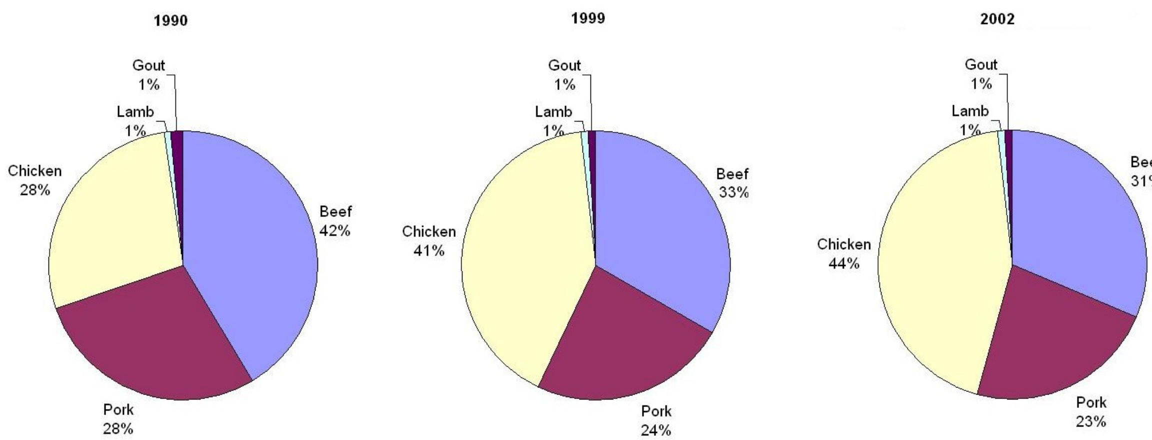

Figure 1.17: Mexican Meat Production by Type.

Source: SIACON-SIAP-SAGARPA, computed by author. Accessed on March 16, 2007.

Table 1.2: The Ten Largest Meat Producing States in Mexico.

1990	**Total Meat Production (MT)**[a]	**Share**	**1999**	**Total Meat Production (MT)**[a]	**Share**	**2002**	**Total Meat Production (MT)**[a]	**Share**
Jalisco	364,693	13.60%	Jalisco	552,098	13.17%	Jalisco	635,426	13.54%
Veracruz	209,229	7.80%	Veracruz	433,194	10.33%	Veracruz	520,233	11.08%
Guanajuato	201,683	7.52%	Guanajuato	264,888	6.32%	Sonora	275,018	5.88%
Sonora	201,478	7.51%	Puebla	263,251	6.28%	Puebla	271,018	5.77%
Puebla	132,435	4.94%	Sonora	253,856	6.05%	Guanajuato	264,773	5.64%
Durango	116,377	4.34%	México	204,116	4.87%	Querétaro	229,213	4.88%
Michoacán	113,949	4.25%	Yucatán	195,535	4.66%	Durango	216,901	4.61%
México	108,929	4.06%	Querétaro	191,918	4.58%	México	196,901	4.19%
Sinaloa	97,804	3.65%	Durango	176,736	4.21%	Yucatán	195,067	4.16%
Chiapas	96,406	3.59%	Chiapas	170,823	4.07%	Nuevo León	177,658	3.78%
Chihuahua	92,694	3.46%	Michoacán	148,542	3.54%	Coahuila	177,477	3.78%
Others	946,817	35.30%	Others	1,228,612	31.92%	Others	1,533,977	32.68%
Total	2,682,494	100.00%	Total	4,193,569	100.00%	Total	4,694,008	100.00%

a. Meat production includes beef, pork, poultry meat, lamb and gout.

Source: SIACON-SIAP-SAGARPA, computed by author. Accessed on March 16, 2007.

1.5 Data

As explained in the beginning of this chapter, the general objective of this research is to provide an understanding of the SUR procedure. Additionally, one of the specific objectives is to provide an empirical application of a SUR model. As an empirical application, this research will study the Mexican meat consumption. In order to study the Mexican meat consumption, this study will employ Mexican data on household income and expenditures. As it will be explained in Chapter IV, this data is published in *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)*. This is nationwide survey published by a Mexican governmental institution.

ENIGH collects data by performing direct interviews through a stratified sampling method. Two instruments are used to collect the data: a questionnaire and a journal. The questionnaire is designed to collect the data concerning the house infrastructure, the members and their household identification, and members' socio-demographic characteristics. In addition, for household members older than 12 years old, the questionnaire will capture occupational activities and related characteristics as well as income and expenditures. On the other hand, the journal is designed to collect at-home and away-from-home expenditures on food, drinks, cigarettes and public transportation. However, food expenditures are recorded for the household unit only. ENIGH also contains information about household incomes, and quantities and prices of goods purchase. However, ENIGH data on food, drinks, cigarettes and public transportation is recorded only when the household made a purchase. Since interviewers collect information from households during the period of one week, those meat cuts that a household did not buy during the week of the interview will not be recorded. Chapter IV will describe what type of information ENIGH contains in more detail.

CHAPTER II
SEEMINGLY UNRELATED REGRESSIONS

Arnold Zellner (1962) derived a method of estimating parameters of a set of regression equations. His method is now widely used and has been generally referred as seemingly unrelated regressions (SUR). The terminology arises because estimating each equation separately will ignore possible relationships of the equations errors (Griffiths et al., 1992, p. 549).

Zellner (1962) found that the regression coefficient estimators are at least asymptotically more efficient than the least squares equation-by-equation estimators. Additionally, Zellner (1962) showed that a quite large gain in efficiency can be obtained when independent variables in different equations are not highly correlated and when error terms in different equations are highly correlated. Finally, Zellner described and showed how to apply a test for the equality of all regression equation coefficients. The test is referred as a test for aggregation bias. This test is used to determine if aggregated data (macro-data) has an aggregation bias problem or if data at the micro-level can be aggregated without suffering from aggregation bias.

2.1 Estimation Procedure

Suppose we are interested in estimating a system of M equations. Each equation contains K_i regression coefficients (parameters), for a total of $\bar{K} = \sum_{i=1}^{M} K_i$. Additionally, the data sample for the dependent and independent variables of each equation consist of T observations.

Let the i^{th} equation be given by

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{u}_i, \quad i = 1, 2, \ldots, M, \tag{2.1}$$

where $\mathbf{y}_i$ is a $(T \times 1)$ vector of observations on the i^{th} dependent variable, $\mathbf{X}_i$ is a $(T \times K_i)$ matrix of observations on the K_i independent variables, $\boldsymbol{\beta_i}$ is a $(K_i \times 1)$ vector of regression coefficients (also called vector of parameters), and $\mathbf{u}_i$ is a $(T \times 1)$

vector of random errors (also called vector of disturbances).[1] We assume $\mathbf{X}_i$ is fixed (i.e., deterministic or non-stochastic) with rank$(\mathbf{X}_i) = K_i$ (i.e., $\mathbf{X}_i$ is of full column rank), $E(\mathbf{u}_i) = 0$, and $\lim_{T\to\infty}\left(\frac{1}{T}\mathbf{X}_i'\mathbf{X}_j\right) = \mathbf{Q}_{ij}$, $i, j = 1, 2, \ldots, M$ where $\mathbf{Q}_{ij}$, $i \neq j$, is a $(K_i \times K_j)$ matrix with finite elements, and $\mathbf{Q}_{ij}$, $i = j$, is non-singular.[2] The latter assumption means $\mathbf{X}_i$ and $\mathbf{X}_j$ depend on T. That is, their sizes change as T changes.

Writing all M equations in equation (2.1) into one model gives

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \\ \vdots \\ \boldsymbol{\beta_M} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_M \end{bmatrix} \tag{2.2}$$

or

$$\underset{(MT\times 1)}{\mathbf{y}} = \underset{(MT\times \bar{K})}{\mathbf{X}} \underset{(\bar{K}\times 1)}{\boldsymbol{\beta}} + \underset{(MT\times 1)}{\mathbf{u}} . \tag{2.3}$$

We assume the errors from different equations in the same "time period" are correlated but the errors from two different equations in different time periods are uncorrelated.[3]

---

[1]We could also write equation (2.1) as

$$\underset{(T\times 1)}{\mathbf{y}_i} = \begin{pmatrix} \mathbf{X}_{i1} & \mathbf{X}_{i2} & \cdots & \mathbf{X}_{iK_i} \end{pmatrix}_{(T\times K_i)} \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{iK_i} \end{pmatrix}_{(K_i\times 1)} + \underset{(T\times 1)}{\mathbf{u}_i}, \qquad i = 1, 2, \ldots, M;$$

or equivalently, $\mathbf{y}_i = \beta_{i1}\mathbf{X}_{i1} + \beta_{i2}\mathbf{X}_{i2} + \cdots + \beta_{iK_i}\mathbf{X}_{iK_i} + \mathbf{u}_i,\ i = 1, 2, \ldots, M$.

If an intercept is desired, add the $(T \times 1)$ vector $\mathbf{X}_{i0} = \mathbf{1}_T$ whose elements are 1, to the $\mathbf{X}_i$ matrix and the parameter β_{i0} to the $\boldsymbol{\beta}_i$ vector increasing their dimensions to $(T \times (K_i+1))$ and $((K_i+1)\times 1)$ respectively. Alternatively, we could treat $\mathbf{X}_{i1}$ as $\mathbf{1}_T$ and keep in mind there are $K_i - 1$ independent variables in the $\mathbf{X}_i$ matrix.

[2]Srivastava and Giles (1987, p. 27) explain that these last two assumptions rule out certain data features, such as the presence of a trend variable.

[3]The use of $t = 1, 2, \ldots, T$ does not necessarily imply time series analysis. It can be applied to cross-sectional data, time series and cross-sectional, and to regression equations in which each equation refers to a particular classification category and the observations refer to different points in space. See Zellner (1962) for a specific example in each of these cases.

That is,

$$E[u_i(t)u_j(s)] = \begin{cases} \sigma_{ij} & \text{if} \quad t = s \\ 0 & \text{if} \quad t \neq s \end{cases} \qquad \text{for } i, j = 1, 2, \ldots, M \text{ and } t, s = 1, 2, \ldots, T.$$

For instance,

$$\begin{aligned}
E[\mathbf{u}_i \mathbf{u}_j'] &= E\left[\begin{pmatrix} u_i(1) \\ u_i(2) \\ \vdots \\ u_i(T) \end{pmatrix} \begin{pmatrix} u_j(1) & u_j(2) & \cdots & u_j(T) \end{pmatrix}\right] \\
&= \begin{bmatrix} E[u_i(1)u_j(1)] & E[u_i(1)u_j(2)] & \cdots & E[u_i(1)u_j(T)] \\ E[u_i(2)u_j(1)] & E[u_i(2)u_j(2)] & \cdots & E[u_i(2)u_j(T)] \\ \vdots & \vdots & \ddots & \vdots \\ E[u_i(T)u_j(1)] & E[u_i(T)u_j(2)] & \cdots & E[u_i(T)u_j(T)] \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{ij} & 0 & \cdots & 0 \\ 0 & \sigma_{ij} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{ij} \end{bmatrix} = \sigma_{ij} \mathbf{I}_T,
\end{aligned}$$

where $\mathbf{I}_T$ is the identity matrix of dimension $(T \times T)$.

Then, the variance-covariance matrix for the disturbance vector is:

$$
\begin{aligned}
\mathbf{W} &= var(\mathbf{u}) = E\left\{(\mathbf{u}-E(\mathbf{u}))(\mathbf{u}-E(\mathbf{u}))'\right\} = E\left\{\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\\ \vdots\\ \mathbf{u}_M\end{bmatrix}\begin{bmatrix}\mathbf{u}_1' & \mathbf{u}_2' & \cdots & \mathbf{u}_M'\end{bmatrix}\right\}\\
&= \begin{bmatrix}E[\mathbf{u}_1\mathbf{u}_1'] & E[\mathbf{u}_1\mathbf{u}_2'] & \cdots & E[\mathbf{u}_1\mathbf{u}_M']\\ E[\mathbf{u}_2\mathbf{u}_1'] & E[\mathbf{u}_2\mathbf{u}_2'] & \cdots & E[\mathbf{u}_2\mathbf{u}_M']\\ \vdots & \vdots & \ddots & \vdots\\ E[\mathbf{u}_M\mathbf{u}_1'] & E[\mathbf{u}_M\mathbf{u}_2'] & \cdots & E[\mathbf{u}_M\mathbf{u}_M']\end{bmatrix} = \begin{bmatrix}\sigma_{11}\mathbf{I}_T & \sigma_{12}\mathbf{I}_T & \cdots & \sigma_{1M}\mathbf{I}_T\\ \sigma_{21}\mathbf{I}_T & \sigma_{22}\mathbf{I}_T & \cdots & \sigma_{2M}\mathbf{I}_T\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{M1}\mathbf{I}_T & \sigma_{M2}\mathbf{I}_T & \cdots & \sigma_{MM}\mathbf{I}_T\end{bmatrix}\\
&= \begin{bmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM}\end{bmatrix} \otimes \mathbf{I}_T = \boldsymbol{\Sigma}_{\boldsymbol{c}} \otimes \mathbf{I}_T,
\end{aligned}
\tag{2.4}
$$

where $\otimes$ (called the Kronecker product) indicates that each element of $\boldsymbol{\Sigma}_{\boldsymbol{c}}$ is multiplied by an identity matrix of dimension $(T \times T)$.

Given the above variance-covariance matrix for the error term in equation (2.3), the appropriate procedure to estimate $\boldsymbol{\beta}$ is the generalized least squares,[4]

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{y}, \tag{2.5}$$

where $\mathbf{W}^{-1}$ is such that there exists a nonsingular $(MT \times MT)$ matrix $\mathbf{H}$ such that $\mathbf{HWH}' = \mathbf{I}_{MT}$.

[4]The purpose of generalized least squares estimation is to estimate $\boldsymbol{\beta}$ in the most efficient possible manner by accounting for the information provided by the knowledge of $\mathbf{W} = E[\mathbf{uu}'] = \boldsymbol{\Sigma}_{\boldsymbol{c}} \otimes \mathbf{I}_T$. The best linear unbiased parameter estimates are obtained if it is possible to transform the original data so that the variance-covariance matrix of the transformed errors equals $\mathbf{I}_{MT}$. Once the data is transformed, application of the Gauss-Markov theorem will provide the best linear unbiased parameter estimates. Assuming that $\mathbf{W}$ is a positive definite matrix guarantees the existence of a nonsingular square matrix $\mathbf{H}$ such that $\mathbf{HWH}' = \mathbf{I}_{MT}$. Hence, the transformation of the data is also guaranteed. For more information about generalized least squares estimation see Aitken (1934-1935) and Telser (1964).

We pre-multiply both sides of equation (2.3) by matrix $\mathbf{H}_{(MT\times MT)}$ to obtain:

$$\mathbf{Hy} = \mathbf{HX}\boldsymbol{\beta} + \mathbf{Hu}$$

or

$$\underset{(MT\times 1)}{\mathbf{y}^*} = \underset{(MT\times \bar{K})}{\mathbf{X}^*} \underset{(\bar{K}\times 1)}{\boldsymbol{\beta}} + \underset{(MT\times 1)}{\mathbf{u}^*}.$$

The system now satisfies the usual properties of the least squares model $E[\mathbf{Hu}] = \mathbf{H}E[\mathbf{u}] = 0$ and $var(\mathbf{Hu}) = E\left\{[\mathbf{Hu} - E(\mathbf{Hu})][\mathbf{Hu} - E(\mathbf{Hu})]'\right\} = E\left\{\mathbf{Huu'H'}\right\} = \mathbf{HWH'} = \mathbf{I}_{MT}$.

Now, the inverse of $\mathbf{W}$ in equation (2.4) is denoted by:

$$\begin{aligned}
\mathbf{W}^{-1} &= var^{-1}(\mathbf{u}) = \begin{bmatrix} \sigma^{11}\mathbf{I}_T & \sigma^{12}\mathbf{I}_T & \cdots & \sigma^{1M}\mathbf{I}_T \\ \sigma^{21}\mathbf{I}_T & \sigma^{22}\mathbf{I}_T & \cdots & \sigma^{2M}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1}\mathbf{I}_T & \sigma^{M2}\mathbf{I}_T & \cdots & \sigma^{MM}\mathbf{I}_T \end{bmatrix} \\
&= \begin{bmatrix} \sigma^{11} & \sigma^{12} & \cdots & \sigma^{1M} \\ \sigma^{21} & \sigma^{22} & \cdots & \sigma^{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1} & \sigma^{M2} & \cdots & \sigma^{MM} \end{bmatrix} \otimes \mathbf{I}_T = \boldsymbol{\Sigma_c}^{-1} \otimes \mathbf{I}_T. \qquad (2.6)
\end{aligned}$$

Hence, $\underset{(\bar{K}\times 1)}{\hat{\boldsymbol{\beta}}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \\ \vdots \\ \hat{\boldsymbol{\beta}}_M \end{bmatrix}$ is given by:

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left\{ \begin{bmatrix} \mathbf{X}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M' \end{bmatrix} \begin{bmatrix} \sigma^{11}\mathbf{I}_T & \sigma^{12}\mathbf{I}_T & \cdots & \sigma^{1M}\mathbf{I}_T \\ \sigma^{21}\mathbf{I}_T & \sigma^{22}\mathbf{I}_T & \cdots & \sigma^{2M}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1}\mathbf{I}_T & \sigma^{M2}\mathbf{I}_T & \cdots & \sigma^{MM}\mathbf{I}_T \end{bmatrix} \right. \\
&\quad \times \left. \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{bmatrix} \right\}^{-1} \begin{bmatrix} \mathbf{X}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M' \end{bmatrix} \\
&\quad \times \begin{bmatrix} \sigma^{11}\mathbf{I}_T & \sigma^{12}\mathbf{I}_T & \cdots & \sigma^{1M}\mathbf{I}_T \\ \sigma^{21}\mathbf{I}_T & \sigma^{22}\mathbf{I}_T & \cdots & \sigma^{2M}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1}\mathbf{I}_T & \sigma^{M2}\mathbf{I}_T & \cdots & \sigma^{MM}\mathbf{I}_T \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} \\
&= \begin{bmatrix} \sigma^{11}\mathbf{X}_1'\mathbf{X}_1 & \sigma^{12}\mathbf{X}_1'\mathbf{X}_2 & \cdots & \sigma^{1M}\mathbf{X}_1'\mathbf{X}_M \\ \sigma^{21}\mathbf{X}_2'\mathbf{X}_1 & \sigma^{22}\mathbf{X}_2'\mathbf{X}_2 & \cdots & \sigma^{2M}\mathbf{X}_2'\mathbf{X}_M \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1}\mathbf{X}_M'\mathbf{X}_1 & \sigma^{M2}\mathbf{X}_M'\mathbf{X}_2 & \cdots & \sigma^{MM}\mathbf{X}_M'\mathbf{X}_M \end{bmatrix}^{-1}_{(\bar{K}\times\bar{K})} \\
&\quad \times \begin{bmatrix} \sum_{i=1}^{M} \sigma^{1i}\mathbf{X}_1'\mathbf{y}_i \\ \sum_{i=1}^{M} \sigma^{2i}\mathbf{X}_2'\mathbf{y}_i \\ \vdots \\ \sum_{i=1}^{M} \sigma^{Mi}\mathbf{X}_M'\mathbf{y}_i \end{bmatrix}_{(\bar{K}\times 1)} \qquad (2.7)
\end{aligned}$$

and is the best linear unbiased estimator.

Now, the variance-covariance matrix of the estimator $\hat{\boldsymbol{\beta}}$ is:

$$\begin{aligned}
var(\hat{\boldsymbol{\beta}}) &= var[(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^*] = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime} var(\mathbf{y}^*)\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} \\
&= (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime} E\left\{[\mathbf{y}^* - E(\mathbf{y}^*)][\mathbf{y}^* - E(\mathbf{y}^*)]'\right\}\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} \\
&= (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime} E\left\{[\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta}][\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta}]'\right\}\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} \\
&= (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime} E\left\{\mathbf{u}^*\mathbf{u}^{*\prime}\right\}\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} \\
&= (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime} E\left\{[\mathbf{u}^* - E(\mathbf{u})][\mathbf{u}^* - E(\mathbf{u})]'\right\}\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} \\
&= (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime} var(\mathbf{u}^*)\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{I}_{MT}\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} \\
&= (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} = (\mathbf{X}'\mathbf{H}'\mathbf{H}\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} \\
&= \left\{ \begin{bmatrix} \mathbf{X}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M' \end{bmatrix} \begin{bmatrix} \sigma^{11}\mathbf{I}_T & \sigma^{12}\mathbf{I}_T & \cdots & \sigma^{1M}\mathbf{I}_T \\ \sigma^{21}\mathbf{I}_T & \sigma^{22}\mathbf{I}_T & \cdots & \sigma^{2M}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1}\mathbf{I}_T & \sigma^{M2}\mathbf{I}_T & \cdots & \sigma^{MM}\mathbf{I}_T \end{bmatrix} \right. \\
&\quad \left. \times \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{bmatrix} \right\}^{-1} \\
&= \begin{bmatrix} \sigma^{11}\mathbf{X}_1'\mathbf{X}_1 & \sigma^{12}\mathbf{X}_1'\mathbf{X}_2 & \cdots & \sigma^{1M}\mathbf{X}_1'\mathbf{X}_M \\ \sigma^{21}\mathbf{X}_2'\mathbf{X}_1 & \sigma^{22}\mathbf{X}_2'\mathbf{X}_2 & \cdots & \sigma^{2M}\mathbf{X}_2'\mathbf{X}_M \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1}\mathbf{X}_M'\mathbf{X}_1 & \sigma^{M2}\mathbf{X}_M'\mathbf{X}_2 & \cdots & \sigma^{MM}\mathbf{X}_M'\mathbf{X}_M \end{bmatrix}^{-1}_{(\bar{K}\times\bar{K})}. \qquad (2.8)
\end{aligned}$$

Nonetheless, the generalized least squares estimators are impossible to use when $\mathbf{W}$ is unknown. Zellner (1962) proposed to replace the unknown σ_{ij} with the estimate[5]

$$s_{ij} = \frac{\tilde{\mathbf{u}}_i'\tilde{\mathbf{u}}_j}{T - K_i} = \frac{(\mathbf{y}_i - \mathbf{X}_i\tilde{\boldsymbol{\beta}}_i)'(\mathbf{y}_j - \mathbf{X}_j\tilde{\boldsymbol{\beta}}_j)}{T - K_i} \quad \text{for} \quad i, j = 1, 2, \ldots, M, \qquad (2.9)$$

[5]When the system of M equations contains the same number of parameters (i.e., $K_1 = K_2 = \ldots = K_M$), the denominator $(T - K_i)$ is not unambiguously defined. When this is not the case, only T can be used in the denominator and s_{ij} will still be consistent (Griffiths et al., 1992, p. 551).

where $\tilde{\boldsymbol{\beta}}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}_i$ is the least-squares estimator.[6] Consequently, $\mathbf{W} = \boldsymbol{\Sigma_c} \otimes \mathbf{I}_T$ in equation (2.4) and $\mathbf{W}^{-1} = \boldsymbol{\Sigma_c}^{-1} \otimes \mathbf{I}_T$ in equation (2.6) will be estimated respectively by

$$\hat{\mathbf{W}} = \begin{bmatrix} s_{11}\mathbf{I}_T & s_{12}\mathbf{I}_T & \cdots & s_{1M}\mathbf{I}_T \\ s_{21}\mathbf{I}_T & s_{22}\mathbf{I}_T & \cdots & s_{2M}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ s_{M1}\mathbf{I}_T & s_{M2}\mathbf{I}_T & \cdots & s_{MM}\mathbf{I}_T \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1M} \\ s_{21} & s_{22} & \cdots & s_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ s_{M1} & s_{M2} & \cdots & s_{MM} \end{bmatrix} \otimes \mathbf{I}_T = \mathbf{S_c} \otimes \mathbf{I}_T \tag{2.10}$$

and

$$\hat{\mathbf{W}}^{-1} = \begin{bmatrix} s^{11}\mathbf{I}_T & s^{12}\mathbf{I}_T & \cdots & s^{1M}\mathbf{I}_T \\ s^{21}\mathbf{I}_T & s^{22}\mathbf{I}_T & \cdots & s^{2M}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ s^{M1}\mathbf{I}_T & s^{M2}\mathbf{I}_T & \cdots & s^{MM}\mathbf{I}_T \end{bmatrix} = \begin{bmatrix} s^{11} & s^{12} & \cdots & s^{1M} \\ s^{21} & s^{22} & \cdots & s^{2M} \\ \vdots & \vdots & \ddots & \vdots \\ s^{M1} & s^{M2} & \cdots & s^{MM} \end{bmatrix} \otimes \mathbf{I}_T = \mathbf{S_c}^{-1} \otimes \mathbf{I}_T, \tag{2.11}$$

where equation (2.11) is obtained by inversion of equation (2.10).[7]

Therefore, we can estimate $\hat{\boldsymbol{\beta}}$ in equation (2.7) and $var(\hat{\boldsymbol{\beta}})$ in equation (2.8) re-

[6] Note that when $i = j$, the estimate s_{ij} (also known as s^2) is the same estimate of σ_{ij} (also known as σ^2) used under least-squares estimation for equation i.

[7] The replacement of $\boldsymbol{\Sigma_c}$ with $\mathbf{S_c}$ leads in the literature to one of the many feasible generalized least squares estimators (FGLS). This particular replacement is called "restricted residuals" and it leads in literature to seemingly unrelated restricted residuals (SURR).

Srivastava and Giles (1987, p. 13) explain another way to obtain the s_{ij}'s. The approach is called "unrestricted residuals" because restrictions on the coefficients of the SUR model which distinguish it from the multivariate regression model are ignored. The unrestricted-residual approach leads in literature to seemingly unrelated unrestricted residuals (SUUR).

spectively by

$$
\begin{aligned}
\hat{\hat{\boldsymbol{\beta}}} &= (\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{y} \\
&= \begin{bmatrix} s^{11}\mathbf{X}_1'\mathbf{X}_1 & s^{12}\mathbf{X}_1'\mathbf{X}_2 & \cdots & s^{1M}\mathbf{X}_1'\mathbf{X}_M \\ s^{21}\mathbf{X}_2'\mathbf{X}_1 & s^{22}\mathbf{X}_2'\mathbf{X}_2 & \cdots & s^{2M}\mathbf{X}_2'\mathbf{X}_M \\ \vdots & \vdots & \ddots & \vdots \\ s^{M1}\mathbf{X}_M'\mathbf{X}_1 & s^{M2}\mathbf{X}_M'\mathbf{X}_2 & \cdots & s^{MM}\mathbf{X}_M'\mathbf{X}_M \end{bmatrix}^{-1}_{(\bar{K}\times\bar{K})} \begin{bmatrix} \sum_{i=1}^{M} s^{1i}\mathbf{X}_1'\mathbf{y}_i \\ \sum_{i=1}^{M} s^{2i}\mathbf{X}_2'\mathbf{y}_i \\ \vdots \\ \sum_{i=1}^{M} s^{Mi}\mathbf{X}_M'\mathbf{y}_i \end{bmatrix}_{(\bar{K}\times 1)} \qquad (2.12)
\end{aligned}
$$

and

$$
var(\hat{\hat{\boldsymbol{\beta}}}) = \begin{bmatrix} s^{11}\mathbf{X}_1'\mathbf{X}_1 & s^{12}\mathbf{X}_1'\mathbf{X}_2 & \cdots & s^{1M}\mathbf{X}_1'\mathbf{X}_M \\ s^{21}\mathbf{X}_2'\mathbf{X}_1 & s^{22}\mathbf{X}_2'\mathbf{X}_2 & \cdots & s^{2M}\mathbf{X}_2'\mathbf{X}_M \\ \vdots & \vdots & \ddots & \vdots \\ s^{M1}\mathbf{X}_M'\mathbf{X}_1 & s^{M2}\mathbf{X}_M'\mathbf{X}_2 & \cdots & s^{MM}\mathbf{X}_M'\mathbf{X}_M \end{bmatrix}^{-1}_{(\bar{K}\times\bar{K})} . \qquad (2.13)
$$

2.1.1 Special Cases

The first special case, explained by Zellner (1962), occurs when $E[\mathbf{u}_i\mathbf{u}_j'] = \sigma_{ij}\mathbf{I}_T = \mathbf{0}_{(T\times T)}$ for $i \neq j$. In this case, the generalized least squares estimator for the model in equation (2.2) or (2.3) will be identical to applying least-squares to each equation. In addition, the generalized least squares variances of the estimators will reduce to the least-squares variances of the estimators. That is, if

$$
\mathbf{W} = \begin{bmatrix} \sigma_{11}\mathbf{I}_T & \sigma_{12}\mathbf{I}_T & \cdots & \sigma_{1M}\mathbf{I}_T \\ \sigma_{21}\mathbf{I}_T & \sigma_{22}\mathbf{I}_T & \cdots & \sigma_{2M}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1}\mathbf{I}_T & \sigma_{M2}\mathbf{I}_T & \cdots & \sigma_{MM}\mathbf{I}_T \end{bmatrix} = \begin{bmatrix} \sigma_{11}\mathbf{I}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_{22}\mathbf{I}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_{MM}\mathbf{I}_T \end{bmatrix}
$$

and

$$
\begin{aligned}
\mathbf{W}^{-1} &= \begin{bmatrix} \sigma^{11}\mathbf{I}_T & \sigma^{12}\mathbf{I}_T & \cdots & \sigma^{1M}\mathbf{I}_T \\ \sigma^{21}\mathbf{I}_T & \sigma^{22}\mathbf{I}_T & \cdots & \sigma^{2M}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1}\mathbf{I}_T & \sigma^{M2}\mathbf{I}_T & \cdots & \sigma^{MM}\mathbf{I}_T \end{bmatrix} = \begin{bmatrix} \sigma^{11}\mathbf{I}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma^{22}\mathbf{I}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma^{MM}\mathbf{I}_T \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sigma_{11}}\mathbf{I}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma_{22}}\mathbf{I}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{\sigma_{MM}}\mathbf{I}_T \end{bmatrix},
\end{aligned}
$$

then

$$
\begin{aligned}
\underset{(\bar{K}\times 1)}{\hat{\boldsymbol{\beta}}} &= \left\{ \begin{bmatrix} \mathbf{X}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M' \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_{11}}\mathbf{I}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma_{22}}\mathbf{I}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{\sigma_{MM}}\mathbf{I}_T \end{bmatrix} \right. \\
&\quad \times \left. \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{bmatrix} \right\}^{-1} \begin{bmatrix} \mathbf{X}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M' \end{bmatrix} \\
&\quad \times \begin{bmatrix} \frac{1}{\sigma_{11}}\mathbf{I}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma_{22}}\mathbf{I}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{\sigma_{MM}}\mathbf{I}_T \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sigma_{11}}\mathbf{X}_1'\mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma_{22}}\mathbf{X}_2'\mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{\sigma_{MM}}\mathbf{X}_M'\mathbf{X}_M \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sigma_{11}}\mathbf{X}_1'\mathbf{y}_1 \\ \frac{1}{\sigma_{22}}\mathbf{X}_2'\mathbf{y}_2 \\ \vdots \\ \frac{1}{\sigma_{MM}}\mathbf{X}_M'\mathbf{y}_M \end{bmatrix} \\
&= \begin{bmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1 \\ (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y}_2 \\ \vdots \\ (\mathbf{X}_M'\mathbf{X}_M)^{-1}\mathbf{X}_M'\mathbf{y}_M \end{bmatrix} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \tilde{\boldsymbol{\beta}}. \qquad (2.14)
\end{aligned}
$$

Then, since $var(\hat{\boldsymbol{\beta}}) = E\{[\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})][\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})]'\} = E\{[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}][\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]'\}$ and in this case $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$ then

$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$. Therefore,

$$
\begin{aligned}
var(\hat{\boldsymbol{\beta}}) &= E\{[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}]'\} = E\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E\{[\mathbf{u} - E(\mathbf{u})][\mathbf{u} - E(\mathbf{u})]'\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \begin{bmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (\mathbf{X}_M'\mathbf{X}_M)^{-1}\mathbf{X}_M' \end{bmatrix} \\
&\quad \times \begin{bmatrix} \sigma_{11}\mathbf{I}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_{22}\mathbf{I}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_{MM}\mathbf{I}_T \end{bmatrix} \\
&\quad \times \begin{bmatrix} \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M(\mathbf{X}_M'\mathbf{X}_M)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{11}(\mathbf{X}_1'\mathbf{X}_1)^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_{22}(\mathbf{X}_2'\mathbf{X}_2)^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_{MM}(\mathbf{X}_M'\mathbf{X}_M)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} var(\tilde{\boldsymbol{\beta}}_1) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & var(\tilde{\boldsymbol{\beta}}_2) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & var(\tilde{\boldsymbol{\beta}}_M) \end{bmatrix}_{(\bar{K}\times\bar{K})}, \qquad (2.15)
\end{aligned}
$$

where the diagonal blocks are the least-squares variances of the estimators for the corresponding equations.

Alternatively, we could have used equation (2.8), where

$$
\begin{aligned}
var(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} \\
&= \left\{ \begin{bmatrix} \mathbf{X}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M' \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_{11}}\mathbf{I}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma_{22}}\mathbf{I}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{\sigma_{MM}}\mathbf{I}_T \end{bmatrix} \right. \\
&\quad \left. \times \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{bmatrix} \right\}^{-1} \\
&= \begin{bmatrix} \sigma_{11}(\mathbf{X}_1'\mathbf{X}_1)^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_{22}(\mathbf{X}_2'\mathbf{X}_2)^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_{MM}(\mathbf{X}_M'\mathbf{X}_M)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} var(\tilde{\boldsymbol{\beta}}_1) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & var(\tilde{\boldsymbol{\beta}}_2) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & var(\tilde{\boldsymbol{\beta}}_M) \end{bmatrix}_{(\bar{K}\times\bar{K})} .
\end{aligned}
$$

The second special case, explained by Zellner (1962), occurs when $\mathbf{X}_1 = \mathbf{X}_2 = \ldots = \mathbf{X}_M = \mathbf{X}_i$, $i = 1, 2, \ldots, M$. In this case the generalized least squares estimators will reduce again to single equation least-squares estimators even when $E[\mathbf{u}_i\mathbf{u}_j'] = \sigma_{ij}\mathbf{I}_T \neq \mathbf{0}_{(T\times T)}$. Similarly, the generalized least squares variances of the estimators will also reduce to the least squares variances of the estimators but the generalized least squares covariances of the estimators are unique to the generalized least squares variance-covariance matrix. That is,[8]

$$
\begin{aligned}
\underset{(\bar{K}\times 1)}{\hat{\boldsymbol{\beta}}} &= (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{y} \\
&= \left[\mathbf{X}'(\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{I}_T)\mathbf{X}\right]^{-1}\mathbf{X}'(\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{I}_T)\mathbf{y} \\
&= \left[\begin{bmatrix} \mathbf{X}_i' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_i' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_i' \end{bmatrix} (\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{I}_T) \begin{bmatrix} \mathbf{X}_i & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_i & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_i \end{bmatrix}\right]^{-1} \\
&\quad \times \begin{bmatrix} \mathbf{X}_i' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_i' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_i' \end{bmatrix} (\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{I}_T)\mathbf{y} \\
&= \left[(\mathbf{I}_M\otimes\mathbf{X}_i')(\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{I}_T)(\mathbf{I}_M\otimes\mathbf{X}_i)\right]^{-1}(\mathbf{I}_M\otimes\mathbf{X}_i')(\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{I}_T)\mathbf{y} \\
&= \left[(\mathbf{I}_M\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{X}_i'\mathbf{I}_T)(\mathbf{I}_M\otimes\mathbf{X}_i)\right]^{-1}(\mathbf{I}_M\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{X}_i'\mathbf{I}_T)\mathbf{y} \\
&= \left[(\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{X}_i')(\mathbf{I}_M\otimes\mathbf{X}_i)\right]^{-1}(\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{X}_i')\mathbf{y} \\
&= \left[(\boldsymbol{\Sigma_c}^{-1}\mathbf{I}_M\otimes\mathbf{X}_i'\mathbf{X}_i)\right]^{-1}(\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{X}_i')\mathbf{y} = (\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{X}_i'\mathbf{X}_i)^{-1}(\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{X}_i')\mathbf{y} \\
&= \left[\boldsymbol{\Sigma_c}\otimes(\mathbf{X}_i'\mathbf{X}_i)^{-1}\right](\boldsymbol{\Sigma_c}^{-1}\otimes\mathbf{X}_i')\mathbf{y} = \left[\boldsymbol{\Sigma_c}\boldsymbol{\Sigma_c}^{-1}\otimes(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\right]\mathbf{y} \\
&= \left[\mathbf{I}_M\otimes(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\right]\mathbf{y} = \left[\mathbf{I}_M\mathbf{I}_M\otimes(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\right]\mathbf{y} \\
&= \left[\mathbf{I}_M\otimes(\mathbf{X}_i'\mathbf{X}_i)^{-1}\right](\mathbf{I}_M\otimes\mathbf{X}_i')\mathbf{y} \\
&= \begin{bmatrix} (\mathbf{X}_i'\mathbf{X}_i)^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}_i'\mathbf{X}_i)^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (\mathbf{X}_i'\mathbf{X}_i)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{X}_i' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_i' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_i' \end{bmatrix} \mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \tilde{\boldsymbol{\beta}}. \qquad (2.16)
\end{aligned}
$$

Once again, since $var(\hat{\boldsymbol{\beta}}) = E\{[\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})][\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})]'\} = E\{[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}][\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]'\}$ and also in this case $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$

[8]Useful Kronecker product properties are found in Harville (1997, Chapter 16).

then $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$. Therefore,

$$
\begin{aligned}
var(\hat{\boldsymbol{\beta}}) &= E\{[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}]'\} = E\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E\{[\mathbf{u} - E(\mathbf{u})][\mathbf{u} - E(\mathbf{u})]'\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \begin{bmatrix} (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i' \end{bmatrix} \\
&\quad \times \begin{bmatrix} \sigma_{11}\mathbf{I}_T & \sigma_{12}\mathbf{I}_T & \cdots & \sigma_{1M}\mathbf{I}_T \\ \sigma_{21}\mathbf{I}_T & \sigma_{22}\mathbf{I}_T & \cdots & \sigma_{2M}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1}\mathbf{I}_T & \sigma_{M2}\mathbf{I}_T & \cdots & \sigma_{MM}\mathbf{I}_T \end{bmatrix} \\
&\quad \times \begin{bmatrix} \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{11}(\mathbf{X}_i'\mathbf{X}_i)^{-1} & \sigma_{12}(\mathbf{X}_i'\mathbf{X}_i)^{-1} & \cdots & \sigma_{1M}(\mathbf{X}_i'\mathbf{X}_i)^{-1} \\ \sigma_{21}(\mathbf{X}_i'\mathbf{X}_i)^{-1} & \sigma_{22}(\mathbf{X}_i'\mathbf{X}_i)^{-1} & \cdots & \sigma_{2M}(\mathbf{X}_i'\mathbf{X}_i)^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1}(\mathbf{X}_i'\mathbf{X}_i)^{-1} & \sigma_{M2}(\mathbf{X}_i'\mathbf{X}_i)^{-1} & \cdots & \sigma_{MM}(\mathbf{X}_i'\mathbf{X}_i)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} var(\tilde{\boldsymbol{\beta}}_1) & \sigma_{12}(\mathbf{X}_i'\mathbf{X}_i)^{-1} & \cdots & \sigma_{1M}(\mathbf{X}_i'\mathbf{X}_i)^{-1} \\ \sigma_{21}(\mathbf{X}_i'\mathbf{X}_i)^{-1} & var(\tilde{\boldsymbol{\beta}}_2) & \cdots & \sigma_{2M}(\mathbf{X}_i'\mathbf{X}_i)^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1}(\mathbf{X}_i'\mathbf{X}_i)^{-1} & \sigma_{M2}(\mathbf{X}_i'\mathbf{X}_i)^{-1} & \cdots & var(\tilde{\boldsymbol{\beta}}_M) \end{bmatrix}, \qquad (2.17)
\end{aligned}
$$

where the diagonal blocks are the least-squares variances of the estimators for the corresponding equations but the off-diagonal blocks are unique to the generalized least squares variance-covariance matrix.

Alternatively, if we use equation (2.8), we get

$$\begin{aligned}
var(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} \\
&= \left\{ \begin{bmatrix} \mathbf{X}_i' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_i' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_i' \end{bmatrix} \begin{bmatrix} \sigma^{11}\mathbf{I}_T & \sigma^{12}\mathbf{I}_T & \cdots & \sigma^{1M}\mathbf{I}_T \\ \sigma^{21}\mathbf{I}_T & \sigma^{22}\mathbf{I}_T & \cdots & \sigma^{2M}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1}\mathbf{I}_T & \sigma^{M2}\mathbf{I}_T & \cdots & \sigma^{MM}\mathbf{I}_T \end{bmatrix} \right. \\
&\quad \left. \times \begin{bmatrix} \mathbf{X}_i & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_i & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_i \end{bmatrix} \right\}^{-1} \\
&= \begin{bmatrix} \sigma^{11}\mathbf{X}_i'\mathbf{X}_i & \sigma^{12}\mathbf{X}_i'\mathbf{X}_i & \cdots & \sigma^{1M}\mathbf{X}_i'\mathbf{X}_i \\ \sigma^{21}\mathbf{X}_i'\mathbf{X}_i & \sigma^{22}\mathbf{X}_i'\mathbf{X}_i & \cdots & \sigma^{2M}\mathbf{X}_i'\mathbf{X}_i \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1}\mathbf{X}_i'\mathbf{X}_i & \sigma^{M2}\mathbf{X}_i'\mathbf{X}_i & \cdots & \sigma^{MM}\mathbf{X}_i'\mathbf{X}_i \end{bmatrix}^{-1}_{(\bar{K}\times\bar{K})} . \qquad (2.18)
\end{aligned}$$

This means that computing the inverse in equation (2.18) will give equation (2.17).

However, when the $\mathbf{X}_i$, $i = 1, 2, \ldots, M$, are not all the same or when $E[\mathbf{u}_i\mathbf{u}_j'] = \sigma_{ij}\mathbf{I}_T \neq \mathbf{0}_{(T\times T)}$, the generalized least squares estimators will be different from the single-equation least squares estimators. In particular, a quite large gain in efficiency can be obtained when independent variables in different equations are not highly correlated and when the error terms in different equations are highly correlated.

2.2 Properties

Zellner (1962) showed the following properties:

- First, $\hat{\hat{\boldsymbol{\beta}}} = \hat{\boldsymbol{\beta}} + \mathcal{O}_p(T^{-1})$.
- Second, $T^{1/2}(\hat{\hat{\boldsymbol{\beta}}} - \boldsymbol{\beta})$ and $T^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ have the same asymptotic multivariate normal distribution.

- Third,

$$
\begin{aligned}
\underset{(\bar{K}\times\bar{K})}{var(\hat{\hat{\boldsymbol{\beta}}})} &= (\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1} + o_p(T^{-1}) \\
&= \begin{bmatrix} s^{11}\mathbf{X}_1'\mathbf{X}_1 & s^{12}\mathbf{X}_1'\mathbf{X}_2 & \cdots & s^{1M}\mathbf{X}_1'\mathbf{X}_M \\ s^{21}\mathbf{X}_2'\mathbf{X}_1 & s^{22}\mathbf{X}_2'\mathbf{X}_2 & \cdots & s^{2M}\mathbf{X}_2'\mathbf{X}_M \\ \vdots & \vdots & \ddots & \vdots \\ s^{M1}\mathbf{X}_M'\mathbf{X}_1 & s^{M2}\mathbf{X}_M'\mathbf{X}_2 & \cdots & s^{MM}\mathbf{X}_M'\mathbf{X}_M \end{bmatrix}^{-1} + o_p(T^{-1}), \quad (2.19)
\end{aligned}
$$

where $\mathcal{O}_p(T^{-1})$ denotes a quantity which is of the order T^{-1} in probability and $o_p(T^{-1})$ denotes terms of higher order of smallness than T^{-1}.[9]

Property 1: $\hat{\hat{\boldsymbol{\beta}}} = \hat{\boldsymbol{\beta}} + \mathcal{O}_p(T^{-1})$.

Let $\hat{\mathbf{W}} = (\boldsymbol{\Sigma}_{\mathbf{c}} + \boldsymbol{\Delta}_{\mathbf{1}}) \otimes \mathbf{I}_T$ where $\boldsymbol{\Sigma}_{\mathbf{c}} \otimes \mathbf{I}_T$ is given in equation (2.4) and $\boldsymbol{\Delta}_{\mathbf{1}}$ is a matrix whose elements are the sampling errors of using the elements of $\mathbf{S}_{\mathbf{c}}$ as estimates

[9]Brockwell and Davis (1987, p. 192, Definition 6.1.3) define convergence in probability and order in probability as follows:

Let $\{a_n,\ n = 1, 2, \ldots\}$ be a sequence of strictly positive real numbers and let $\{X_n,\ n = 1, 2, \ldots\}$ be a sequence of random variables all defined on the same probability space. Then,

(i) X_n converges in probability to the random variable X, written $X_n \xrightarrow{P} X$, if and only if $X_n - X = o_p(1)$. That is, we say that $X_n - X$ converges in probability to zero, written $X_n - X = o_p(1)$ or $X_n - X \xrightarrow{P} 0$, if for every $\varepsilon > 0$, $Pr(|X_n - X| > \varepsilon) \to 0 \quad \text{as} \quad n \to \infty$.

(ii) $X_n = o_p(a_n)$ if and only if $a_n^{-1}X_n = o_p(1)$. That is, we say that $\frac{X_n}{a_n}$ converges in probability to zero, written $\frac{X_n}{a_n} = o_p(1)$ or $\frac{X_n}{a_n} \xrightarrow{P} 0$, if for every $\varepsilon > 0$, $Pr\left(\left|\frac{X_n}{a_n}\right| > \varepsilon\right) \to 0 \quad \text{as} \quad n \to \infty$.

(iii) $X_n = \mathcal{O}_p(a_n)$ if and only if $a_n^{-1}X_n = \mathcal{O}_p(1)$. That is, we say that the sequence $\{\frac{X_n}{a_n}\}$ is bounded in probability (or tight), written $X_n = \mathcal{O}_p(1)$, if for every $\varepsilon > 0$ there exists $\delta(\varepsilon) \in (0, \infty)$ such that $Pr\left(\left|\frac{X_n}{a_n}\right| > \delta(\varepsilon)\right) < \varepsilon \quad \text{for all} \quad n$.

of the elements of $\mathbf{\Sigma_c}$. That is, $\mathbf{\Delta_1} = \mathbf{S_c} - \mathbf{\Sigma_c}$. Then,

$$\begin{aligned}
\mathbf{\Delta_1} \otimes \mathbf{I}_T &= \mathbf{S_c} \otimes \mathbf{I}_T - \mathbf{\Sigma_c} \otimes \mathbf{I}_T \\
&= \hat{\mathbf{W}} - \mathbf{W} \\
&= \begin{bmatrix} s_{11}\mathbf{I}_T & s_{12}\mathbf{I}_T & \cdots & s_{1M}\mathbf{I}_T \\ s_{21}\mathbf{I}_T & s_{22}\mathbf{I}_T & \cdots & s_{2M}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ s_{M1}\mathbf{I}_T & s_{M2}\mathbf{I}_T & \cdots & s_{MM}\mathbf{I}_T \end{bmatrix} - \begin{bmatrix} \sigma_{11}\mathbf{I}_T & \sigma_{12}\mathbf{I}_T & \cdots & \sigma_{1M}\mathbf{I}_T \\ \sigma_{21}\mathbf{I}_T & \sigma_{22}\mathbf{I}_T & \cdots & \sigma_{2M}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1}\mathbf{I}_T & \sigma_{M2}\mathbf{I}_T & \cdots & \sigma_{MM}\mathbf{I}_T \end{bmatrix} \\
&= \begin{bmatrix} (s_{11} - \sigma_{11}) & (s_{12} - \sigma_{12}) & \cdots & (s_{1M} - \sigma_{1M}) \\ (s_{21} - \sigma_{21}) & (s_{22} - \sigma_{22}) & \cdots & (s_{2M} - \sigma_{2M}) \\ \vdots & \vdots & \ddots & \vdots \\ (s_{M1} - \sigma_{M1}) & (s_{M2} - \sigma_{M2}) & \cdots & (s_{MM} - \sigma_{MM}) \end{bmatrix} \otimes \mathbf{I}_T \\
&= \begin{bmatrix} \delta_{11}^{(1)} & \delta_{12}^{(1)} & \cdots & \delta_{1M}^{(1)} \\ \delta_{21}^{(1)} & \delta_{22}^{(1)} & \cdots & \delta_{2M}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{M1}^{(1)} & \delta_{M2}^{(1)} & \cdots & \delta_{MM}^{(1)} \end{bmatrix} \otimes \mathbf{I}_T,
\end{aligned}$$

where each $\delta_{ij}^{(1)}$ is $\mathcal{O}_p(T^{-1/2})$ according to Zellner (1962).[10]

Thus,

$$\hat{\mathbf{W}}^{-1} = (\mathbf{\Sigma_c} + \mathbf{\Delta_1})^{-1} \otimes \mathbf{I}_T. \tag{2.20}$$

Now we use a theorem from Harville (1997, p. 429, Theorem 18.2.16) for the

[10] Brockwell and Davis (1987, p. 193, Definition 6.1.4) define order in probability for random vectors as follows:

Suppose that $\{\mathbf{X}_n, n = 1, 2, \ldots\}$ is a sequence of random vectors, all defined on the same probability space such that $\mathbf{X_n}$ has k components $X_{n1}, X_{n2}, \ldots, X_{nk}$, $n = 1, 2, \ldots$. Then,

(i) $\mathbf{X}_n = o_p(a_n)$ if and only if $X_{nj} = o_p(a_n)$, $j = 1, 2, \ldots, k$.

(ii) $\mathbf{X}_n = \mathcal{O}_p(a_n)$ if and only if $X_{nj} = \mathcal{O}_p(a_n)$, $j = 1, 2, \ldots, k$.

(iii) $\mathbf{X}_n$ converges in probability to the random vector $\mathbf{X}$, written $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$, if and only if $\mathbf{X}_n - \mathbf{X} = o_p(1)$.

Therefore, by applying (ii) to $\mathbf{\Delta_1}$ we conclude that $\mathbf{\Delta_1}$ and consequently $\mathbf{\Delta_1} \otimes \mathbf{I}_T$ are $\mathcal{O}_p(T^{-1/2})$.

geometric series of a square matrix. The theorem states that for $\mathbf{C}_{(M\times M)}$, the infinite series $\mathbf{I}+\mathbf{C}+\mathbf{C}^2+\mathbf{C}^3+\ldots$ converges if and only if $\lim_{k\to\infty}\mathbf{C}^k=0$, in which case $\mathbf{I}-\mathbf{C}$ is nonsingular and

$$(\mathbf{I}-\mathbf{C})^{-1}=\sum_{k=0}^{\infty}\mathbf{C}^k=\mathbf{I}+\mathbf{C}+\mathbf{C}^2+\mathbf{C}^3+\ldots$$

(where $\mathbf{C}^0=\mathbf{I}$).

Let $\mathbf{C}=\mathbf{A}^{-1}(-\mathbf{B})$ and suppose $\lim_{k\to\infty}(\mathbf{A}^{-1}\mathbf{B})^k=0$. Then,

$$\begin{aligned}
(\mathbf{A}+\mathbf{B})^{-1} &= [\mathbf{A}-(-\mathbf{B})]^{-1}=\left[\mathbf{A}(\mathbf{I}_M-\mathbf{A}^{-1}(-\mathbf{B}))\right]^{-1}\\
&= \left[\mathbf{I}_M-\mathbf{A}^{-1}(-\mathbf{B})\right]^{-1}\mathbf{A}^{-1}\\
&= \left[\mathbf{I}_M+(\mathbf{A}^{-1}(-\mathbf{B}))+(\mathbf{A}^{-1}(-\mathbf{B}))^2+(\mathbf{A}^{-1}(-\mathbf{B}))^3+\ldots\right]\mathbf{A}^{-1}\\
&= \left[\mathbf{A}^{-1}-(\mathbf{A}^{-1}\mathbf{B})\mathbf{A}^{-1}+(\mathbf{A}^{-1}\mathbf{B})^2\mathbf{A}^{-1}-(\mathbf{A}^{-1}\mathbf{B})^3\mathbf{A}^{-1}+\ldots\right].
\end{aligned}$$

Now, let $\mathbf{A}=\boldsymbol{\Sigma}_{\boldsymbol{c}(M\times M)}$ and $\mathbf{B}=\boldsymbol{\Delta}_{\mathbf{1}(M\times M)}$. In order to apply this previous result to equation (2.20), we need to show $\lim_{k\to\infty}(\mathbf{A}^{-1}\mathbf{B})^k=0$. Using another theorem from Harville (1997, p. 431, Theorem 18.2.19), if $||\mathbf{C}_{(n\times n)}||<1$ then $\lim_{k\to\infty}\mathbf{C}^k=0$.

Since

$$\underset{(M\times M)}{\mathbf{S_c}}=\begin{bmatrix} s_{11} & s_{12} & \ldots & s_{1M}\\ s_{21} & s_{22} & \ldots & s_{2M}\\ \vdots & \vdots & \ddots & \vdots\\ s_{M1} & s_{M2} & \ldots & s_{MM}\end{bmatrix}$$

and $\boldsymbol{\Delta}_{\mathbf{1}}=\mathbf{S_c}-\boldsymbol{\Sigma}_{\boldsymbol{c}}$, then

$$||\mathbf{C}||=||\mathbf{A}^{-1}\mathbf{B}||=||\boldsymbol{\Sigma}_{\boldsymbol{c}}{}^{-1}\boldsymbol{\Delta}_{\mathbf{1}}||=||\boldsymbol{\Sigma}_{\boldsymbol{c}}{}^{-1}(\mathbf{S_c}-\boldsymbol{\Sigma}_{\boldsymbol{c}})||=||\boldsymbol{\Sigma}_{\boldsymbol{c}}{}^{-1}\mathbf{S_c}-\mathbf{I}_M||.$$

Griffiths et al. (1992, pp. 447–451 and p. 551) explain that $\underset{T\to\infty}{lim}\ Pr\left(|s_{ij}-\sigma_{ij}|<\varepsilon\right)=$

1 for any $\varepsilon > 0$.[11] Consequently, [12]

$$\mathbf{S_c} \xrightarrow{P} \boldsymbol{\Sigma_c} \quad \text{as } T \to \infty\,;$$

$$\boldsymbol{\Sigma_c}^{-1}\mathbf{S_c} \xrightarrow{P} \mathbf{I}_M \quad \text{as } T \to \infty\,;$$

$$\boldsymbol{\Sigma_c}^{-1}\mathbf{S_c} - \mathbf{I}_M \xrightarrow{P} \mathbf{0}_{(M\times M)} \quad \text{as } T \to \infty\,;$$

$$||\boldsymbol{\Sigma_c}^{-1}\mathbf{S_c} - \mathbf{I}_M|| \xrightarrow{P} \mathbf{0}_{(M\times M)} \quad \text{as } T \to \infty\,,$$

where $\xrightarrow{P}$ denotes convergence in probability.

Therefore, for sufficiently large T,[13] equation (2.20) becomes

$$\begin{aligned}
\hat{\mathbf{W}}^{-1} &= [\boldsymbol{\Sigma_c}^{-1} - \boldsymbol{\Sigma_c}^{-1}\boldsymbol{\Delta_1}\boldsymbol{\Sigma_c}^{-1} - (\boldsymbol{\Sigma_c}^{-1}\boldsymbol{\Delta_1})^2\boldsymbol{\Sigma_c}^{-1} - \cdots] \otimes \mathbf{I}_T \\
&= \boldsymbol{\Sigma_c}^{-1} \otimes \mathbf{I}_T - (\boldsymbol{\Sigma_c}^{-1}\boldsymbol{\Delta_1}\boldsymbol{\Sigma_c}^{-1}) \otimes \mathbf{I}_T + \left[(\boldsymbol{\Sigma_c}^{-1}\boldsymbol{\Delta_1})^2\boldsymbol{\Sigma_c}^{-1}\right] \otimes \mathbf{I}_T - \cdots \\
&= \mathbf{W}^{-1} - \boldsymbol{\Delta_2} + \left[(\boldsymbol{\Sigma_c}^{-1}\boldsymbol{\Delta_1})^2\boldsymbol{\Sigma_c}^{-1}\right] \otimes \mathbf{I}_T - \ldots.
\end{aligned}$$

Neglecting the terms of higher order of smallness,[14] we have

$$\hat{\mathbf{W}}^{-1} \approx \mathbf{W}^{-1} - \boldsymbol{\Delta_2}, \tag{2.21}$$

where $\boldsymbol{\Delta_2} = (\boldsymbol{\Sigma_c}^{-1}\boldsymbol{\Delta_1}\boldsymbol{\Sigma_c}^{-1}) \otimes \mathbf{I}_T$.

---

[11] Equivalently, we can write

$$\underset{T\to\infty}{lim}\,[1 - Pr(|s_{ij} - \sigma_{ij}| > \varepsilon)] = 1\,;$$

$$\underset{T\to\infty}{lim}\,Pr(|s_{ij} - \sigma_{ij}| > \varepsilon) = 0\,;$$

$$s_{ij} - \sigma_{ij} = o_p(1) \quad \text{or} \quad s_{ij} \xrightarrow{P} \sigma_{ij}\,.$$

[12] Harville (1997, p. 59) defines the norm $||\mathbf{D}||$ of a matrix $\mathbf{D}_{(m\times n)} = (d_{ij})$ as

$$||\mathbf{D}|| = (\mathbf{D} \bullet \mathbf{D})^{1/2} = [\mathrm{tr}(\mathbf{D}'\mathbf{D})]^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij}^2\right)^{1/2}.$$

[13] For sufficiently large T means $\exists\, T_0 \ni \forall\, T > T_0$ our claim holds.

[14] $(\boldsymbol{\Sigma_c}^{-1}\boldsymbol{\Delta_1})^2, (\boldsymbol{\Sigma_c}^{-1}\boldsymbol{\Delta_1})^3, \ldots \xrightarrow{P} \mathbf{0}_{(M\times M)}$ faster than $(\boldsymbol{\Sigma_c}^{-1}\boldsymbol{\Delta_1}) \xrightarrow{P} \mathbf{0}_{(M\times M)}$ as $T \to \infty$.

That is,

$$
\begin{aligned}
\mathbf{\Delta_2} &= \left\{ \begin{bmatrix} \sigma^{11} & \sigma^{12} & \cdots & \sigma^{1M} \\ \sigma^{21} & \sigma^{22} & \cdots & \sigma^{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1} & \sigma^{M2} & \cdots & \sigma^{MM} \end{bmatrix} \begin{bmatrix} \delta_{11}^{(1)} & \delta_{12}^{(1)} & \cdots & \delta_{1M}^{(1)} \\ \delta_{21}^{(1)} & \delta_{22}^{(1)} & \cdots & \delta_{2M}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{M1}^{(1)} & \delta_{M2}^{(1)} & \cdots & \delta_{MM}^{(1)} \end{bmatrix} \right. \\
&\quad \left. \times \begin{bmatrix} \sigma^{11} & \sigma^{12} & \cdots & \sigma^{1M} \\ \sigma^{21} & \sigma^{22} & \cdots & \sigma^{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1} & \sigma^{M2} & \cdots & \sigma^{MM} \end{bmatrix} \right\} \otimes \mathbf{I}_T \\
&= \begin{bmatrix} \sum_{i=1}^{M} \sigma^{1i}\delta_{i1}^{(1)} & \sum_{i=1}^{M} \sigma^{1i}\delta_{i2}^{(1)} & \cdots & \sum_{i=1}^{M} \sigma^{1i}\delta_{iM}^{(1)} \\ \sum_{i=1}^{M} \sigma^{2i}\delta_{i1}^{(1)} & \sum_{i=1}^{M} \sigma^{2i}\delta_{i2}^{(1)} & \cdots & \sum_{i=1}^{M} \sigma^{2i}\delta_{iM}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{M} \sigma^{Mi}\delta_{i1}^{(1)} & \sum_{i=1}^{M} \sigma^{Mi}\delta_{i2}^{(1)} & \cdots & \sum_{i=1}^{M} \sigma^{Mi}\delta_{iM}^{(1)} \end{bmatrix} \begin{bmatrix} \sigma^{11} & \sigma^{12} & \cdots & \sigma^{1M} \\ \sigma^{21} & \sigma^{22} & \cdots & \sigma^{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1} & \sigma^{M2} & \cdots & \sigma^{MM} \end{bmatrix} \otimes \mathbf{I}_T \\
&= \begin{bmatrix} \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{1i}\delta_{ij}^{(1)}\sigma^{j1} & \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{1i}\delta_{ij}^{(1)}\sigma^{j2} & \cdots & \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{1i}\delta_{ij}^{(1)}\sigma^{jM} \\ \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{2i}\delta_{ij}^{(1)}\sigma^{j1} & \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{2i}\delta_{ij}^{(1)}\sigma^{j2} & \cdots & \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{2i}\delta_{ij}^{(1)}\sigma^{jM} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{Mi}\delta_{ij}^{(1)}\sigma^{j1} & \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{Mi}\delta_{ij}^{(1)}\sigma^{j2} & \cdots & \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{Mi}\delta_{ij}^{(1)}\sigma^{jM} \end{bmatrix} \otimes \mathbf{I}_T \\
&= \begin{bmatrix} \delta_{11}^{(2)} & \delta_{12}^{(2)} & \cdots & \delta_{1M}^{(2)} \\ \delta_{21}^{(2)} & \delta_{22}^{(2)} & \cdots & \delta_{2M}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{M1}^{(2)} & \delta_{M2}^{(2)} & \cdots & \delta_{MM}^{(2)} \end{bmatrix} \otimes \mathbf{I}_T,
\end{aligned}
$$

where $\delta_{kl}^{(2)} = \sum_{i=1}^{M}\sum_{j=1}^{M} \sigma^{ki}\delta_{ij}\sigma^{jl}$ is $\mathcal{O}_p(T^{-1/2})$ according to Zellner (1962).[15]

[15]Consequently, $\mathbf{\Delta_2}$ is $\mathcal{O}_p(T^{-1/2})$.

In addition, Srivastava and Giles (1987, p. 42) explain $\mathbf{X}'\mathbf{X} = \mathcal{O}(T)$, where $\mathcal{O}(\cdot)$ denotes the

Now, the generalized least-squares estimator is

$$\begin{aligned}
\hat{\hat{\boldsymbol{\beta}}} &= (\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{y} \\
&= (\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{W}}^{-1}(\mathbf{X}\boldsymbol{\beta}+\mathbf{u}) \\
&= (\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})\boldsymbol{\beta}+(\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{u}.
\end{aligned}$$

Thus,

$$\hat{\hat{\boldsymbol{\beta}}}-\boldsymbol{\beta} = (\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{u}.$$

Substitution of equation (2.21) yields

$$\begin{aligned}
\hat{\hat{\boldsymbol{\beta}}}-\boldsymbol{\beta} &\approx [\mathbf{X}'(\mathbf{W}^{-1}-\boldsymbol{\Delta_2})\mathbf{X}]^{-1}\mathbf{X}'(\mathbf{W}^{-1}-\boldsymbol{\Delta_2})\mathbf{u} \\
&= \left\{[\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}][\mathbf{I}-(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X}]\right\}^{-1}\mathbf{X}'(\mathbf{W}^{-1}-\boldsymbol{\Delta_2})\mathbf{u} \\
&= \left\{(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})-(\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X})\right\}^{-1}\mathbf{X}'(\mathbf{W}^{-1}-\Delta_2)\mathbf{u}. \qquad (2.22)
\end{aligned}$$

Now we use another result derived from the geometric series of a square matrix. Let $\mathbf{C}_{(\bar{K}\times\bar{K})} = \mathbf{A}^{-1}_{(\bar{K}\times\bar{K})}\mathbf{B}_{(\bar{K}\times\bar{K})}$. Then,

$$\begin{aligned}
(\mathbf{A}-\mathbf{B})^{-1} &= \left[\mathbf{A}(\mathbf{I}_{\bar{K}}-\mathbf{A}^{-1}\mathbf{B})\right]^{-1} = \left(\mathbf{I}_{\bar{K}}-\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{A}^{-1} \\
&= \left[\mathbf{I}_{\bar{K}}+(\mathbf{A}^{-1}\mathbf{B})+(\mathbf{A}^{-1}\mathbf{B})^2+\ldots\right]\mathbf{A}^{-1} \\
&= \left[\mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}+(\mathbf{A}^{-1}\mathbf{B})^2\mathbf{A}^{-1}+(\mathbf{A}^{-1}\mathbf{B})^3\mathbf{A}^{-1}+\ldots\right].
\end{aligned}$$

In order to apply this result to (2.22), let $\mathbf{A}_{(\bar{K}\times\bar{K})} = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})$ and $\mathbf{B}_{(\bar{K}\times\bar{K})} =$

---

order arising from mathematical convergence. Therefore, by Proposition 6.1.1 (i) and Definition 6.1.4 (ii) from Brockwell and Davis (1987, pp. 192–193), and because $\boldsymbol{\Delta_2} = \mathcal{O}_p(T^{-1/2})$, we have $(\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X}) = \mathcal{O}_p(T^{-1/2}T) = \mathcal{O}_p(T^{1/2})$.

$(\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X})$.[16] Then,

$$\hat{\hat{\boldsymbol{\beta}}}-\boldsymbol{\beta} \approx \left\{(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X})(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} + \ldots\right\} \mathbf{X}'(\mathbf{W}^{-1}-\boldsymbol{\Delta_2})\mathbf{u}$$

or after deleting the terms of higher order of smallness than $\mathcal{O}_p\left(T^{-3/2}\right)$ gives[17]

$$\hat{\hat{\boldsymbol{\beta}}} - \boldsymbol{\beta} \approx \left\{(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Delta_2}\ \mathbf{X})(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\right\} \mathbf{X}'(\mathbf{W}^{-1} - \boldsymbol{\Delta_2})\mathbf{u}. \tag{2.23}$$

Rearranging terms and neglecting those of higher order of smallness gives

$$\hat{\hat{\boldsymbol{\beta}}} - \boldsymbol{\beta} \approx \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} + \boldsymbol{\Delta_3}, \tag{2.24}$$

where

$$\boldsymbol{\Delta_3} = -(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{u} + (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X}(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{u}. \tag{2.25}$$

[16] To use this result we need to show $||(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X})|| < 1$. Since $\underset{T\to\infty}{lim} \frac{1}{T}\mathbf{X}'\mathbf{X} = \mathbf{Q}_{ij}$ for all $i, j = 1, 2, \ldots, M$ and $\mathbf{W}^{-1}$ is positive definite, then $\underset{T\to\infty}{lim} \frac{1}{T}\mathbf{X}'\mathbf{W}^{-1}\mathbf{X} = \mathbf{G}$, where $\mathbf{G}$ is a $(\bar{K} \times \bar{K})$ positive definite matrix that does not depend on T. Hence, $\underset{T\to\infty}{lim} \left(\frac{1}{T}\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\right)^{-1} = \mathbf{G}^{-1}$, where $\mathbf{G}^{-1}$ is also $(\bar{K} \times \bar{K})$ positive definite. Then, $\underset{T\to\infty}{lim}\, T\left(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\right)^{-1} = \mathbf{G}^{-1}$ and thus $(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} = \mathcal{O}_p(T^{-1})$. Therefore, $(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X}) = \mathcal{O}_p(T^{-1}T^{1/2}) = \mathcal{O}_p(T^{-1/2})$ and thus $(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X}) = o_p(1)$. Consequently, $||(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X})|| < 1$ for sufficiently large T.

[17] Brockwell and Davis (1987, p. 192, Proposition 6.1.1) propose:

If X_n and Y_n, $n = 1, 2, \ldots$, are random variables defined on the same probability space and $a_n > 0$, $b_n > 0$, $n = 1, 2, \ldots$, then

(i) if $X_n = o_p(a_n)$ and $Y_n = o_p(b_n)$, we have $X_nY_n = o_p(a_nb_n)$, $X_n + Y_n = o_p(\max(a_n, b_n))$, and $|X_n| = o_p(a_n^r)$, for $r > 0$;

(ii) if $X_n = o_p(a_n)$ and $Y_n = \mathcal{O}_p(b_n)$, we have $X_nY_n = o_p(a_nb_n)$. Moreover

(iii) the statement (i) remains valid if o_p is everywhere replaced by $\mathcal{O}_p$.

Therefore, by Proposition 6.1.1 (i) and Definition 6.1.4 (ii) from Brockwell and Davis (1987),

$$\begin{aligned}
(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X}) &= \mathcal{O}_p(T^{-1}T^{1/2}) = \mathcal{O}_p(T^{-1/2}) \\
\left[(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X})\right]^2 &= \mathcal{O}_p(T^{-1/2}T^{-1/2}) = \mathcal{O}_p(T^{-1}).
\end{aligned}$$

Similarly, $\left[(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X})\right]^3 = \mathcal{O}_p(T^{-3/2})$, $\left[(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Delta_2}\mathbf{X})\right]^4 = \mathcal{O}_p(T^{-2})$, and so on.

According to Zellner (1962), the order of $\mathbf{X}'\boldsymbol{\Delta}_\mathbf{2}\mathbf{u}$ is equal to the order of $\boldsymbol{\Delta}_\mathbf{2}$ multiplied by the order of $\mathbf{X}'\mathbf{u}$.[18] Therefore, the order of $\mathbf{X}'\boldsymbol{\Delta}_\mathbf{2}\mathbf{u}$ is $\mathcal{O}_p(1)$.[19] In addition, according to Zellner (1962), the first term as a whole in the right had side of (2.25) is $\mathcal{O}_p(T^{-1})$,[20] and $\mathbf{X}'\mathbf{W}^{-1}\mathbf{u}$ is $\mathcal{O}_p(T^{1/2})$. Therefore, the second term as a whole in the right hand side of equation (2.25) is $\mathcal{O}_p(T^{-1})$.[21] Finally, $\boldsymbol{\Delta}_\mathbf{3}$ is $\mathcal{O}_p(T^{-1})$.[22]

Therefore, $\hat{\hat{\boldsymbol{\beta}}} = \hat{\boldsymbol{\beta}} + \mathcal{O}_p(T^{-1})$.

Property 2: $T^{1/2}(\hat{\hat{\boldsymbol{\beta}}} - \boldsymbol{\beta})$ and $T^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ have the same asymptotic normal distribution.

We use equation (2.23) and the facts that $(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} = \mathcal{O}_p(T^{-1})$, $\mathbf{X}'\boldsymbol{\Delta}_\mathbf{2}\mathbf{X} = \mathcal{O}_p(T^{1/2})$, $\mathbf{X}'\boldsymbol{\Delta}_\mathbf{2}\mathbf{u} = \mathcal{O}_p(1)$, and $E(\mathbf{u}) = \mathbf{0}$. If we drop the term $(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}$ $(\mathbf{X}'\boldsymbol{\Delta}_\mathbf{2}\mathbf{X})$ $(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} = \mathcal{O}_p(T^{-3/2})$ in equation (2.23) and then take expectation in both sides of the equation we get

$$\begin{aligned}
E(\hat{\hat{\boldsymbol{\beta}}} - \boldsymbol{\beta}) &\approx (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}E[\mathbf{X}'(\mathbf{W}^{-1} - \boldsymbol{\Delta}_\mathbf{2})\mathbf{u}] \\
&= (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}E[(\mathbf{X}'\mathbf{W}^{-1}\mathbf{u}) - (\mathbf{X}'\boldsymbol{\Delta}_\mathbf{2}\mathbf{u})] \\
&= (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}[\mathbf{X}'\mathbf{W}^{-1}E(\mathbf{u}) - E(\mathbf{X}'\boldsymbol{\Delta}_\mathbf{2}\mathbf{u})] \\
&= -(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}E(\mathbf{X}'\boldsymbol{\Delta}_\mathbf{2}\mathbf{u}) \\
&= \mathcal{O}_p(T^{-1}\,1) = \mathcal{O}_p(T^{-1}).
\end{aligned}$$

Similarly, we could have taken expectation in both sides of the equation (2.24) to get

$$E(\hat{\hat{\boldsymbol{\beta}}} - \boldsymbol{\beta}) \approx E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + E(\boldsymbol{\Delta}_\mathbf{3}).$$

According to Zellner (1962), $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ is $\mathcal{O}_p(T^{-1/2})$. Since $\boldsymbol{\Delta}_\mathbf{3}$ is $\mathcal{O}_p(T^{-1})$, then $bias\left(\hat{\hat{\boldsymbol{\beta}}}\right) =$

[18] Srivastava and Giles (1987, p. 42) explain $\mathbf{X}'\mathbf{u} = \mathcal{O}_p(T^{1/2})$.

[19] Since $\boldsymbol{\Delta}_\mathbf{2} = \mathcal{O}_p(T^{-1/2})$ and $\mathbf{X}'\mathbf{u} = \mathcal{O}_p(T^{1/2})$, then $(\mathbf{X}'\boldsymbol{\Delta}_\mathbf{2}\mathbf{u}) = \mathcal{O}_p(T^{-1/2}T^{1/2}) = \mathcal{O}_p(1)$.

[20] Therefore, $(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Delta}_\mathbf{2}\mathbf{u}) = \mathcal{O}_p(T^{-1}\,1) = \mathcal{O}_p(T^{-1})$. Srivastava and Giles (1987, p. 43, equation (3.7)) also explain $(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} = \mathcal{O}_p(T^{-1})$.

[21] Because $(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Delta}_\mathbf{2}\mathbf{X})(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}^{-1}\mathbf{u}) = \mathcal{O}_p(T^{-1}T^{1/2}T^{-1}T^{1/2}) = \mathcal{O}_p(T^{-1})$.

[22] By Proposition 6.1.1 (i) from Brockwell and Davis (1987), $\boldsymbol{\Delta}_\mathbf{3} = \mathcal{O}_p(\max(T^{-1}, T^{-1})) = \mathcal{O}_p(T^{-1})$.

$E(\hat{\hat{\boldsymbol{\beta}}} - \boldsymbol{\beta})$ is at most $\mathcal{O}_p(T^{-1/2})$.[23] Alternatively, using the result in the paragraph above, $bias\left(\hat{\hat{\boldsymbol{\beta}}}\right)$ is as well at most $\mathcal{O}_p(T^{-1})$. Furthermore, since $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ is $\mathcal{O}_p(T^{-1/2})$ and $\boldsymbol{\Delta_3}$ is $\mathcal{O}_p(T^{-1})$, the asymptotic covariance matrix of $\hat{\hat{\boldsymbol{\beta}}} - \boldsymbol{\beta}$ is the same as $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$. Finally, since under general conditions the asymptotic distribution of $T^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is multivariate normal, the asymptotic distribution of $T^{1/2}(\hat{\hat{\boldsymbol{\beta}}} - \boldsymbol{\beta})$ is the same as that of $T^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. This is because

$$\lim_{T\to\infty} \Pr\left[\left|T^{1/2}(\hat{\hat{\boldsymbol{\beta}}} - \boldsymbol{\beta}) - T^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right| > \varepsilon\right] = \lim_{T\to\infty} \Pr\left[\left|T^{1/2}\boldsymbol{\Delta_3}\right| > \varepsilon\right] = 0,$$

where ε is a very small quantity.[24]

Property 3: $var(\hat{\hat{\boldsymbol{\beta}}}) = (\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1} + o_p(T^{-1})$.

Intuitively, this property might follow from Property 1 and equation (2.8).

2.3 Efficiency Gain

We have shown that the generalized least squares estimator, $\hat{\boldsymbol{\beta}} = \left(\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2', \ldots, \hat{\boldsymbol{\beta}}_M'\right)'$, is different from applying least-squares to each equation, $\tilde{\boldsymbol{\beta}} = \left(\tilde{\boldsymbol{\beta}}_1', \tilde{\boldsymbol{\beta}}_2', \ldots, \tilde{\boldsymbol{\beta}}_M'\right)'$. As shown before, this difference occurs because the least-squares estimator assumes $E[\mathbf{u}_i\mathbf{u}_j] = \sigma_{ij}\mathbf{I}_T = \mathbf{0}_{(T\times T)}$ for $i \neq j$ while the generalized least squares estimator does not. That is, the least-squares estimator assumes errors from different equations in the same time period are uncorrelated.

To show how the generalized least squares estimator differs from the least-squares estimator, consider the example given by Zellner (1962, p. 354) where he supposed [25]

$$\sigma_{ij} = \begin{cases} \sigma_{ij} & = \sigma^2 \quad \text{if} \quad i = j \\ \sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}\,\rho & = \sigma^2\rho \quad \text{if} \quad i \neq j \end{cases} \quad \text{for } i, j = 1, 2, \ldots, M.$$

[23] $(\hat{\hat{\boldsymbol{\beta}}} - \boldsymbol{\beta}) = \mathcal{O}_p(\max(T^{-1/2}, T^{-1})) = \mathcal{O}_p(T^{-1/2})$.

[24] Zellner (1962) referred to the convergence theorem in Cramér (1946, p. 254).

[25] This example is only for illustrative purposes. For estimation purposes, $\boldsymbol{\Sigma_c}$ is estimated with $\mathbf{S_c}$.

In addition, Zellner (1962, p. 360) explains ρ is the correlation of contemporaneous disturbances between equations. The correlation between two contemporaneous random variables $u_i(t)$ and $u_j(t)$, is defines as $\rho = \frac{cov[u_i(t),u_j(t)]}{\sqrt{var[u_i(t)]}\sqrt{var[u_j(t)]}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$. Then, $\sigma_{ij} = \sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}\,\rho$, $i \neq j$ and $-1 \leq \rho \leq 1$.

As it can be seen in equation (2.4), this implies that

$$\underset{(M\times M)}{\boldsymbol{\Sigma_c}} = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \cdots & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 & \cdots & \sigma^2\rho \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2\rho & \sigma^2\rho & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

$$= \sigma^2 \left\{ \begin{bmatrix} 1-\rho & 0 & \cdots & 0 \\ 0 & 1-\rho & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1-\rho & \cdots & 1-\rho \end{bmatrix} + \rho \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \right\}$$

$$= \sigma^2 \left[(1-\rho)\mathbf{I}_M + \rho \mathbf{1}_M \mathbf{1}'_M\right],$$

where $\mathbf{1_M}$ is a $(M \times 1)$ vector whose elements are 1.

Then, $\underset{(M\times M)}{\boldsymbol{\Sigma_c^{-1}}} = [\varrho \mathbf{I}_M - \gamma \mathbf{1}_M \mathbf{1}'_M]$ with $\varrho = \dfrac{1}{\sigma^2(1-\rho)}$ and $\gamma = \dfrac{\varrho\rho}{[1+(M-1)\rho]} = \dfrac{\rho}{\sigma^2(1-\rho)\,[1+(M-1)\rho]}$. That is,[26]

$$\underset{(M\times M)}{\boldsymbol{\Sigma_c^{-1}}} = \begin{bmatrix} \frac{1+M\rho-2\rho}{\sigma^2(1-\rho)[1+(M-1)\rho]} & \frac{-\rho}{\sigma^2(1-\rho)[1+(M-1)\rho]} & \cdots & \frac{-\rho}{\sigma^2(1-\rho)[1+(M-1)\rho]} \\ \frac{-\rho}{\sigma^2(1-\rho)[1+(M-1)\rho]} & \frac{1+M\rho-2\rho}{\sigma^2(1-\rho)[1+(M-1)\rho]} & \cdots & \frac{-\rho}{\sigma^2(1-\rho)[1+(M-1)\rho]} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-\rho}{\sigma^2(1-\rho)[1+(M-1)\rho]} & \frac{-\rho}{\sigma^2(1-\rho)[1+(M-1)\rho]} & \cdots & \frac{1+M\rho-2\rho}{\sigma^2(1-\rho)[1+(M-1)\rho]} \end{bmatrix}.$$

Alternatively, a simplified expression is

$$\boldsymbol{\Sigma_c^{-1}} = \begin{bmatrix} \varrho & 0 & \cdots & 0 \\ 0 & \varrho & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varrho \end{bmatrix} - \begin{bmatrix} \gamma & \gamma & \cdots & \gamma \\ \gamma & \gamma & \cdots & \gamma \\ \vdots & \vdots & \ddots & \vdots \\ \gamma & \gamma & \cdots & \gamma \end{bmatrix} = \begin{bmatrix} (\varrho-\gamma) & -\gamma & \cdots & -\gamma \\ -\gamma & (\varrho-\gamma) & \cdots & -\gamma \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma & -\gamma & \cdots & (\varrho-\gamma) \end{bmatrix}.$$

[26] In addition, observe that when multiplying $\boldsymbol{\Sigma_c}$ with $\boldsymbol{\Sigma_c^{-1}}$:
$\frac{\sigma^2(1+M\rho-2\rho)}{\sigma^2(1-\rho)[1+(M-1)\rho]} - \frac{(M-1)\sigma^2\rho^2}{\sigma^2(1-\rho)[1+(M-1)\rho]} = 1$, and
$\frac{-\sigma^2\rho}{\sigma^2(1-\rho)[1+(M-1)\rho]} + \frac{(1+M\rho-2\rho)\sigma^2\rho}{\sigma^2(1-\rho)[1+(M-1)\rho]} - \frac{(M-2)\sigma^2\rho^2}{\sigma^2(1-\rho)[1+(M-1)\rho]} = 0.$

Then, the variance-covariance matrix of the estimator $\hat{\boldsymbol{\beta}}$ is:

$$\begin{aligned}
\underset{MK\times MK}{var(\hat{\boldsymbol{\beta}})} &= (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}')^{-1} \\
&= [\mathbf{X}'(\boldsymbol{\Sigma_c}^{-1}\otimes \mathbf{I}_T)\mathbf{X}]^{-1} \\
&= \begin{bmatrix} (\varrho-\gamma)\mathbf{X}_1'\mathbf{X}_1 & -\gamma\mathbf{X}_1'\mathbf{X}_2 & \cdots & -\gamma\mathbf{X}_1'\mathbf{X}_M \\ -\gamma\mathbf{X}_2'\mathbf{X}_1 & (\varrho-\gamma)\mathbf{X}_2'\mathbf{X}_2 & \cdots & -\gamma\mathbf{X}_2'\mathbf{X}_M \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma\mathbf{X}_M'\mathbf{X}_1 & -\gamma\mathbf{X}_M'\mathbf{X}_2 & \cdots & (\varrho-\gamma)\mathbf{X}_M'\mathbf{X}_M \end{bmatrix}^{-1}. \qquad (2.26)
\end{aligned}$$

Since $var(\hat{\boldsymbol{\beta}})$ is nonsingular, when $M = 2$, we can partition $var(\hat{\boldsymbol{\beta}})$ and solve for the leading sub-matrix, $var(\hat{\boldsymbol{\beta}}_1)$. According to Zellner (1962) this variance-covariance matrix is

$$var(\hat{\boldsymbol{\beta}}_1) = \left[(\varrho-\gamma)\mathbf{X}_1'\mathbf{X}_1 - \frac{\gamma^2}{\varrho-\gamma}\mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{X}_1\right]^{-1} \qquad (2.27)$$

and Zellner and Hwang (1962, p. 308) showed

$$\left|var(\hat{\boldsymbol{\beta}}_1)\right| = \frac{(1-\rho^2)^{K_1}}{\prod_{i=1}^{K_1}(1-\rho^2 r_i^2)}\left|\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}\right|, \qquad (2.28)$$

where K_1 is the number of independent variables in the first equation ($K_1 \leq K_2$), r_i is the i^{th} canonical correlation coefficient associated with the sets of variables in $\mathbf{X}_1$ and $\mathbf{X}_2$, and $|\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}|$ is the determinant of $\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}$ or the generalized variance of the "single-equation" least squares estimator of the coefficient vector of the first equation.[27] Since $0 \leq r_i^2 \leq 1$, the generalized variance of $\boldsymbol{\beta}_1$ will be smaller than the generalized variance of the "single-equation" least squares estimator of the coefficient vector of the first equation. That is,

$$\left|var(\hat{\boldsymbol{\beta}}_1)\right|_{M=2} = \underbrace{\frac{(1-\rho^2)^{K_1}}{\prod_{i=1}^{K_1}(1-\rho^2 r_i^2)}\left|\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}\right|}_{\text{Generalized Variance of } \hat{\boldsymbol{\beta}}_1} \leq \underbrace{\left|\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}\right|}_{\substack{\text{``Single-equation''} \\ \text{Generalized Variance of } \hat{\boldsymbol{\beta}}_1}} = \left|var(\hat{\boldsymbol{\beta}}_1)\right|_{M=1}.$$

[27]Suppose we want to estimate $\mathbf{y}_1 = \mathbf{X}_1\,\boldsymbol{\beta}_1 + \mathbf{u}_1$. Then, $var(\hat{\boldsymbol{\beta}}_1) = (\mathbf{X}_1'\mathbf{W}^{-1}\mathbf{X}_1)^{-1} = [\mathbf{X}_1'(\sigma_{11}\mathbf{I}_{\mathbf{K_1}})^{-1}\mathbf{X}_1]^{-1} = \sigma_{11}(\mathbf{X}_1'\mathbf{X}_1)^{-1} = \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}$.

Now, suppose $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$ (which implies $r_i = 0$ for all i). Plugging $r_i = 0$ for all i in equation (2.28) gives $\left|var(\hat{\boldsymbol{\beta}}_1)\right| = (1-\rho^2)^{K_1} \left|\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}\right|$. Since in general $\mathbf{X}_1'\mathbf{X}_2 \neq \mathbf{0}$, this latter equation represents the minimum value $\left|var(\hat{\boldsymbol{\beta}}_1)\right|$ can take given σ^2 and ρ (as stated at the beginning of this section).

Similarly, as in the two-dimensional case, in the M-dimensional case, the generalized variance of $\hat{\boldsymbol{\beta}}_1$ will be smaller than the "single-equation" least squares generalized variance of $\hat{\boldsymbol{\beta}}_1$. In addition, in the M-dimensional case, the generalized variance of $\hat{\boldsymbol{\beta}}_1$ is smaller than the least squares variance of $\tilde{\boldsymbol{\beta}}_1$. The assumption $cov(\mathbf{u}_i, \mathbf{u}_j) = E\{[\mathbf{u}_i - E(\mathbf{u}_i)][\mathbf{u}_j - E(\mathbf{u}_j)]'\} = E[\mathbf{u}_i\mathbf{u}_j'] = \sigma_{ij}\mathbf{I}_T = \mathbf{0}_{(T\times T)}$ for $i \neq j$ is needed in order for the Gauss-Markov theorem to apply to the least squares model in the M-dimensional case.[28]

However, when we consider the M dimensional case with $\mathbf{X}_i'\mathbf{X}_j = 0$ for $i \neq j$, $i,\ j = 1, 2, \ldots, M$, as long as $var(\hat{\boldsymbol{\beta}})$ is nonsingular, equation (2.26) reduces to

$$
\begin{aligned}
var(\hat{\boldsymbol{\beta}}) &= \begin{bmatrix} (\varrho-\gamma)\mathbf{X}_1'\mathbf{X}_1 & 0 & \cdots & 0 \\ 0 & (\varrho-\gamma)\mathbf{X}_2'\mathbf{X}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\varrho-\gamma)\mathbf{X}_M'\mathbf{X}_M \end{bmatrix}^{-1} \\
&= \begin{bmatrix} [(\varrho-\gamma)\mathbf{X}_1'\mathbf{X}_1]^{-1} & 0 & \cdots & 0 \\ 0 & [(\varrho-\gamma)\mathbf{X}_2'\mathbf{X}_2]^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [(\varrho-\gamma)\mathbf{X}_M'\mathbf{X}_M]^{-1} \end{bmatrix}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
var(\hat{\boldsymbol{\beta}}_1) &= \frac{1}{(\varrho-\gamma)}(\mathbf{X}_1'\mathbf{X}_1)^{-1} = \frac{1}{\left(\frac{1}{\sigma^2(1-\rho)}\right) - \left(\frac{\rho}{\sigma^2(1-\rho)[1+(M-1)\rho]}\right)}(\mathbf{X}_1'\mathbf{X}_1)^{-1} \\
&= \frac{(1-\rho)\sigma^2}{1-\frac{\rho}{[1+(M-1)\rho]}}(\mathbf{X}_1'\mathbf{X}_1)^{-1} = \left[\frac{1-\rho}{1-\frac{\rho}{1+\rho(M-1)}}\right]\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}
\end{aligned}
$$

[28] If this assumption does not hold, the error variance-covariance matrix of the least-squares model involves a special form of heteroskedasticity and autocorrelation (Griffiths et al., 1992, p. 551).

Now, as M approaches infinity,

$$\begin{aligned}\lim_{M\to\infty} var(\hat{\boldsymbol{\beta}}_1) &= \left[\frac{1-\rho}{1-\frac{\rho}{1+\rho}\lim_{M\to\infty}\frac{1}{(M-1)}}\right]\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}, \\ var(\hat{\boldsymbol{\beta}}_1) &= [1-\rho]\,\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}.\end{aligned}$$

2.4 Test for Aggregation Bias

2.4.1 Micro-Data, Macro-Data and the Aggregation Problem

Micro-data refers to data that is not aggregated. For example, data on individuals, such as firms or consumers (Theil, 1954, p. 2). However, as is frequently the case, data are available only for aggregates of consumers and firms. Macro-data refers to data that has being aggregated over individuals, commodities or time.[29] The transition from micro-economics of individuals to the analysis of economic aggregates is referred as the "aggregation problem." As explained by Deaton and Muellbauer (1980, p. 148), "[a]ggregation is seen as a nuisance, a temporary obstacle lying in the way of a straightforward application of the theory to the data. In this view, the role of aggregation theory is to provide the necessary conditions under which it is possible to treat aggregate consumer behavior as if it were the outcome of the decision of a ['representative'] consumer."[30]

As explained by Theil (1954, p. 2), the relations postulated by the economic theory of individual households (the micro-theory) are called micro-relations or micro-equations. The micro-equations are composed of micro-variables and micro-parameters. Aggregation implies that micro-variables are replaced by aggregates or macro-variables. Similar to the micro-theory, the macro-theory postulates that macro-variables are connected by macro-relations or macro-equations. The macro-equations are composed of macro-variables and macro-parameters.

Theil (1954, p. 3) distinguishes between three types of aggregation: aggregation over individuals, such as firms and consumers, aggregation over several sets of com-

[29] In general, aggregated data is a function of micro-data.

[30] In this example, the aggregate consumer behavior is the average behavior of all consumers.

modities, and aggregation over time. Consider the following examples provided by Theil (1954, pp. 1–4) for each of these cases.

a) Aggregation over individuals.

> Consider the sugar consumption by families of some country during a certain period. Suppose that according to economic theory each family's demand for sugar is a function of its income, of the price of sugar and of the number of family members. Suppose furthermore that the empirical research worker decides to consider total sugar consumption as a function of total personal income, the price of sugar and the population size.

b) Aggregation over commodities.

> Consider for instance, an entrepreneur who uses several factors of production. Suppose that his demand for each of these factors depends on the level of production and on the price of this factor. Suppose also that an econometrician wants to combine some of these factors to a group and that he considers an input index of this group as a function of the level of production and of an input price index.

c) Aggregation over time periods.

> Suppose that an entrepreneur bases his demand for labour on the quantity of products sold during past periods. More specifically, we assume that the demand for labour during a certain month depends on the quantity sold during the preceding four months. Suppose furthermore that only quarterly sales figures are available. Then an obvious procedure is to aggregate monthly periods to quarterly periods and to postulate that the demand for labour during a quarterly period is a function of the quantity sold during the same and the preceding quarterly periods.

2.4.2 Testing with Micro-Data

Suppose micro-data is available for $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M$ and $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M$. In addition, suppose that each $\mathbf{X}_i$, $i = 1, 2, \ldots, M$, is a $(T \times K)$ matrix containing observations on K independent variables.[31] For instance, consider a general version of the

[31] That is, $K_1 = K_2 = \ldots = K_M = K$.

investment example provided by Zellner (1962, pp. 357–362). Suppose $\mathbf{y}_1$ is a $(T \times 1)$ vector of observations on firm one's current gross investment, $\mathbf{y}_2$ is a $(T \times 1)$ vector of observations on firm two's current gross investment, ..., $\mathbf{y}_M$ is a $(T \times 1)$ vector of observations on firm M's current gross investment. Then, $\mathbf{X}_1$ is a matrix of observations on the K independent variables affecting $\mathbf{y}_1$. For example, $\mathbf{X}_{11}$ is a $(T \times 1)$ vector of observations on firm one's capital stock, $\mathbf{X}_{12}$ is a $(T \times 1)$ vector of observations on firm one's outstanding shares, ..., $\mathbf{X}_{1K}$ is a $(T \times 1)$ vector of observations on firm one's k^{th} independent variable affecting gross investment. Similarly, $\mathbf{X}_2$ is a matrix of observation on the K independent variables affecting $\mathbf{y}_2$. So, $\mathbf{X}_{21}$ is a $(T \times 1)$ vector of observations on firm two's capital stock, $\mathbf{X}_{22}$ is a $(T \times 1)$ vector of observations on firm two's outstanding shares, ..., $\mathbf{X}_{2K}$ is a $(T \times 1)$ vector of observations on firm two's k^{th} independent variable affecting gross investment. Similar for $\mathbf{X}_3$ to $\mathbf{X}_M$. You would like to know whether you can work with aggregated data. Or equivalently, you would like to know whether the M firms' current gross investment react in the same way to changes in its capital stock, outstanding shares, and so on. That is, you would like to know if the M firms are characterized by the same regression parameters. If they do, then you can aggregate your micro-data without suffering from aggregation bias. Therefore, you need to test

$$\begin{aligned} &H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ldots = \boldsymbol{\beta}_M \\ &H_a : \text{at least one } \boldsymbol{\beta}_i \neq \boldsymbol{\beta}_j,\ i \neq j,\ i,j = 1,2,\ldots,M. \end{aligned} \tag{2.29}$$

Or equivalently,

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \begin{bmatrix} \mathbf{I}_K & -\mathbf{I}_K & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K & -\mathbf{I}_K & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_K & -\mathbf{I}_K & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{I}_K & -\mathbf{I}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \tag{2.30}$$

H_a : at least one $\boldsymbol{\beta}_i \neq \boldsymbol{\beta}_j,\ i \neq j,\ i,j = 1,2,\ldots,M,$

where $\mathbf{C}$ is $[(M-1)K \times MK]$ matrix, $\boldsymbol{\beta}$ is a $(MK \times 1)$ vector, $\mathbf{I}_K$ is a $(K \times K)$ identity matrix, and $\mathbf{0}$ is a $[(M-1)K \times 1]$ zero matrix. Therefore, there are $(M-1)K$ restrictions under H_0.

According to Zellner (1962), if H_0 is true, there will be no aggregation bias involved in the simple linear aggregation.[32] To show this is the case, suppose we estimate $\bar{\boldsymbol{\beta}}$ in

$$\bar{\mathbf{y}} = \bar{\mathbf{X}}\bar{\boldsymbol{\beta}} + \bar{\mathbf{u}}, \tag{2.31}$$

where $\bar{\mathbf{y}} = \sum_{i=1}^{M} \mathbf{y}_i$ is a $(T \times 1)$ vector of observations on the dependent variable, $\bar{\mathbf{X}} = \sum_{i=1}^{M} \mathbf{X}_i$ is a $(T \times K)$ matrix of observations on the K independent variables, $\bar{\boldsymbol{\beta}}$ is the $(K \times 1)$ vector of parameters to be estimated, and $\bar{\mathbf{u}} = \sum_{i=1}^{M} \mathbf{u}_i$ is a $(T \times 1)$ vector of disturbances.[33]

The least-squares estimator of $\bar{\boldsymbol{\beta}}$ is given by

$$\underset{(K\times 1)}{\tilde{\bar{\boldsymbol{\beta}}}} = (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\bar{\mathbf{y}}.$$

Taking expected value in both sides of the equation gives

$$\begin{aligned}
E\left(\tilde{\bar{\boldsymbol{\beta}}}\right) &= E\left[(\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\bar{\mathbf{y}}\right] = E\left[(\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\left(\sum_{i=1}^{M}\mathbf{y}_i\right)\right] \\
&= E\left[(\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\left(\sum_{i=1}^{M}(\mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{u}_i)\right)\right] \\
&= E\left[(\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\left(\sum_{i=1}^{M}\mathbf{X}_i\boldsymbol{\beta}_i\right) + (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\left(\sum_{i=1}^{M}\mathbf{u}_i\right)\right] \\
&= \sum_{i=1}^{M}(\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'(\mathbf{X}_i\boldsymbol{\beta}_i) + (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\sum_{i=1}^{M}E(\mathbf{u}_i) \\
&= \sum_{i=1}^{M}(\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'(\mathbf{X}_i\boldsymbol{\beta}_i).
\end{aligned} \tag{2.32}$$

---

[32] Zellner (1962) followed Kloek's (1960) matrix representation of the aggregation problem.

[33] We could also write equation (2.31) as

$$\underset{(T\times 1)}{\bar{\mathbf{y}}} = \begin{pmatrix} \bar{\mathbf{X}}_1 & \bar{\mathbf{X}}_2 & \cdots & \bar{\mathbf{X}}_K \end{pmatrix}_{(T\times K)} \begin{pmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \\ \vdots \\ \bar{\beta}_K \end{pmatrix}_{(K\times 1)} + \underset{(T\times 1)}{\bar{\mathbf{u}}}.$$

Thus, under $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ldots = \boldsymbol{\beta}_M = \bar{\boldsymbol{\beta}}$,

$$\begin{aligned} E\left(\tilde{\bar{\boldsymbol{\beta}}}\right) &= (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\left(\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \ldots + \mathbf{X}_M\boldsymbol{\beta}_M\right) \\ &= (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\left(\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_M\right)\bar{\boldsymbol{\beta}} \\ &= (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\bar{\mathbf{X}}\bar{\boldsymbol{\beta}} \\ &= \bar{\boldsymbol{\beta}} \\ &= \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ldots = \boldsymbol{\beta}_M. \end{aligned}$$

That is, under H_0, the expectation of macro-estimator, $\tilde{\bar{\boldsymbol{\beta}}}$, will be equal to the micro-parameter vector $\bar{\boldsymbol{\beta}}$, where $\bar{\boldsymbol{\beta}} = \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ldots = \boldsymbol{\beta}_M$.

The hypotheses in (2.29) or (2.30) can be tested using the model in equation (2.3) and employing Roy's (1957) F-statistic[34]

$$\begin{aligned} F_{q,MT-MK} &= \frac{MT - MK}{q} \\ &\times \frac{\mathbf{y}'\mathbf{W}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{y}}{\mathbf{y}'\mathbf{W}^{-1}\mathbf{y} - \mathbf{y}'\mathbf{W}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{y}}. \end{aligned} \tag{2.33}$$

Now, replacing $\mathbf{W}^{-1}$ with $\hat{\mathbf{W}}^{-1}$ gives

$$\begin{aligned} \tilde{F}_{q,MT-MK} &= \frac{MT - MK}{q} \\ &\times \frac{\mathbf{y}'\hat{\mathbf{W}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}(\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{y}}{\mathbf{y}'\hat{\mathbf{W}}^{-1}\mathbf{y} - \mathbf{y}'\hat{\mathbf{W}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{y}}, \end{aligned} \tag{2.34}$$

where $q = (M-1)K$ is the number of restrictions under H_0 and $\hat{\mathbf{W}}^{-1}$ is given in equation (2.11).

[34] Roy (1957, p. 82) derived the F-statistic without involving the likelihood-ratio approach. However, determining how the F-statistic is distributed requires the assumption that $\bar{\mathbf{u}}$ is normally distributed. Zellner (1962, Appendix A) showed the likelihood-ratio approach asymptotically leads to the same test statistic. Similar to Roy (1957), Zellner (1962, Appendix A) assumed normality of $\mathbf{u}$ when formulating the likelihood function for the null and alternative hypotheses of the system in equations (2.2) or (2.3).

Zellner (1962, Appendix B) showed that

$$\tilde{F} = F_{q,MT-MK} + \mathcal{O}_p((MT)^{-1/2}) \tag{2.35}$$

and made use of one of Cramér's (1946, p. 254) theorem to conclude that $\tilde{F}$ has the same asymptotic distribution as F. Therefore, we reject H_0 if $\tilde{F} > F^*_{q,MT-MK}(\alpha)$ with at most $\alpha\,100\%$ probability of Type I error. The quantity $F^*_{q,MT-MK}(\alpha)$ is a critical value from an F-distribution with q degrees of freedom in the numerator, $(MT - MK)$ degrees of freedom in the denominator, and level α.[35]

Zellner (1962, Appendix A) also showed that $-2\log\lambda = q\,F_{q,MT-MK} + \mathcal{O}_p((MT)^{-1})$, where λ is the likelihood ratio for testing H_0 in (2.29).[36] Then Zellner (1962) used results from Mood (1950, p. 259) and Wilks (1943, p. 151) to conclude that $-2\log\lambda$, $q\,F_{q,MT-MK}$ and $q\,\tilde{F}_{q,MT-MK}$ are asymptotically distributed as χ^2_q. Therefore, we could also reject H_0 if $\left|q\,\tilde{F}\right| > \chi^2_q(\alpha)$ with level α or alternatively reject H_0 if $-2\log\lambda > \chi^2_q(\alpha)$ with level α.[37]

Zellner (1962) emphasized that for small samples we could proceed in two ways. One way is to compute $q\,\tilde{F}$ and assume it is asymptotically distributed as χ^2_q. Another way is to compute $\tilde{F}$ and assume is closely distributed as $F^*_{q,MT-MK}$.[38]

2.4.3 Testing with Macro-Data

Now, suppose only macro-data is available for $\bar{\mathbf{y}}$ and $\bar{\mathbf{X}}$. For example, you only have information on the total current gross investment by the M firms and the $\bar{\mathbf{X}}_{(T\times K)}$ matrix of observations in the K independent aggregated variables affecting $\bar{\mathbf{y}}$. That is, $\bar{\mathbf{X}}_1$ is a $(T \times 1)$ vector of observations on total capital stock of the M firms, $\bar{\mathbf{X}}_2$ is

[35] I.e., $\alpha = Pr_{\,\nu_1,\nu_2}\left(\tilde{F} > F^*_{\nu_1,\nu_2}(\alpha)\right)$.

[36] I.e., $\lambda = \dfrac{\sup\limits_{H_0} L(\boldsymbol{\beta},\sigma^2)}{\sup\limits_{H_0\cup H_a} L(\boldsymbol{\beta},\sigma^2)}$.

[37] Now, $\alpha = Pr_{\,\nu}\left(\chi^2 > \chi^2_\nu(\alpha)\right)$.

[38] Zellner (1962, pp. 357–362) provided an example of testing for aggregation bias with micro-data for a system of two equations ($M = 2$) with two independent variables ($K = 2$) when the sample size is small ($T = 20$). In his example, he showed how both procedures lead to the same conclusion.

a $(T \times 1)$ vector of observations on total outstanding shares of the M firms, ..., $\bar{\mathbf{X}}_K$. You would like to know whether aggregation bias is present in your aggregated data. Hence, you are going to test the hypotheses in (2.29), but employing your macro-data.

Suppose we have a system of M equations, each with K_i independent variables and an intercept. Let's consider the t^{th} observation from each equation. That is,

$$\begin{aligned}
y_1(t) &= \beta_{10} + \beta_{11}x_{11}(t) + \beta_{12}x_{12}(t) + \cdots + \beta_{1K_1}x_{1K_1}(t) + u_1(t) \\
y_2(t) &= \beta_{20} + \beta_{21}x_{21}(t) + \beta_{22}x_{22}(t) + \cdots + \beta_{2K_2}x_{2K_2}(t) + u_2(t) \\
&\vdots \\
y_M(t) &= \beta_{M0} + \beta_{M1}x_{M1}(t) + \beta_{M2}x_{M2}(t) + \cdots + \beta_{MK_M}x_{MK_M}(t) + u_M(t)\,.
\end{aligned}$$

Since $K_1 = K_2 = \ldots = K_M$, we drop the subscript and only use K. Now we proceed to aggregate the data. That is,

$$\begin{aligned}
(y_1(t) + y_2(t) + \ldots + y_M(t)) &= (\beta_{10} + \beta_{20} + \ldots + \beta_{M0}) \\
&\quad + (\beta_{11}x_{11}(t) + \beta_{21}x_{21}(t) + \ldots + \beta_{M1}x_{M1}(t)) \\
&\quad + (\beta_{12}x_{12}(t) + \beta_{22}x_{22}(t) + \ldots + \beta_{M2}x_{M2}(t)) \\
&\qquad \vdots \\
&\quad + (\beta_{1K}x_{1K}(t) + \beta_{2K}x_{2K}(t) + \ldots + \beta_{MK}x_{MK}(t)) \\
&\quad + (u_1(t) + u_2(t) + \ldots + u_M(t))\,.
\end{aligned}$$

Now letting $\bar{y}(t) = \sum_{i=1}^{M} y_i(t)$, $\beta_0 = \sum_{i=1}^{M} \beta_{i0}$, $\bar{x}_k(t) = \sum_{j=1}^{M} x_{jk}(t)$, and $\bar{u}(t) = \sum_{i=1}^{M} u_i(t)$ for $k = 1, 2, \ldots, K$ and adjusting terms gives

$$\begin{aligned}
\bar{y}(t) &= \beta_0 + \left[\frac{\beta_{11}x_{11}(t) + \beta_{21}x_{21}(t) + \ldots + \beta_{M1}x_{M1}(t)}{x_{11}(t) + x_{21}(t) + \ldots + x_{M1}(t)}\right]\bar{x}_1(t) \\
&\quad + \left[\frac{\beta_{12}x_{12}(t) + \beta_{22}x_{22}(t) + \ldots + \beta_{M2}x_{M2}(t)}{x_{12}(t) + x_{22}(t) + \ldots + x_{M2}(t)}\right]\bar{x}_2(t) \\
&\qquad \vdots \\
&\quad + \left[\frac{\beta_{1K}x_{1K}(t) + \beta_{2K}x_{2K}(t) + \ldots + \beta_{MK}x_{MK}(t)}{x_{1K}(t) + x_{2K}(t) + \ldots + x_{MK}}(t)\right]\bar{x}_K + \bar{u}(t).
\end{aligned}$$

Now, we expand terms in the brackets to obtain

$$
\begin{aligned}
\bar{y}(t) \;\; = \;\; & \beta_0 + \beta_{11}\left[\frac{x_{11}(t)}{\sum\limits_{j=1}^{M} x_{j1}(t)}\right]\bar{x}_1(t) + \ldots + \beta_{M1}\left[\frac{x_{M1}(t)}{\sum\limits_{j=1}^{M} x_{j1}(t)}\right]\bar{x}_1(t) \\
& + \beta_{12}\left[\frac{x_{12}(t)}{\sum\limits_{j=1}^{M} x_{j2}(t)}\right]\bar{x}_2(t) + \ldots + \beta_{M2}\left[\frac{x_{M2}(t)}{\sum\limits_{j=1}^{M} x_{j2}(t)}\right]\bar{x}_2(t) \\
& \vdots \\
& + \beta_{1K}\left[\frac{x_{1K}(t)}{\sum\limits_{j=1}^{M} x_{jK}(t)}\right]\bar{x}_K(t) + \ldots + \beta_{MK}\left[\frac{x_{MK}(t)}{\sum\limits_{j=1}^{M} x_{jK}(t)}\right]\bar{x}_K(t) \\
& + \bar{u}(t).
\end{aligned}
$$

Now, let

$$
w_{ik}(t) = \frac{x_{ik}(t)}{\sum\limits_{j=1}^{M} x_{jk}(t)}, \quad i = 1, 2, \ldots, M, \quad k = 1, 2, \ldots, K.
$$

Note that,

$$
w_{Mk}(t) = 1 - \sum_{i=1}^{M-1} w_{ik}(t), \quad k = 1, 2, \ldots, K.
$$

Then,

$$\begin{aligned}
\bar{y}(t) &= \beta_0 + \beta_{11}w_{11}(t)\bar{x}_1(t) + \ldots + \beta_{(M-1)1}w_{(M-1)1}\bar{x}_1(t) + \beta_{M1}\left(1 - \sum_{i=1}^{(M-1)} w_{i1}(t)\right)\bar{x}_1(t) \\
&\quad +\beta_{12}w_{12}(t)\bar{x}_2(t) + \ldots + \beta_{(M-1)2}w_{(M-1)2}\bar{x}_2(t) + \beta_{M2}\left(1 - \sum_{i=1}^{(M-1)} w_{i2}(t)\right)\bar{x}_2(t) \\
&\qquad\qquad \vdots \\
&\quad +\beta_{1K}w_{1K}(t)\bar{x}_K(t) + \ldots + \beta_{(M-1)K}w_{(M-1)k}\bar{x}_K(t) + \beta_{MK}\left(1 - \sum_{i=1}^{(M-1)} w_{iK}(t)\right)\bar{x}_K(t) \\
&\quad +\bar{u}(t) \\
&= \beta_0 + (\beta_{11} - \beta_{M1})\, w_{11}(t)\bar{x}_1(t) + \ldots + \left(\beta_{(M-1)1} - \beta_{M1}\right) w_{(M-1)1}\bar{x}_1(t) + \beta_{M1}\bar{x}_1(t) \\
&\quad + (\beta_{12} - \beta_{M2})\, w_{12}(t)\bar{x}_2(t) + \ldots + \left(\beta_{(M-1)2} - \beta_{M2}\right) w_{(M-1)2}\bar{x}_2(t) + \beta_{M2}\bar{x}_2(t) \\
&\qquad\qquad \vdots \\
&\quad + (\beta_{1K} - \beta_{MK})\, w_{1K}(t)\bar{x}_K(t) + \ldots + \left(\beta_{(M-1)K} - \beta_{MK}\right) w_{(M-1)k}\bar{x}_K(t) + \beta_{MK}\bar{x}_K(t) \\
&\quad +\bar{u}(t) \qquad (2.36) \\
&= \beta_0 + \beta_{M1}\bar{x}_1(t) + \left[\sum_{i=1}^{(M-1)} (\beta_{i1} - \beta_{M1})\, w_{i1}(t)\right]\bar{x}_1(t) \\
&\quad +\beta_{M2}\bar{x}_2(t) + \left[\sum_{i=1}^{(M-1)} (\beta_{i2} - \beta_{M2})\, w_{i2}(t)\right]\bar{x}_2(t) \\
&\qquad\qquad \vdots \\
&\quad +\beta_{MK}\bar{x}_K(t) + \left[\sum_{i=1}^{(M-1)} (\beta_{iK} - \beta_{MK})\, w_{iK}(t)\right]\bar{x}_K(t) \\
&\quad +\bar{u}(t). \qquad (2.37)
\end{aligned}$$

Now, let

$$\underset{(T\times 1)}{\bar{\mathbf{X}}_1} = \begin{bmatrix} \bar{x}_1(1) \\ \bar{x}_1(2) \\ \vdots \\ \bar{x}_1(T) \end{bmatrix}, \quad \underset{(T\times 1)}{\bar{\mathbf{X}}_2} = \begin{bmatrix} \bar{x}_2(1) \\ \bar{x}_2(2) \\ \vdots \\ \bar{x}_2(T) \end{bmatrix}, \quad \ldots, \quad \underset{(T\times 1)}{\bar{\mathbf{X}}_K} = \begin{bmatrix} \bar{x}_K(1) \\ \bar{x}_K(2) \\ \vdots \\ \bar{x}_K(T) \end{bmatrix},$$

$$\underset{(T\times T)}{\bar{\mathbf{w}}_{ik}} = \begin{bmatrix} \bar{w}_{ik}(1) & 0 & \dots & 0 \\ 0 & \bar{w}_{ik}(2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \bar{w}_{ik}(T) \end{bmatrix},$$

where $i = 1, 2, \dots, (M-1)$ and $k = 1, 2, \dots, K$.

Hence, if data is available on $w_{i1}(t)$, $w_{i2}(t)$, ..., $w_{iK}(t)$ for $i = 1, 2, \dots, M$ and $t = 1, 2, \dots, T$,[39] it is possible to do a least-squares regression of $\bar{\mathbf{y}}$ on $\mathbf{w}_{11}\bar{\mathbf{X}}_1$, ..., $\mathbf{w}_{(M-1)1}\bar{\mathbf{X}}_1$, $\bar{\mathbf{X}}_1$, $\mathbf{w}_{12}\bar{\mathbf{X}}_2$, ..., $\mathbf{w}_{(M-1)2}\bar{\mathbf{X}}_2$, $\bar{\mathbf{X}}_2$, ..., $\mathbf{w}_{1K}\bar{\mathbf{X}}_K$, ..., $\mathbf{w}_{(M-1)K}\bar{\mathbf{X}}_K$, $\bar{\mathbf{X}}_K$ for equation (2.36) and perform K different F tests involving equality of the coefficients of equation (2.36). These K different F-tests are:

$$\begin{aligned} &H_{01} : (\beta_{11} - \beta_{M1}) = \dots = \left(\beta_{(M-1)1} - \beta_{M1}\right) = 0 \\ &H_{a1} : \text{at least one } \beta_{i1} \neq \beta_{j1},\ i \neq j,\ i, j = 1, 2, \dots, M \\ &\qquad\qquad\vdots \\ &H_{0K} : (\beta_{iK} - \beta_{MK}) = \dots = \left(\beta_{(M-1)K} - \beta_{MK}\right) = 0 \\ &H_{aK} : \text{at least one } \beta_{iK} \neq \beta_{jK},\ i \neq j,\ i, j = 1, 2, \dots, M. \end{aligned}$$

Or equivalently,

$$\begin{aligned} &H_{01} : \text{There is no Aggregation Bias in } \bar{\mathbf{X}}_1 \\ &H_{a1} : \text{There is Aggregation Bias in } \bar{\mathbf{X}}_1 \\ &\qquad\qquad\vdots \\ &H_{0K} : \text{There is no Aggregation Bias in } \bar{\mathbf{X}}_K \\ &H_{aK} : \text{There is Aggregation Bias in } \bar{\mathbf{X}}_K. \end{aligned}$$

If we consider the k^{th} F-test, notice that we are testing whether the M firms are affected in the same way by the $\mathbf{X}_{1k}, \mathbf{X}_{2k}, \dots, \mathbf{X}_{Mk}$ independent variables. That is,

[39] Zellner (1962) provided the following example. Suppose you have aggregated sales of M different firms from one particular industry at time year t. Consider the aggregated sales of firm one, $\bar{x}_1(t) = x_{11}(t) + x_{21}(t) + \dots + x_{M1}(t)$. Then, $w_{11}(t)$ might be firm one's proportion of industry sales in year t, $w_{21}(t)$ is firm two's proportion of industry sales in year t, ..., $w_{M1}(t)$ is firm M's proportion of industry sales in year t.

we are testing if the M firms can be characterized by a common regression parameter that corresponds to the independent variables $\mathbf{X}_{1k}, \mathbf{X}_{2k}, \ldots, \mathbf{X}_{Mk}$. If they can, then there is no aggregation problem but if they can not, then not all firms react in the same way to the $\mathbf{X}_{1k}, \mathbf{X}_{2k}, \ldots, \mathbf{X}_{Mk}$ independent variables; therefore, it is incorrect to aggregate the data.

In order to understand how we can apply one of these K F-tests, suppose we would like to perform the first test. For simplicity and convenience, we let $\pi_{11} = (\beta_{11} - \beta_{M1})$, $\ldots$, $\pi_{(M-1)1} = (\beta_{(M-1)1} - \beta_{M1})$, $\pi_{M1} = \beta_{M1}$, $\pi_{12} = (\beta_{12} - \beta_{M2})$, $\ldots$, $\pi_{(M-1)2} = (\beta_{(M-1)2} - \beta_{M2})$, $\pi_{M2} = \beta_{M2}$, $\pi_{1K} = (\beta_{1K} - \beta_{MK})$, $\ldots$, $\pi_{(M-1)K} = (\beta_{(M-1)K} - \beta_{MK})$, $\pi_{MK} = \beta_{MK}$. Hence, we can write the first test as

$$\begin{aligned} &H_{01} : \pi_{11} = \ldots = \pi_{(M-1)1} = 0 \\ &H_{a1} : \text{at least one } \pi_{i1} \neq 0, \ i = 1, 2, \ldots, (M-1). \end{aligned}$$

Then, if H_{01} is false, we can rewrite equation (2.36) as: [40]

[40] Equation (2.36) can also be written as:

$$\begin{aligned} \bar{\mathbf{y}} \quad = \quad & \beta_0 \mathbf{1}_T + \pi_{11}\mathbf{w}_{11}\bar{\mathbf{X}}_1 + \ldots + \pi_{(M-1)1}\mathbf{w}_{(M-1)1}\bar{\mathbf{X}}_1 + \pi_{M1}\bar{\mathbf{X}}_1 \\ & + \pi_{12}\mathbf{w}_{12}\bar{\mathbf{X}}_2 + \ldots + \pi_{(M-1)2}\mathbf{w}_{(M-1)2}\bar{\mathbf{X}}_2 + \pi_{M2}\bar{\mathbf{X}}_{M2} \\ & \quad \vdots \\ & + \pi_{1K}\mathbf{w}_{1K}\bar{\mathbf{X}}_K + \ldots + \pi_{(M-1)K}\mathbf{w}_{(M-1)K}\bar{\mathbf{X}}_K + \pi_{MK}\bar{\mathbf{X}}_K + \bar{\mathbf{u}}. \end{aligned}$$

$$
\begin{aligned}
\underset{(T\times 1)}{\bar{\mathbf{y}}} &= \begin{bmatrix} \left[\mathbf{1}_T\right]' \\ ------- \\ \left[\mathbf{w}_{11}\bar{\mathbf{X}}_1\right]' \\ \vdots \\ \left[\mathbf{w}_{(M-1)1}\bar{\mathbf{X}}_1\right]' \\ \left[\bar{\mathbf{X}}_1\right]' \\ ------- \\ \left[\mathbf{w}_{12}\bar{\mathbf{X}}_2\right]' \\ \vdots \\ \left[\mathbf{w}_{(M-1)2}\bar{\mathbf{X}}_2\right]' \\ \left[\bar{\mathbf{X}}_2\right]' \\ ------- \\ \vdots \\ ------- \\ \left[\mathbf{w}_{1K}\bar{\mathbf{X}}_K\right]' \\ \vdots \\ \left[\mathbf{w}_{(M-1)K}\bar{\mathbf{X}}_K\right]' \\ \left[\bar{\mathbf{X}}_K\right]' \end{bmatrix}'_{(T\times(MK+1))} \begin{bmatrix} \beta_0 \\ ----- \\ \pi_{11} \\ \vdots \\ \pi_{(M-1)1} \\ \pi_{M1} \\ ----- \\ \pi_{12} \\ \vdots \\ \pi_{(M-1)2} \\ \pi_{M2} \\ ----- \\ \vdots \\ ----- \\ \pi_{1K} \\ \vdots \\ \pi_{(M-1)K} \\ \pi_{MK} \end{bmatrix}_{((MK+1)\times 1)} + \underset{(T\times 1)}{\bar{\mathbf{u}}} \\
&= \mathcal{X}\boldsymbol{\pi}_a + \bar{\mathbf{u}}.
\end{aligned}
$$

However, if H_{01} is true, we rewrite this previous equation as

$$
\begin{aligned}
\underset{(T\times 1)}{\bar{\mathbf{y}}} &= \begin{bmatrix} [\mathbf{1}_T]' \\ -------- \\ [\bar{\mathbf{X}}_1]' \\ -------- \\ [\mathbf{w}_{12}\bar{\mathbf{X}}_2]' \\ \vdots \\ [\mathbf{w}_{(M-1)2}\bar{\mathbf{X}}_2]' \\ [\bar{\mathbf{X}}_2]' \\ -------- \\ \vdots \\ -------- \\ [\mathbf{w}_{1K}\bar{\mathbf{X}}_K]' \\ \vdots \\ [\mathbf{w}_{(M-1)K}\bar{\mathbf{X}}_K]' \\ [\bar{\mathbf{X}}_K]' \end{bmatrix}'_{(T\times(MK-M))} \begin{bmatrix} \beta_0 \\ ------ \\ \pi_{M1} \\ ------ \\ \pi_{12} \\ \vdots \\ \pi_{(M-1)2} \\ \pi_{M2} \\ ------ \\ \vdots \\ ------ \\ \pi_{1K} \\ \vdots \\ \pi_{(M-1)K} \\ \pi_{MK} \end{bmatrix}_{((MK-M)\times 1)} + \underset{(T\times 1)}{\bar{\mathbf{u}}} \\
&= \mathcal{X}_{\mathbf{0}}\boldsymbol{\pi}_0 + \bar{\mathbf{u}}.
\end{aligned}
$$

To test the H_{01}, we employ the following F-statistic from a theorem in Christensen (2002, p. 58):[41]

$$F_{rank(\mathbf{M}-\mathbf{M}_0),rank(\mathbf{I}_T-\mathbf{M})} = \frac{\dfrac{\bar{\mathbf{y}}'(\mathbf{M}-\mathbf{M}_0)\bar{\mathbf{y}}}{rank(\mathbf{M}-\mathbf{M}_0)}}{\dfrac{\bar{\mathbf{y}}'(\mathbf{I}_T-\mathbf{M})\bar{\mathbf{y}}}{rank(\mathbf{I}_T-\mathbf{M})}},$$

where $\mathbf{M}_{(T\times T)} = \mathcal{X}(\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}'$ and $\mathbf{M}_{\mathbf{0}(T\times T)} = \mathcal{X}_{\mathbf{0}}(\mathcal{X}_{\mathbf{0}}'\mathcal{X}_{\mathbf{0}})^{-1}\mathcal{X}_{\mathbf{0}}'$.

Therefore, we reject H_0 if $F > F^*_{rank(\mathbf{M}-\mathbf{M}_0),rank(\mathbf{I}_T-\mathbf{M})}(\alpha)$ with at most $\alpha\,100\%$ probability of Type I error. The quantity $F^*_{rank(\mathbf{M}-\mathbf{M}_0),rank(\mathbf{I}_T-\mathbf{M})}(\alpha)$ is a critical value from an F-dsitribution with $rank(\mathbf{M}-\mathbf{M}_0)$ degrees of freedom in the numerator,

[41] We assume $\bar{\mathbf{u}} \sim N(0,\sigma^2\mathbf{I})$ under H_{a1} and H_{01}.

$rank(\mathbf{I}_T - \mathbf{M})$ degrees of freedom in the denominator, and α level.[42]

However, if data is not available on $w_{i1}(t)$, $w_{i2}(t)$, ..., $w_{iK}(t)$ for $i = 1, 2, \ldots, M$ and $t = 1, 2, \ldots, T$, it might be that $w_{i1}(t), w_{i2}(t), \ldots, w_{iK}(t)$ are functions of variables for which data are available. For instance, suppose[43]

$$\begin{aligned} w_{i1}(t) &= \theta_{0i1} + \theta_{i1} z_1(t), \quad && \text{for } i = 1, 2, ..., M \\ w_{i2}(t) &= \theta_{0i2} + \theta_{i2} z_2(t), \quad && \text{for } i = 1, 2, ..., M \\ &\vdots && \qquad \vdots \\ w_{iK}(t) &= \theta_{0iK} + \theta_{iK} z_K(t), && \text{for } i = 1, 2, ..., M. \end{aligned}$$

[42] I.e., $\alpha = Pr_{\nu_1,\nu_2}\left(F > F^*_{\nu_1,\nu_2}(\alpha)\right)$.

[43] Zellner (1962) assumed this relation is non-stochastic. That is, there is no disturbance term at the end of each function. However, as he explained, if this relation is stochastic, say, $w_{ik}(t) = \theta_{0ik} + \theta_{ik} z_k(t) + \nu(t)$, the approach to take is to consider a regression model in which one (or some) of the independent variables have "measurement error."

Then, equation (2.37) becomes

$$
\begin{aligned}
\bar{y}(t) &= \beta_0 + \beta_{M1}\bar{x}_1(t) + \left[\sum_{i=1}^{(M-1)} \left(\beta_{i1} - \beta_{M1}\right)\left(\theta_{0i1} + \theta_{i1} z_1(t)\right)\right] \bar{x}_1(t) \\
&\quad + \beta_{M2}\bar{x}_2(t) + \left[\sum_{i=1}^{(M-1)} \left(\beta_{i2} - \beta_{M2}\right)\left(\theta_{0i2} + \theta_{i2} z_2(t)\right)\right] \bar{x}_2(t) \\
&\quad \vdots \\
&\quad + \beta_{MK}\bar{x}_K(t) + \left[\sum_{i=1}^{(M-1)} \left(\beta_{iK} - \beta_{MK}\right)\left(\theta_{0iK} + \theta_{iK} z_K(t)\right)\right] \bar{x}_K(t) \\
&\quad + \bar{u}(t) \\
&= \beta_0 + \left[\beta_{M1} + \sum_{i=1}^{(M-1)} \left(\beta_{i1} - \beta_{M1}\right)\theta_{0i1} + \left(\beta_{i1} - \beta_{M1}\right)\theta_{i1} z_1(t)\right] \bar{x}_1(t) \\
&\quad + \left[\beta_{M2} + \sum_{i=1}^{(M-1)} \left(\beta_{i2} - \beta_{M2}\right)\theta_{0i2} + \left(\beta_{i2} - \beta_{M2}\right)\theta_{i2} z_2(t)\right] \bar{x}_2(t) \\
&\quad \vdots \\
&\quad + \left[\beta_{MK} + \sum_{i=1}^{(M-1)} \left(\beta_{iK} - \beta_{MK}\right)\theta_{0iK} + \left(\beta_{iK} - \beta_{MK}\right)\theta_{iK} z_K(t)\right] \bar{x}_K(t) \\
&\quad + \bar{u}(t) \\
&= \beta_0 + \left[\beta_{M1} + \sum_{i=1}^{(M-1)} \left(\beta_{i1} - \beta_{M1}\right)\theta_{0i1}\right] \bar{x}_1(t) + \left[\sum_{i=1}^{(M-1)} \left(\beta_{i1} - \beta_{M1}\right)\theta_{i1}\right] z_1(t)\bar{x}_1(t) \\
&\quad + \left[\beta_{M2} + \sum_{i=1}^{(M-1)} \left(\beta_{i2} - \beta_{M2}\right)\theta_{0i2}\right] \bar{x}_2(t) + \left[\sum_{i=1}^{(M-1)} \left(\beta_{i2} - \beta_{M2}\right)\theta_{i2}\right] z_2(t)\bar{x}_2(t) \\
&\quad \vdots \\
&\quad + \left[\beta_{MK} + \sum_{i=1}^{(M-1)} \left(\beta_{iK} - \beta_{MK}\right)\theta_{0iK}\right] \bar{x}_K(t) + \left[\sum_{i=1}^{(M-1)} \left(\beta_{iK} - \beta_{MK}\right)\theta_{iK}\right] z_K(t)\bar{x}_K(t) \\
&\quad + \bar{u}(t). \qquad (2.38)
\end{aligned}
$$

Now, let

$$\underset{(T\times 1)}{\bar{\mathbf{X}}_1} = \begin{bmatrix} \bar{x}_1(1) \\ \bar{x}_1(2) \\ \vdots \\ \bar{x}_1(T) \end{bmatrix}, \quad \underset{(T\times 1)}{\bar{\mathbf{X}}_2} = \begin{bmatrix} \bar{x}_2(1) \\ \bar{x}_2(2) \\ \vdots \\ \bar{x}_2(T) \end{bmatrix}, \ldots, \quad \underset{(T\times 1)}{\bar{\mathbf{X}}_K} = \begin{bmatrix} \bar{x}_K(1) \\ \bar{x}_K(2) \\ \vdots \\ \bar{x}_K(T) \end{bmatrix},$$

$$\underset{(T\times T)}{\bar{\mathbf{Z}}_1} = \begin{bmatrix} z_1(1) & 0 & \ldots & 0 \\ 0 & z_1(2) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & z_1(T) \end{bmatrix}, \ldots, \underset{(T\times T)}{\bar{\mathbf{Z}}_K} = \begin{bmatrix} z_K(1) & 0 & \ldots & 0 \\ 0 & z_K(2) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & z_K(T) \end{bmatrix}.$$

Now, a simple least-squares regression of $\bar{\mathbf{y}}$ on $\bar{\mathbf{X}}_1$, $\mathbf{Z}_1\bar{\mathbf{X}}_1$, $\mathbf{X}_2$, $\mathbf{Z}_2\bar{\mathbf{X}}_2$, …, $\mathbf{X}_K$, and $\mathbf{Z}_K\bar{\mathbf{X}}_K$ is all that is needed to test K hypotheses of micro-parameter equality. These K possible t-tests are:[44]

$$\begin{aligned} H_{01} &: \sum_{i=1}^{(M-1)} (\beta_{i1} - \beta_{M1})\, \theta_{i1} = 0 \\ H_{a1} &: \sum_{i=1}^{(M-1)} (\beta_{i1} - \beta_{M1})\, \theta_{i1} \neq 0 \\ &\vdots \\ H_{0K} &: \sum_{i=1}^{(M-1)} (\beta_{iK} - \beta_{MK})\, \theta_{iK} = 0 \\ H_{aK} &: \sum_{i=1}^{(M-1)} (\beta_{iK} - \beta_{MK})\, \theta_{iK} \neq 0. \end{aligned}$$

[44] We assume $\bar{\mathbf{u}} \sim N(0, \sigma^2\mathbf{I})$.

Or equivalently, [45]

$$
\begin{aligned}
&H_{01} : \text{There is no Aggregation Bias in } \bar{\mathbf{X}}_1\\
&H_{a1} : \text{There is Aggregation Bias in } \bar{\mathbf{X}}_1\\
&\qquad\qquad\vdots\\
&H_{0K} : \text{There is no Aggregation Bias in } \bar{\mathbf{X}}_K\\
&H_{aK} : \text{There is Aggregation Bias in } \bar{\mathbf{X}}_K.
\end{aligned}
$$

Therefore, we reject the k^{th} null hypothesis ($k = 1, 2, \ldots, K$) if $|t| = \left|\frac{\tilde{\pi}_k - \pi_k}{\widehat{\mathrm{SE}}(\tilde{\pi}_k)}\right| = \left|\frac{\tilde{\pi}_k}{\widehat{\mathrm{SE}}(\tilde{\pi}_k)}\right| > t^*_{(T-K)}(\alpha)$ with at most $\alpha\,100\%$ probability of Type I error. Where $\pi_k = \sum_{i=1}^{(M-1)} (\beta_{ik} - \beta_{Mk})$, $\tilde{\pi}_k$ is the least-squares estimator of π_k, $\widehat{\mathrm{SE}}(\tilde{\pi}_k)$ is the estimate of the standard error of $\tilde{\pi}_k$, and the quantity $t^*_{(T-K)}(\alpha)$ is a critical value from a t-distribution with $(T - K)$ degrees of freedom, and α level.[46]

Now, if we consider a system of M equations, each with only one independent variable ($K = 1$),[47] equation (2.38) reduces to

$$\bar{y}(t) = \beta_0 + \beta_{M1}\bar{x}_1(t) + \left[\sum_{i=1}^{(M-1)} (\beta_{i1} - \beta_{M1})\, w_{i1}(t)\right] \bar{x}_1(t) + \bar{u}(t) \tag{2.39}$$

and equation (2.37) reduces to

$$
\begin{aligned}
\bar{y}(t) \;=\;& \beta_0 + \left[\beta_{M1} + \sum_{i=1}^{(M-1)} (\beta_{i1} - \beta_{M1})\, \theta_{0i1}\right] \bar{x}_1(t)\\
&+ \left[\sum_{i=1}^{(M-1)} (\beta_{i1} - \beta_{M1})\, \theta_{i1}\right] z_1(t)\bar{x}_1(t) + \bar{u}(t).
\end{aligned} \tag{2.40}
$$

---

[45] Since all θ_{ik}, $i = 1, 2, \ldots, M$, are different from zero, H_{0k} is true if each $(\beta_{ik} - \beta_{Mk})$ equals zero, which implies that $\beta_{ik} = \beta_{Mk}$ for $i = 1, 2, \ldots, M$, which then implies that $\beta_{1k} = \beta_{2k} = \ldots = \beta_{Mk}$. Hence, if all micro-parameters corresponding to $\mathbf{X}_{1k}, \mathbf{X}_{2k}, \ldots, \mathbf{X}_{Mk}$ are equal, there is no aggregation bias in $\bar{\mathbf{X}}_k$ where $k = 1, 2, \ldots, K$. However, if $\beta_{jk} \neq \beta_{Mk}$ for $i, j = 1, 2, \ldots, M$ and $i \neq j$, there is aggregation bias in $\bar{\mathbf{X}}_k$.

[46] I.e., $\alpha = Pr(t > t^*_\nu(\alpha))$.

[47] This is Zellner's (1962, p. 357) second example.

In equation (2.39) we can apply one F-test as we did for equation (2.36). In equation (2.40) we can apply one t-test on the corresponding parameter of $\mathbf{Z}_1\bar{\mathbf{X}}_1$ as we did for equation (2.38) to test for aggregation bias in $\bar{\mathbf{X}}_1$.

In the most simple system involving two micro-equations ($M = 2$), each with only one independent variable ($K = 1$),[48] equation (2.38) reduces to

$$\bar{y}(t) = \beta_0 + \beta_{11}\bar{x}_1(t) + (\beta_{11} - \beta_{21})\, w_{11}(t)\bar{x}_1(t) + \bar{u}(t) \tag{2.41}$$

and equation (2.37) reduces to

$$\bar{y}(t) = \beta_0 + [\beta_{21} + (\beta_{11} - \beta_{21})\,\theta_{011}]\,\bar{x}_1(t) + [(\beta_{11} - \beta_{21})\,\theta_{11}]\,z_1(t)\bar{x}_1(t) + \bar{u}(t). \tag{2.42}$$

In equation (2.41), if data is available on $w_{11}(t)$, $t = 1, 2, ..., T$, it is possible to do a least-squares regression of $\bar{\mathbf{y}}$ on $\bar{\mathbf{X}}_1$ and $\mathbf{w}_{11}\bar{\mathbf{X}}_1$. However, due to the simplicity of equation (2.41), not only an F-test can be performed to test for aggregation bias in $\bar{\mathbf{X}}_1$ as we did for equation (2.36) but also a t-test on the corresponding parameter of $\mathbf{w}_{11}\bar{\mathbf{X}}_1$.

Therefore, we reject the null hypothesis if $|t| = \left|\frac{\tilde{\pi} - \pi}{\widehat{\mathrm{SE}}(\tilde{\pi})}\right| = \left|\frac{\tilde{\pi}}{\widehat{\mathrm{SE}}(\tilde{\pi})}\right| > t^*_{(T-1)}(\alpha)$ with at most $\alpha\,100\%$ probability of Type I error. Where $\pi = (\beta_{11} - \beta_{21})$, $\tilde{\pi}$ is the least-squares estimator of π, $\widehat{\mathrm{SE}}(\tilde{\pi})$ is the estimate of the standard error of $\tilde{\pi}$, and the quantity $t^*_{(T-1)}(\alpha)$ is a critical value from a t-distribution with $(T - 1)$ degrees of freedom and α level.[49]

If data is not available on $w_{11}(t)$ but on $z_1(t)$, $t = 1, 2, ..., T$, equation (2.42) can be used instead of equation (2.41) to test for aggregation bias in $\bar{\mathbf{X}}_1$. As we did for equation (2.38), we apply one t-test for the corresponding parameter of $\mathbf{Z}_1\bar{\mathbf{X}}_1$.

[48] This is Zellner's (1962, p. 356) first example.

[49] I.e., $\alpha = Pr(t > t^*_\nu(\alpha))$.

CHAPTER III

LITERATURE REVIEW

The primary objective of this chapter is to review previous research on seemingly unrelated regressions (SUR) and relevant issues in consumer survey data that will provide insight in the formulation and empirical application of Mexican household meat consumption model. Section 3.1 starts with a review of censored data, which is a frequently encountered problem in consumer survey data—the same nature of the data used in this study. This issue is our first experience before familiarizing with a second problem encountered in consumer survey data: adult equivalence scales. In order to understand how other researchers have modeled and estimated adult equivalence scales, it is good to have knowledge on the censored expenditures problem. This is recommended because the literature reviewed in Section 3.2 implicitly assumes the reader was familiar with censored expenditures. In addition, the models from Section 3.1 are then used in Section 3.4 as examples of parametric models of item nonresponse on the dependent variable. Section 3.3 expands on the topic discussed in Chapter 2 by reviewing seemingly unrelated regression with unequal number of observations. Section 3.4 provides basic concepts related to missing data and then it moves on to explain how to deal with it. Some of the techniques learned in Section 3.4 were intended to be used in Section 4.2, but a simpler approach was adopted. Finally, Section 3.5 begins with an introduction to stratified sampling and finishes with a discussion on how to estimate linear models with stratified sample data—the type of probability sampling technique that was used to collect the data used in this study. Finally, Section 3.6 briefly explains the bootstrap, a general bootstrap algorithm, and different bootstrap sampling methods.

3.1 Censored Expenditures

Censored expenditures are common in consumer survey data. Generally, the censoring is due to survey design and implementation or institutional constraints. **Cen-**

sored expenditures occur when the value is partially known. It is partially known because even though you do not have the actual value (it might be coded as zero or omitted) on the variable of interest (e.g. the dependent variable); you do have information on related variables (e.g. the independent variables). As it will be explained in Section 3.4, this is also referred as item nonresponse on the dependent variable. In literature, when information is missing on both dependent and independent variables, the dependent variable is referred as **truncated** (Wooldridge, 2006, p. 613; Pindyck and Rubinfeld, 1997, p. 325). Again, using Section 3.4's terminology, when information is missing on both dependent and independent variables and there is no more information collected, it is also referred as unit nonresponse. A **truncated regression model** differs from a censored regression model in that in a truncated regression model we do not observe any information about a certain segment of the population (Wooldridge, 2006, p. 613). In addition, truncated regression is a special case of a general problem known as nonrandom sample selection (Wooldridge, 2006, p. 616).

Wooldridge (2006, p. 609) explains **censored data** is an issue of data observability. Wooldridge (2006, p. 609) explains the use of a **censored regression model** when there is missing data on the response variable (the dependent variable) but there is information about when the variable is missing (above or below some known threshold). For instance, consider the example provided by Wooldridge (2006, p. 610) where we know the value of a family's wealth up to a certain threshold. This censoring problem might occur, Wooldridge (2006) explains, when respondents are asked for their wealth, but people are allowed to respond with "more than $500,000." Then, we observe actual wealth for those respondents whose wealth is less than $500,000 but not for those whose wealth is greater than $500,000. In this case, the censoring threshold is fixed for all families whose wealth is greater than $500,000. However, the censoring threshold may also change depending on individual or family characteristics. For instance, consider another example provided by Wooldridge (2006, p. 611) where we know the time in months until an inmate is arrested after being released from prison. By the end of the period in which you investigate if an inmate was arrested again, not all of them would have been rearrested; therefore, the observations from the inmates not yet arrested would be censored. In other words, some felons may never be arrested again or they may be arrested after such a long time that there is a need to censor the number of days in order to analyze the data. In addition, in this case, the censoring time is different for each inmate. By providing an empirical application of the second example, Wooldridge showed that applying Ordinary Least Squares (OLS) will result in coefficient estimates markedly different from those of a censoring regression model where coefficients and the variance of the error term are estimated by maximum likelihood. In his example, OLS coefficient estimates were all shrunk toward zero. Furthermore, Wooldridge (2006, p. 613) emphasized that an application of a censored regression model will be more reliable.

The second example provided by Wooldridge (2006, p. 611) is very similar to a problem encountered in this study with the Mexican survey data on household income and weekly expenditures. At the end of the period in which the interviewer recorded all items purchased by a household, there will be items that would have not been purchased, which the household consumes, or were purchased away from home and the interviewer did not record them. Therefore, items not purchased during the week of the interview, which the household consumes, will be censored.

Pindyck and Rubinfeld (1997, p. 325) explain that censoring occurs when "the dependent variable has been constructed on the basis of an underlying continuous variable for which there are a number of observations about which we do not have information." Pindyck and Rubinfeld (1997, p. 325) provide the following examples.

> Suppose, for example, that we are studying the wages of women. We know the actual wages of those women who are working, but we do not know the "reservation wage" (the minimum wage at which an individual will work) for those who are not. The latter group is simply recorded as not working. Or suppose that we are studying automobile purchasing behavior using a random survey of the population. For those who happened to buy a car, we can record their expenditure, but for those who did not we have no measure of the maximum amount they would have been willing to pay at the time of survey.

Pindyck and Rubinfeld (1997, p. 325) explain ordinary least-squares estimation of the censored regression model results in biased and inconsistent parameter estimates. They emphasized a maximum-likelihood estimator as a preferred alternative.

Pindyck and Rubinfeld's (1997, p. 325) examples provide insight into the data used in this study, the Mexican survey data on household income and weekly expenditures. For those households that happen to buy a particular item, their expenditure was recorded, but for those who did not we have no measure of the maximum amount they would have been willing to pay at the time of the survey. As it will be explained later, the Mexican survey data on household income and weekly expenditures omit this transaction (i.e., does not make any record of items not purchased). Hence, expenditure on that particular item is censored because if the price goes below the maximum price they may have been willing to pay, the household would have purchased that item. Consequently, for those who did not buy an item, we have no measure of the maximum amount they would have been willing to pay at the time of the survey.

Some researchers more specifically point out the importance of addressing the presence of censored food expenditures when working with weekly food expenditures. If weekly expenditures are reported for at-home expenditures and away-from-home expenditures, then not all households might purchased and consumed food away from home. When this is the case, expenditures on food away from home are censored in nature (Sabates et al., 2001; Gould and Villarreal, 2002).[1] For example, a household expenditure on food away from home that is not recorded (sometimes, but not in the dataset used in this study, expenditure on food away from home are recorded as zero dollars) is censored because this household might have bought and consumed this commodity a week later after the interviewer left. However, if the commodity was not bought at all, then there is no censoring. It is worth while mentioning that Gould and Villarreal (2002) and Sabates et al. (2001) both used data corresponding to the

[1]This is the same idea of the censored data problem mentioned above but this time distinguishing between at-home expenditures and away-from-home expenditures.

year 1996 from the Mexican survey of household income and expenditures, *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)*, published by a Mexican governmental institution (Instituto Nacional de Estadística, Geografía e Informática or INEGI). The study presented in this report uses data from the same source.

3.2 Adult Equivalence Scales

Adult equivalence scales are measures that show how much an individual household member of a given age and sex contributes to household expenditures or consumption of goods relative to a standard household member. As explained by **Deaton and Muellbauer (1986)** adult equivalence scales assign different weights to household members according to their age and gender; whereas a simple count of household members, the most common practice, implicitly assumes each household member has the same marginal impact. The purpose of scales is to capture economies of size associated with larger households, the different impacts of children versus adults and to permit welfare comparisons across households of different size and composition (Lazear and Michael, 1980; Deaton and Muellbauer, 1986; Blaylock, 1991; Perali, 1993).

Deaton and Muellbauer (1986) note that equivalence scales can be determined from nutritional and psychological studies, sociological relationships, or the use of revealed consumption or purchase patterns. They note that the last approach appears to be the most reasonable but there continues to be the dilemma of how to use expenditure data to develop these scales (Brown and Deaton, 1972).

Gould and Villarreal (2002) analyzed Mexican adult equivalence scales and weekly food expenditures for Mexican beef and pork purchases in 1996. They endogenously determined adult equivalence scales and allowed marginal impact to vary by age and gender. Their study estimated commodity-specific adult equivalence scales using the single equation approach suggested by Tedford et al. (1986). Moreover, their two-stage econometric model was an extension of the model implemented by Dong et al. (1998) but first formulated by Wales and Woodland (1980). Their model accounted for censored meat expenditures and endogenously determined commodity specific unit values and therefore product quality. Gould and Villarreal (2002) proceeded to examine whether there are differences in the impacts of household composition on food expenditures and whether there are differences in equivalent profiles across meat commodity.

Gould and Villarreal (2002) used Deaton's (1988) suggestion of using unit values, which are obtained by dividing expenditures by quantities purchased. They cautioned that these values should not be interpreted as the market prices. This is because unit values may not only reflect quantity but also quality.[2]

To represent the consumer maximization problem in terms of composite goods, Gould and Villarreal (2002) used the Hicksian composite commodity theorem (Deaton, 1988; Nelson, 1991). Then they combined it with a quality indicator (Theil, 1952–1953; Dong et al., 1998) and a measure of average quality within each commodity group to derive an expression for composite commodity unit value and expenditures on composite commodity. Then, Gould and Villarreal (2002) referred to Dong et al. (1998) and Wales and Woodland (1980) to derive an expression for censored household expenditures.

Gould and Villarreal (2002) specified an unconditional probability that a particular household will not purchase a particular commodity, the log-likelihood function

[2]The market price of a commodity refers to the price of such commodity in the market, assuming this commodity is homogeneous everywhere in the market. Hence, there is only one price for the commodity in the market. However, in practice, even if we are dealing with homogeneous commodities there are going to be differences in the price of a commodity in different locations due to different profit margins charged by different sellers. Sellers charging higher profit margins most likely reflect 'better' store's conditions. For example, even if it is the same commodity, it is very likely that a supermarket will have a different price than a flea market. This difference in price is translated into a quality attribute. Therefore, calculating unit values by dividing expenditure by quantity will pick up these differences in quality. However, one could argue that market prices, which are some kind of average of prices in the market, include differences in qualities anyways; therefore, in practice unit values and market prices are the same thing. Nonetheless, when the difference between the two is clear, this distinction is usually made.

for the entire set of households, and the expected values of conditional expenditures and unit values. Their study combined expressions for conditional probability to obtain an estimate of the expected value of unconditional expenditures and unit values. Model parameters were obtained by maximizing the likelihood function.

Estimates were reported for the coefficients of the unit value equation, the expenditure equation and the equation for the total of adult equivalence scales in each household. Additionally, income and adult equivalent elasticities and marginal regional impacts were reported. They found that household composition is an important determinant of total household expenditures as well as product quality. They rejected the null hypothesis that the marginal impact of an additional household member on meat expenditures is invariant to the member's age or gender. They found a small but positive impact of the number of adult equivalents in the household on expenditures for beef and pork. They also found a negative impact of the number of adult equivalents in the household on endogenous unit values.

However, their study could not reject the null hypothesis that the female and male adult equivalent profiles are the same. Even more surprising, they found that female adult equivalence scale consistently exceeds the male adult equivalence scale in consumption of beef for females of 40-65 years old. They attributed this result to the high participation of males in the labor force compared to adult females. Adult males working more time outside their home tend to purchase and consume more food away from home than adult females who stay at home. This finding is similar to Sabates et al. (2001) who found that adult female equivalence scales in Argentina and Brazil were either no different or lower than adult male equivalence scales over the age of 40 years. Since the data they used in the analysis did not allow them to identify who purchased and consumed food away from home, Gould and Villarreal (2002) further examine this result by regressing the percentage of total food expenditures originating from food-away-from home purchases on household income, household size, percentage of adult males working full and part time, and percentage of adult females working full and part time. They found insignificant male adult impacts and significant female adult impacts.

Sabates et al. (2001) studied weekly food expenditures in Argentina, Brazil, and Mexico for the corresponding time periods of 1996-1997, 1995-1996, and 1996. They analyzed the impacts of household member counts versus endogenously determined equivalence scales at the per capita aggregated food expenditure level. They estimated country specific expenditure functions to obtain parameter estimates and perform several non-nested hypothesis tests. For instance, hypothesis tests were elaborated to know whether male and female adult equivalent profiles are the same; or whether the use of a simple count of household members provides as much information as the use of adult equivalence scales in explaining food purchase behavior; or whether adult equivalence scales are the same across Argentina, Brazil and Mexico. In addition, they created interaction variables with income to calculate and report income and adult equivalent elasticities. Finally, Sabates et al. (2001) also compared the distribution of weekly per capita food expenditures based on the simple count of household members with the distribution of weekly per capita food expenditures based on the number of adult equivalence scales.

Sabates et al. (2001) made use of the model proposed by Tedford et al. (1986) and categorized each household member as being in a developmental or transitional period (Levinson et al., 1978). Then, cubic spline functions were used to join the developmental periods with the transitional periods (Tedford et al., 1986, pp. 323–325). The number of adult equivalent scales was derived from the cubic spline functions based on the gender and age-based categories (Tedford et al., 1986, pp. 325–236). In addition, the number of adult equivalents in each household equaled the sum of adult equivalence scales over all household members. A total food expenditure function was specified. Log-likelihood functions for the total number of adult equivalence scales in each household and for the simple count of household members in each household were also specified. Parameter estimates were obtained by maximizing the above log-likelihood functions using an iterated procedure. Each log-likelihood function required a specification for the variance and the assumption of normality for the error term of the total food expenditure function.

Sabates et al. (2001) found that adult male equivalent profiles are statistically different from adult female profiles. Male household members in general placed greater demands on household food supplies than female members. In particular, for both Argentina and Brazil the female adult equivalent value was below the male value; however, for Mexico, the male profile was greater than the female profile for up to age 35. After this age, the male and female profiles followed a similar pattern. The male profile in Mexico increased in adult equivalence scale values up to the mid-50s and then declined. They found the oldest male age category in Mexico has an adult equivalence scale value of 1.15 but it was not statistically different from 1. The female profiles for Argentina and Brazil were consistently less than 1.0. Similar to the male profile for Mexico, the male profile for Argentina and Brazil increased in adult equivalence scale values until the mid-50s and then declined.

Sabates et al. (2001) also found that a simple count of household members does not provide the same information as the use of equivalence scales in explaining food purchase behavior. Age and gender information has a statistical significant effect in food expenditures. Furthermore, Sabates et al. (2001) graphically showed and statistically proved that the distribution of weekly per capita food expenditures based on the simple count of household members is consistently above and statistically different than the distribution of weekly per capita food expenditures based on the number of adult equivalence scales. Therefore, using the former variable as a measure of poverty will result in a significant increase in the number of households below a defined poverty line.

Tedford et al. (1986) developed a model to calculate adult equivalence scales, which they named after their last names as the TCH model. Their model was based on concepts from the fields of psychology as well as child and human development. Their conception and components of the life cycle was based upon research by Levinson et al. (1978) and upon concepts from child and human development described by Duvall (1977) and by Vander Zanden (1978). They presented a model where the life cycle was comprised of a sequence of developmental and transitional phases. Tedford et al. (1986) also compared adult scale parameter estimates for total food expenditure from their model with estimates from Blokland's (1976) and Buse-Salathe's (1978) models. Tedford et al. (1986) used U.S. household weekly data from 1977-1978 to obtain parameter estimates of the three models. In addition they reported estimates of the income elasticity for food and household equivalence scale elasticity for food. They also considered geographical regions and whether household were located in central city or non-metropolitan area.

First, Tedford et al. (1986) presented different ways in which the life cycle can be delineated by ages or important events. They presented the view of Levinson et al. (1978) of the life cycle as a sequence of developmental and transitional periods and as a sequence of eras. They also presented the view of Duvall (1977) of the life cycle as a sequence of important events, and the National Research Council's recommendations of the different food energy allowances for males and/or females during the life cycle. Tedford et al. (1986) primarily adopted Levinson's et al. (1978) developmental and transitional periods to specify adult scale functions. Since consumption, expenditure, and socio-demographic information are reported only for the household unit, Tedford et al. (1986) estimated adult scale parameters indirectly as components of household equivalence scales. Household equivalence scales were then aggregates of adult equivalence scales and expressed explicitly as functions of the adult scale parameters. Tedford et al. (1986) specified household equivalence scales as functions of weighted sum variables dependent upon the age-sex composition of the household. An expenditure function was then estimated using a nonlinear procedure as a function of expenditure, income, education, sex, age, geographical region variables, level of urbanization variables, seasonality variables, race variables, the household equivalence scale function, and square of the household equivalence scale function. The latter variable was introduced to account for the possible existence of economies of size. Inclusion of socio-demographic variables reflected the recognition of heterogeneous tastes and preferences. Finally, households that did not report relevant income or socio-demographic information were excluded. Tedford et al. (1986) claimed that sample selection bias was not going to be a problem because the frequencies for the usable sample are quite similar to the frequencies for the overall sample.

Based on the statistical significance of some key parameter estimates and the statistical significance from each other, Tedford et al. (1986) found that the Buse-Salathe's (1978) life-cycle-age-class specification was inconsistent with Blokland's (1976) specification. However, in the analysis of Tedford et al. (1986), despite differences in the age-class delineations and despite the fact that TCH model constitutes a more general specification than Buse-Salathe's (1978) model, the empirical findings of the scale parameters based on the TCH model were similar to those based on Buse-Salathe's (1978) model. Additionally, Buse-Salathe's (1978) model was also a more general specification than Blokland's (1976) model. Hence, the most general specification is found in the TCH model while the simplest specification is found in Blokland's (1976) model.

Tedford et al. (1986) also found that food expenditure behavior for males and females is generally different at various developmental and transitional stages of the life cycle. The TCH model even indicated that food expenditure behavior is different from males and females within the same developmental and transitional stages of the life cycle. They also found differences in household food expenditures by regions, seasons, and by population density (city or non-metropolitan location).

Based on the life cycle pattern of the three models, Tedford et al. (1986) concluded that the adult equivalence scale specification by Blokland (1976) may be too restrictive. Second, the TCH and the Buse-Salathe's (1978) equivalence scales during the life cycle profile were reasonably similar, although noticeably differences resulted in the equivalence scales for females as well as for household members greater than sixty years of age.

Summarizing, the last three articles presented models where adult equivalence scales are determined endogenously within the model. All these models require the specification of an expenditure function which incorporates adult equivalence scales.

Specifically, the first two articles used the Levinson's et al. (1978) sequence of transitional and developmental periods of the life cycle. Hence, adult equivalence scales were estimated by linking concepts from psychology, child development, and human development to economic concepts. Tedford et al. (1986) repeatedly remarked the explicit rationale and consistency of their TCH model with the life-cycle developmental concepts. However, although perhaps lacking some of this rationale and consistency, Tedford et al. (1986) also presented alternative models such as the Blokland's (1976) and Buse-Salathe's (1978) models and the National Research Council's recommendations on food energy allowances for males and/or females. Despite the model used, it is required to use a measure of adult equivalence scale in per capita meat expenditures. In addition, it is important to differentiate between males and females as it has been statistically shown that male and female household members place different demands on household food supplies (Gould and Villarreal, 2002; Sabates et al., 2001; Tedford et al., 1986). It also important to notice that we cannot use estimates of adult equivalence scales in another country for Mexico or estimates of similar commodities because these scales change across countries (Sabates et al., 2001) and across commodities (Gould and Villarreal, 2002).

In addition, it can be observed that these adult equivalence scales tend to be smaller for female household members than male household members (Sabates et al., 2001; Tedford et al., 1986) but it might not always be the case specially when there is high participation of males in the labor force compared to adult females (Gould and Villarreal, 2002). In addition, these scales tend to be smaller than one for members younger or older than the standard adults (Gould and Villarreal, 2002; Sabates et al., 2001; Tedford et al., 1986).

3.3 Seemingly Unrelated Regressions with Unequal Number of Observations

Chapter 2 explained Zellner's (1962) method of estimating parameters of a set of regression equations with equal number of observations. In this section we briefly review SUR models with unequal number of observations. The literature reviewed in this section will provide an idea how SUR with unequal number of observations is handled, inform about the alternative estimators of the variance-covariance matrix of the error term ($\mathbf{\Sigma}$) and the conditions under which one estimator will perform better than another, the main findings about the different feasible-GLS-regression-coefficient estimators and whether or not it is relevant to use better estimates of $\mathbf{\Sigma}$, and other relevant issues related to SUR and unequal number of observations.

In addition, in Chapter 2, it was mentioned that the SUR model was not necessarily restricted to time-series data. Specifically, it was said that Zellner (1962) provides specific examples when the data could be time series, cross sectional or both. It is important to be clear that the SUR model with unequal number of observations is also not restricted to time-series data.

In times-series data, the extra observations of one equation with respect to a second equation will necessarily be missing in the second equation. That is, in terms of Section 3.4, we have unit nonresponse because the entire observation unit is missing—either because the observation unit did not provide any information or simply because no information was collected from the observation unit during the time period under consideration. In consumer surveys of cross-sectional data, the extra observations of one equation with respect to a second equation could be missing in the second equation for several reasons. For example, consider a consumer survey where interviewers make journal entries of any consumption item purchased by the interviewees (households) during the time of interview (say one week).[3] During the week of the interview not all possible consumption items will be purchased by the households; hence, it is easy to record only those that are purchased rather than making a list of all items consumed by the households and record those that were purchased and those that were not. Given that interviewers record only those that were purchased during the week of the interview, it is natural that an expenditure equation (or a demand equation) of a consumption item will have more (or less) observations than a second expenditure

[3] This is the case for the data used in this study. Section 4.2 discusses how ENIGH collects expenditure data on items purchased by households by using a weekly journal.

equation on another item. As explained in Section 3.1, expenditure on the second item is censored and there are several explanations why. In this example, there are extra observations in an expenditure equation of an item compared to an equation for another item because not all households purchase the same items. For all items to have the same observations, households will have to buy the same items. Now, if journals were sent by mail and households were asked to record all items purchased during a week, there would be households that will refuse to write the journal, and among those who participate, there would be households that may refuse to record all items. This will also result in unequal number of observations for some equations. Finally, we could combine time-series with cross-sectional data and give examples of panel data.

Whether there is time-series data, cross-sectional data, or panel data, there will be examples of equations with unequal number of observations. The literature reviewed in this section includes both time-series data (Sharma, 1993; Baltagi et al., 1989; Brown and Kadiyala, 1985) and cross-sectional data (Baltagi et al., 1989).

Sharma (1993) attempted to estimate two-equations using seemingly unrelated regression models when the number of observations in each equation were unequal. He studied two cases. In the first case, there were $n_1 + n$ time series observations for the first equation and $n + n_2$ for the second equation; the last n observations of the first equation match in time with the first n observations of the second equation. In the second case, there were n time-series observations for the first equation and $n_1 + n + n_2$ for the second; the observations for the first equation match in time with those for the second, starting from the $(n_1 + 1)$-th observation.

In the first case, Sharma (1993) partitioned the matrices of each regression equation strategically by the number of observations in the first equation that match in time with the second equation. He computed the variance-covariance matrix for the error term, and then he used generalized least-squares to estimate the vector of parameters. He showed that when the observations on both equations start from the same point in time ($n_1 = 0$), his results reduced to those given by Schmidt (1977).

Similarly, when the observation on both equations terminate together in time series ($n_2 = 0$), his results were similar to those given by Schmidt (1977).

In the second case, there was only the need to partition the second equation strategically by n_1, n, and n_2. Sharma (1993) again computed the variance-covariance matrix of the error term and used generalized least-squares to estimate the vector of parameters. When $n_1 = 0$, his results reduced once again to those of Schmidt (1977). When $n_2 = 0$, similar results to Schmidt (1977) were obtained.

In both cases, Sharma (1993) also indicated how to proceed when the elements of the variance-covariance matrix were unknown. In general, partition of the error terms of each equation and the use of least-squares residuals was necessary for the first case, but only partition of the error term of the second equation was necessary for the second case.

Sharma (1993) explained that if first order autoregressive errors were present, his analysis could be modified analogously to Parks (1967). He also remarked that his results apply when the order of observations is important (e.g. time-series data) or the order of observations is irrelevant (e.g. cross-sectional data) while Schmidt's (1977) results only apply to the latter situation.

Hwang (1990) studied several alternative estimators of the variance-covariance matrix of the error term ($\mathbf{\Sigma}$) in the seemingly unrelated regressions (SUR) model when sample sizes vary for a two-equation SUR model. The purpose of his study was threefold. First, he wanted to clarify the amount of sample information that enters the generalized least squares (GLS) estimation procedure through the alternative estimators. In particular, he wanted to emphasize that the sample information contained in each estimator of $\mathbf{\Sigma}$ is misleading. Second, Hwang (1990) identified a sample statistic (α) which differs among alternative estimators of $\mathbf{\Sigma}$ presented in his study.[4] His sample statistic was used to investigate the conditions under which an estimator of $\mathbf{\Sigma}$ performs better than the other estimators. Hence, Hwang (1990) showed that his sample statistic was a useful guide for the choice of estimator in practice. Finally,

[4]A sample statistic is a function of the sample.

Hwang (1990) proposed an alternative estimator of $\mathbf{\Sigma}$ based on the specification of Telser (1964).

Hwang (1990) examined four of the five alternative estimators of $\mathbf{\Sigma}$ presented by Schmidt (1977). The first alternative was the "usual" estimator of Zellner (1962), but instead of dividing the sum of squared OLS residuals by the difference of the number of observations in each equation and the number of independent variables (as originally presented in Zellner (1962)), Schmidt (1977) divided only by the number of observations in each equation, except for the second equation where the extra observations were ignored. The second, third and fourth alternatives were Wilks' (1932) estimator, Srivastava-Zaatar's (1973) estimator, and Hocking-Smith's (1968) estimator respectively.

As explained by Hwang (1990), the "usual" estimator ignores the extra observations on the second equation; the Wilks' (1932) estimator uses them only in the estimation of the variance of the error term in the second equation (σ_{22}); the Srivastava-Zaatar's (1973) estimator uses them in the estimation of the covariance of the contemporaneous errors of the two equations (σ_{12}) and the variance of the error term in the second equation (σ_{22}) and the Hocking-Smith's (1968) estimator fully uses the extra observations on second equation in the estimation of all σ_{ij}.

Hwang (1990) first parameterized $\mathbf{\Sigma}^{-1}$ by a set of three parameters (θ, δ, σ_{22}). The first parameter is the ratio of the covariance of the contemporaneous errors of the two equations to variance of the error term in the second equation $\left(\theta = \frac{\sigma_{12}}{\sigma_{22}}\right)$, the second parameter $\delta = \sigma_{11} - \theta^2\sigma_{22}$, and the third parameter is the variance of the error term in the second equation (σ_{22}). Hwang (1990) explained that his parameterization is commonly used in multivariate statistical analysis; for example, he explained, in the orthogonal transformation of a normal random vector, in the conditional distribution, in the partial correlation coefficient, etc. Then Hwang (1990) presented the alternative estimators of $\mathbf{\Sigma}$ in terms of their estimates of (θ, δ, σ_{22}). Hwang (1990) proceeded to explain that differences among the alternative estimators of hinge on the value of the sample statistic (α). Then, before turning to a sampling experiment, he proposed an alternative procedure for the estimation of (θ, δ, σ_{22}), the Telser-type (1964) estimator.

In his sampling experiment, Hwang (1990) measured relative efficiencies of alternative estimators by the ratios of their mean square errors (MSEs) of the estimate of the i^{th} equation parameter ($\hat{\beta}_i$) to that of the "usual" estimator. Since MSE ratios of individual coefficient estimates varied from one coefficient to another, Hwang (1990) decided to compute the average of the MSE ratios of the three coefficients in each equation.

Hwang (1990) computed the sample statistics of α, the average MSE ratios of the alternative estimators to the usual estimator in the full sample, and the average MSE ratios of the alternative estimator to the usual estimator in subsamples. The latter was constructed to investigate the effects of α. Hwang (1990) found that the sample distribution of the coefficient estimators are sensitive to the values of α.

> In particular, the [Hocking-Smith's (1968)] estimator of $\mathbf{\Sigma}$ may yield significantly more efficient coefficient estimates than the "usual" estimator of $\mathbf{\Sigma}$ when α is significantly larger than one and [the contemporaneous correlation between the error terms of the two equations (ρ)] is high.
>
> In practice, ρ is generally unknown. Therefore, a reasonable procedure is to estimate $\hat{\rho}$ from the joint observations first, and then use the [Hocking-Smith's (1968)] estimator if α and $\hat{\rho}$ are large. If $\hat{\rho}$ is small and/or α is smaller than one, the "usual" estimator is the proper choice. Alternatively, when $\hat{\rho}$ is high, one may use the Telser estimator of $\mathbf{\Sigma}$, which dominates other estimators regardless of the value of α.

Then, Hwang (1990) proceeded to briefly mention and provide some results on how to extend the model when there are more than two equations. Specifically, he considered a three-equation system for which the first two equations have N observations and the third equation has T observations.

Baltagi et al. (1989) used the same system of two equations of Schmidt (1977) and Kmenta and Gilbert (1968) to replicate the Monte Carlo experiments performed by Schmidt (1977) with the objective of providing additional support or counter-evidence to his findings. Baltagi et al. (1989) also explored whether the performance of estimates of the variance-covariance matrix of the error term in the two-equation seemingly unrelated regression model ($\mathbf{\Sigma}$) leads to better estimates of the regression coefficients. In addition, Baltagi et al. (1989) considered the re-parameterization by Hwang (1987) in order to check whether better estimates of $\mathbf{\Sigma}^{-1}$ yield better estimates of the regression coefficients. Finally, Baltagi et al. (1989) focused on the type of additional observations available. That is, whether the results will change if the type of additional information is time-series or cross-sectional.

Baltagi et al. (1989) examined four of the five different estimators of $\mathbf{\Sigma}$ presented by Schmidt (1977). All of the estimators of $\mathbf{\Sigma}$ presented by Schmidt (1977) are consistent and the corresponding coefficient estimates are asymptotically efficient. Baltagi et al. (1989) examined Wilks' (1932) estimator, Srivastava and Zaatar's (1973) estimator and the Hocking and Smith's (1968) estimator. The fifth estimator presented by Schmidt (1977), the maximum likelihood estimator, was not considered. Baltagi et al. (1989) rather focused on the true GLS estimator instead of the maximum likelihood estimator as a basis for comparison. In the design of their experiment, Baltagi et al. (1989) followed Schmidt's (1977) and Kmenta and Gilbert's (1968) model of two equations.

> Following [Schmidt (1977)], we set the variance of [the error term of the first equation (ϵ_1)] and [the error term of the second equation (ϵ_2)] equal to one ($\sigma_{11} = \sigma_{22} = 1$) and consider three alternative values of the correlation between ϵ_1 and ϵ_2: namely $\epsilon_{12} = \rho = 0.3$, 0.6, 0.925. Three different values of the extra observations are used: E = 5, 10 and 20. All the extra observations are on the second equation. Also, three different sample sizes are considered: T = 10, 20, and 50. For our study all possible combinations of T, E and ρ are entertained.
>
> For each experiment, ($\mathbf{X}$ matrix, value of ρ, value of T and value of E), a sample was generated using a pseudo-random normal deviate generator and the four feasible GLS estimators described in the previous section along with the true GLS and OLS are performed. Each experiment is replicated 500 times and the MSE's are obtained for the σ's and the regression coefficients. Also, a count measure is obtained which gives the number of times an estimator is close to the true value of the parameter than another estimator, and whether this frequency count is significantly different from 50%.

Wilks' (1932) estimator was dropped for comparison purposes of the various estimators because the estimate of $\mathbf{\Sigma}$ was not necessarily positive definite. That is, it frequently gives negative definite estimates of $\mathbf{\Sigma}$. Baltagi et al. (1989) reported the mean squared error (MSE) of the remaining three feasible GLS estimators of $\mathbf{\Sigma}$, the number of times a specific estimator (additionally including the OLS estimator) of β_{11} (the coefficient of one of the independent variables ($\mathbf{X_{11}}$) in the first equation) is closer to β_{11} than the true GLS estimator, and the mean-squared error (MSE) of various estimators (the usual, Srivastava-Zaatar (SZ), Hocking-Smith (HS), and OLS estimators) of β_{11} to that of true GLS. Baltagi et al. (1989) found evidence that estimators of the variances which use the extra observations have better MSE and better simple count performance[5] than those estimators that do not use the extra observations fully. They also found that better estimates of the variances need not imply better estimates of the regression coefficients.

With respect to Hwang's (1987) study, Baltagi et al. (1989) explained that Hwang's (1987) re-parameterization of the estimation problem in terms of the elements of $\mathbf{\Sigma}^{-1}$ rather than $\mathbf{\Sigma}$ was different from the original $\mathbf{\Sigma}$ parameterization presented by Schmidt (1977). In particular, the Hocking-Smith estimator presented by Schmidt (1977) was shown to use extra observations in estimating all the elements of $\mathbf{\Sigma}$, while the Hocking-Smith (1968) estimator presented by Hwang's (1987) differed from the usual estimator only in its estimate of θ_2 (one of the three re-parameterizations of the estimation problem for the second equation).

Baltagi et al. (1989) used Hwang's (1987) re-parameterization to examine comparisons of the various regression coefficient estimators according to the performance of the corresponding estimate of θ. Once again they found that "better estimate of a certain crucial parameter of $\mathbf{\Sigma}^{-1}$ (that differentiates between two feasible GLS estimators) does not necessarily lead to a better estimate of the corresponding regression

[5]A count of the number of times that an estimator of the variance of the error term of the first (σ_{11}) or second equation (σ_{22}) or the covariance of the error term of the first equation with the error term of the second equation (σ_{12}) was close to the true variance or covariance.

coefficients." In addition, Baltagi et al. (1989) results indicated that for larger ρ and larger T, the MSE performance of $\hat{\theta}_{1,HS}$ better than that of $\hat{\theta}_{1,SZ}$ but that this dominance does not necessarily translates into dominance of $\hat{\beta}_{11,HS}$ over $\hat{\beta}_{11,SZ}$. Finally, they explained their conjecture that better estimates of the variances need not imply better estimates of regression coefficients has also been obtained in panel data studies by Maddala and Mount (1973), Taylor (1980) and Baltagi (1981).

Finally, when the type of additional observations changes from time-series to cross-sectional data, Baltagi et al. (1989) found that in both data sets that "[f]easible GLS estimators that seem to ignore the extra observations in estimating $\mathbf{\Sigma}$ (but not necessarily in estimating $\mathbf{\Sigma}^{-1}$ or $\boldsymbol{\beta}$) do not generally do badly relative to feasible GLS estimators that seem to use extra observations fully." That is, "[Schmidt's (1977)] results are shown to be robust to the type of additional observation available i.e., whether they are time series or cross-sectional in nature."

Brown and Kadiyala (1985) also studied the estimation of missing observations with time series data. They referred to missing observations as the difference of the number of observations between two-equations in a seemingly-unrelated-regressions (SUR) model with time series data. That is, the extra observations of one equation were referred to as missing observations in the other equation. The objectives of their study were twofold. First, they wanted to design a test statistic for the significance of the prediction efficiency of a seemingly unrelated regression (SUR) model. Their test statistic consisted of a likelihood ratio test for the predictive ability of the SUR method against the single equation alternative. Second, the cumulative residual procedure described by Fama et al. (1969) was used as a special case within the class of "missing data estimation problems" in an empirical application. In their latter objective, Brown and Kadiyala (1985) adopted a two step process for predicting missing observations from a stock revenue series from the utilities and airline industry over the period 1968-1977. In the first step, a portion of the returns to an asset was deleted while in a second step of the process a cumulative average of the residuals was computed from the single estimates of the "missing" data and the actual values.

In general, Brown and Kadiyala (1985) found that the SUR procedure was able to substantially reduce prediction error and that the sum of squared prediction errors for the single equation model was over a quarter larger than the SUR model. In their study, Brown and Kadiyala (1985) illustrated that it is possible to use a SUR model to assess information treated as unknown.

In summary, first it is important to recognize that researchers refer to the extra observations of one equation with respect to a second equation in a seemingly unrelated regressions (SUR) model as missing observations. Second, there are alternative estimators of the variance-covariance matrix of the error term $\mathbf{\Sigma}$ in the seemingly unrelated regressions (SUR) model. Zellner (1962) who first derived the seemingly-unrelated-regressions (SUR) method of estimating parameters proposed one alternative to estimate $\mathbf{\Sigma}$. Schmidt (1977) then proposed another five consistent estimators when there are unequal number of observations for each regression equation: the "usual" estimator which was similar to the original estimator presented by Zellner (1962), Wilks' (1932) estimator, Srivastava and Zaatar's (1973) estimator, the Hocking and Smith's (1968) and the maximum likelihood estimator (MLE). Then, Hwang (1990) proposed another alternative estimator based on Telser (1964).

With so many alternative ways to estimate $\mathbf{\Sigma}$, some researchers (Hwang, 1990; Baltagi et al., 1989; Schmidt, 1977) were motivated to study under what conditions one estimator will be better than the others. Surprising results were found. Baltagi et al. (1989) confirmed Schmidt's (1977) result that a feasible GLS estimator of the regression coefficients that ignores the extra observations in estimating $\mathbf{\Sigma}$ (but not necessarily in estimating $\mathbf{\Sigma}^{-1}$ or $\boldsymbol{\beta}$) compares favorably to a feasible GLS estimator of the regression coefficients that seem to use all extra observations. However, according to Hwang (1990) this does not mean that it can not be shown that under certain conditions an alternative estimator of $\mathbf{\Sigma}$ will perform better. Hwang (1990) showed that when the contemporaneous correlation between the error terms in a two-equation SUR model is high, the Telser's (1964) estimator of $\mathbf{\Sigma}$ dominates all the other estimators. Nonetheless, Baltagi et al. (1989), who did not used Telser's

(1964) estimator, showed that better estimates of $\mathbf{\Sigma}$ or $\mathbf{\Sigma}^{-1}$ need not imply better estimates of regression coefficients. This final result is supported by studies in panel data by Maddala and Mount (1973), Taylor (1980) and Baltagi (1981). Therefore, even though better estimates of $\mathbf{\Sigma}$ or $\mathbf{\Sigma}^{-1}$ can be used, better estimates of regression coefficients are not guaranteed.

Third, the SUR model with unequal number of observations is not restricted to time-series data. SUR model with unequal number of observations in panel-data studies have been studied among others by Fiebig and Kim (2000) and Baltagi et al. (1989). Finally, compared to the literature reviewed in Section 3.1, we now find alternative procedures to deal with censored data. However, no matter what of the procedures presented in this chapter is used, a feasible GLS estimator of the regression coefficients that ignores the extra observations in estimating $\mathbf{\Sigma}$ (but not necessarily in estimating $\mathbf{\Sigma}^{-1}$ or $\boldsymbol{\beta}$) compares favorably to a feasible GLS estimator of the regression coefficients that seem to use all extra observations. Consequently, econometric Software such as SAS, when estimating a SUR model with unequal number of observations, simply ignore the extra observations of one equation with respect to another one.

3.4 Missing Data

The term missing data is generally used instead of nonresponse. When the nonresponse rate is not negligible, inference based upon only the respondents may be seriously flawed. **Lohr (1999, p. 255)** explains two types of nonresponse: unit nonresponse and item nonresponse. **Unit nonresponse** occurs when when the entire observation unit is missing. For instance, the person provides no information for the survey. **Item nonresponse** occurs when some measurements are present for the observation unit but at least one item is missing. For instance, the person does not respond to a particular item in the questionnaire.

Lohr (1999, pp. 264–265) explains three different ways how the type of nonresponse (unit or item nonresponse) could be missing. Lohr (1999, p. 264) uses Little and

Rubin's (1987) terminology of nonresponse classification.

> **Missing Completely at Random** If [the probability that a unit i is selected for the sample and it will respond] does not depend on [the vector of known information about the unit i in the sample], [the response of interest], or the survey design, the missing data are missing completely at random (MCAR). Such a situation occurs if, for example, someone at the laboratory drops a test tube containing the blood sample of one of the survey participants—there is no reason to think that the dropping of the test tube had anything to do with the white blood cell count. If data are MCAR, the respondents are representative of the selected sample.
>
> **Missing at Random Given Covariates, or Ignorable Nonresponse** If [the probability that a unit i is selected for the sample and it will respond] depends on [the vector of known information about the unit i in the sample] but not on [the response of interest], the data are missing at random (MAR); the nonresponse depends only on observed variables. We can successfully model the nonresponse, since we know the values of [the vector of known information about the unit i in the sample] for all sample units. Persons in the [National Crime Victimization Survey (NCVS)] would be missing at random if the probability of responding to the survey depends on race, sex, and age-all known quantities-but does not vary with victimization experience within each age/race/sex class. This is sometimes termed **ignorable nonresponse**: Ignorable means that a model can explain the nonresponse mechanism and that the nonresponse can be ignored after the model accounts for it, [but it does not mean] that the nonresponse can be completely ignored and complete-data methods used.
>
> **Nonignorable Nonresponse** If the probability of nonresponse depends on the value of a response variable and cannot be completely explained by values of the [vectors of known information about the unit i in the sample], then the nonresponse is **nonignorable**. This is likely the situation for the NCVS: It is suspected that a person who has been victimized by crime is less likely to respond to the survey than a nonvictim, even if they share the values of all known variables such as race, age, and sex. Crime victims may be more likely to move after a victimization and thus not be included in subsequent NCVS interviews. Models can help in this situation, because the nonresponse probability may also depend on known variables but cannot completely adjust for the nonresponse.

Lohr (1999, pp. 255–288) discusses four approaches to deal with nonresponse:

1. Ignoring the nonresponse. This is not recommended.

2. Preventing the nonresponse by designing a survey so that the nonresponse is low. This is highly recommended.

3. Taking a representative subsample of the nonrespondents and use it to make inferences about the other nonrespondents.

4. Using models to predict values for the nonrespondents. Among these models Lohr (1999, pp. 265–288) discusses weighting methods, imputation methods, and parametric models for nonresponse.

The main problem caused by the nonresponse is potential bias of population estimates. The bias results when we estimate the population mean by using only the sample respondent mean and the population mean in the nonrespondent group differs from the population mean in the respondent group. Lohr (1999, p. 258) shows that the bias is small if either (1) the mean of the population nonrespondents is close to the mean for the population respondents or (2) the proportion of the population nonrespondents to the entire population is small (i.e., there is little nonresponse). Since it not possible to know (1), the only alternative is to reduce the nonresponse rate.

Designing the survey such that the nonresponse is low refers to carefully studying the best way to collect the data. This includes being able to anticipate and prevent reasons for nonresponse as much as possible. Lohr (1999, pp. 260–262) provides and discusses a list of factors that need to be examined: survey content, time of survey, interviewers, data-collection method, questionnaire design, respondent burden, survey introduction, incentives and disincentives, and follow up.

Lohr (1999, p. 263) explains Hansen and Hurwitz's (1946) procedure to subsample nonrespondents and to use two-phase sampling (also called double sampling) for stratifying and then estimating the population mean or total. In this procedure, an estimate of the population mean is obtained from a portion of the sample average of the original respondents and a portion of the average of the subsampled nonrespondents. These portions are the percentages of the sample that responded and not responded respectively. Similarly, an estimate of the population total can be obtained from a portion of the sample units in the respondent stratum and a portion of the sampled units in the nonrespondent stratum.

Weighting methods for nonresponse refer to incorporating weights in calculating population estimates of interest or to the use of weights to adjust for the nonresponse. Some weighting methods are weighting-class adjustment methods, postratification using weights, and weights that are the reciprocal of the estimated probability of response. A discussion and further references of these weighting methods are found in Lohr (1999, pp. 265–272). Lohr (1999, p. 272) explains weighting adjustments are usually used for unit nonresponse, not for item nonresponse (which would require a different weight for each item).

Imputation methods refer to alternative ways in which a nonresponse is replaced. The word imputation refers to substituting a missing value for a replacement value. Imputation methods are commonly used for item nonresponse. Lohr (1999, pp. 272–278) explains deductive imputation, cell mean imputation, hot-deck imputation, regression imputation, cold-deck imputation, and multiple imputation. In particular, regression imputation uses a regression of the item of interest on variables observed for all cases to predict the missing value. However, Lohr (1999, p. 278) explains that "[v]ariances computed using the data together with the imputed values are always too small, partly because of the artificial increase in the sample size and partly because the imputed values are treated as though they were really obtained in the data collection." Lohr (1999, p. 278) refers to Rao (1996) and Fay (1996) for a discussion on methods for estimating the variances after imputation.

Finally, parametric models for nonresponse refer to models that estimate within the model the nonresponse by using information on both known values of the variable of interest and missing values of the variable of interest (i.e., the nonresponse). That is, a model for the complete data is developed and components are added to the model to account for the proposed nonresponse mechanism. Depending on how good the model describes the data, the estimates of the variances that result from fitting the model may be better or worse. Examples can be found in Wooldridge (2006, pp. 609–613) and Pindyck and Rubinfeld (1997, pp. 325–331) who explain a censoring model and a maximum likelihood model respectively to address item non-response on the dependent variable.

3.5 Stratified Sampling

Lohr (1999, pp. 23–24) explains three basic types of probability samples.

- A **simple random sample** (SRS) is the simplest form of probability sample. An SRS of size n is taken when every possible subset of n units in the population has the same chance of being the sample... In taking a random sample, the investigator is in effect mixing up the population before grabbing n units. The investigator does not need to examine every member of the population for the same reason that a medical technician does not need to drain you of blood to measure your red blood cell count. Your blood is sufficiently well mixed that any sample should be representative.

- In a **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. The strata are often subgroups of interest to the investigator—for example, the strata might be different ethnic or age groups in a survey of people, different types of terrain in an ecological survey, or sizes of firms in a business survey. Element in the same stratum often tend to be more similar than randomly selected elements from the whole population, so stratification often increases precision.

- In a **cluster sample**, observation units in the population are aggregated into larger sampling units, called clusters. Suppose you want to survey Lutheran church members in Minneapolis but do not have a list of all church members in the city, so you cannot take an SRS of church members. However, you do have a list of all the Lutheran churches. You can then take an SRS of the churches and then subsample all or some church members in the selected churches. In this case, the churches form the clusters, and the church members are the observation units.

As it can be read, all these methods involve random selection of units to be in the sample. The key difference among them is in the level at which the random selection of units takes place. For instance, in an SRS, the observation units are randomly sampled from the population of observation units; in a stratified random sample, the strata are first selected and then the observation units within each stratum are randomly sampled; in a cluster sample, the clusters are first randomly selected from the population of all clusters and then all or some of the observation units are sampled. To illustrate this further, Lohr (1999, p. 24) provides a very useful example. Suppose you want to estimate the number of journal publications that professors at your university have. In an SRS, construct a list of all professors in your sample and randomly select n of them and ask them for the number of journal publications. In a stratified sample, classify faculty by college (agricultural sciences and natural resources, architecture, arts and sciences, business, education, engineering, human sciences, mass communications, etc.) and then take an SRS of faculty in the agricultural sciences and natural resources, another SRS of faculty in architecture, and so on. Finally, in a cluster sample, randomly select 10 of the 50 academic departments in the university and ask each professor in each selected department for his/her number of journal publications.

Lohr (1999, p. 95) further explains stratified random sampling. In stratified random sampling the strata do not overlap, and they constitute the whole population so that each sampling unit belongs to exactly one stratum. Lohr (1999, pp. 95–96) provides the following reasons to use stratified sampling:

1. To be protected from the possibility of obtaining a really bad sample that is not representative of the population.

2. To obtain data of known precision for subgroups. These subgroups should be the strata, which coincide with the domain of the study.

3. To reduce cost and increase ease of administration.

4. To obtain more precise (having lower variance) estimates for the whole population.

The sampling weight in stratified sampling is given by $w_{hj} = (N_h/n_h)$ (Lohr, 1999, p. 103), where $N = N_1 + N_2 + \ldots + N_H$ is the total number of units in the entire population, H is the number of "layers" (also called strata), N_h is the population units in the h^{th} stratum, and n_h is number of observations randomly sampled from the population units in stratum h. The sampling weight w_{hj} can be thought of as the number of units in the population represented by the sample unit j in stratum h or simply the sample member (h, j).[6] Additionally, Lohr (1999, p. 103) explains the probability of selecting the j^{th} unit in the h^{th} stratum to be in the sample is $\pi_{hj} = n_h/N_h$, which is also the sampling fraction in the h^{th} stratum. Hence, the sampling weight is the reciprocal of the probability of selection. That is, $w_{hj} = 1/\pi_{hj}$. Then, the sum of the sampling weights equals the population size. That is, $N = \sum_{h=1}^{H} \sum_{j \in \mathcal{S}_h} w_{hj}$, where $\mathcal{S}_h$ is the set of n_h units in the SRS for stratum h. "[If] each sampled unit 'represents' a certain number of units in the population,... the whole sample 'represents' the whole population" (Lohr, 1999, p. 103).[7]

It is very important that a statistician does not ignore the weights in a stratified sampling. A statistician who designs a survey to be analyzed using weights has implicitly visualized a model for the data. A sample is usually stratified and subpopulations oversampled precisely because researchers believe there will be differences among the subpopulations. Such differences also need to be included in the model. "A data analyst who ignores stratification variables and dependence among observations is not fitting a good model to the data but is simply being lazy" (Lohr, 1999, p. 229).

Lohr (1999, p. 229) recommends incorporating weights in calculating quantities

[6] As it will be discussed in Section 4.2, ENIGH calls the sampling weight the "expansion factor" (i.e., the number of households that a particular household represent nationally).

[7] As it will be mentioned in Section 4.2, according to ENIGH—Síntesis Metodológica (2006), the results obtained from ENIGH survey can be generalized to the entire Mexican population.

such as means, medians, quantiles, totals, and ratios. One way to estimate these quantities is by incorporating the stratification variables (Lohr, 1999, pp. 95–130). Another way to estimate these quantities (but not their standard errors) is by constructing an empirical distribution for the population from the sampling weights. "The statistics calculated using weights are much closer to the population quantities" (Lohr, 1999, p. 234).

Lohr (1999, pp. 347–378) also explores how to do regression in complex survey samples. She explains that even though there is debate whether the sample sampling weights are relevant for inference in regression (Lohr, 1999, p. 363), the data structure needs to be taken into account in either approach. She explains two things can happen in complex surveys (Lohr, 1999, pp. 352–253):

> 1. Observations may have different probabilities of selection, π_i. If the probability of selection is related to the response variable y_i, then an analysis that does not account for the different probabilities of selection may lead to biases in the estimated regression parameters.
>
> 2. Even if the estimators of the regression parameters are approximately design unbiased, the standard errors given by SAS or SPSS will likely be wrong if the survey design involves clustering. Usually, with clustering, the design effect (deff) for regression coefficients will be greater than 1.

Lohr (1999, p. 355) recommends, "[i]n practice, use professional software designed for estimating regression parameters in complex surveys. If you do not have access to such software, use any statistical regression package that calculates weighted least squares estimates. If you use weights w_i in weighted least squares estimation, you will obtain the same point estimates...; however, in complex surveys, the standard errors and hypothesis tests the software provides will be incorrect and should be ignored." Lohr (1999, pp. 289–318) explains several methods for estimating variances of estimated totals and other statistics from complex surveys. She explains linearization (Taylor Series) methods, random group and resampling methods (balanced repeated replication, the Jacknife, and the Bootstrap) for calculating variances of nonlinear statistics. In addition, she also explains the calculation of generalized variance functions and how to construct confidence intervals. For more information on these methods refer to Lohr (1999, pp. 298–318).

Wooldridge (2002, p. 551) explains that there are a variety of selection mechanisms that result in **nonrandom samples** (also called **selected samples**). Some of these are due to sample design, while others are due to the behavior of the units being sampled, including nonresponse on survey questions and attrition from social programs (i.e., in panel data where people leave the sample entirely and usually do not reappear in later years). Wooldridge (2002, p. 552) explains that in some cases, the fact that we have a nonrandom sample does not affect the way we estimate population parameters. Wooldridge (2002, pp. 552–558) provides conditions under which estimating the population model with linear and nonlinear models using nonrandom sample is consistent for the population parameters. For an explanation of these conditions refer to Wooldridge (2002). Wooldridge (2002, pp. 558–590) also explains how to deal with nonrandom samples on the basis of the response variable, how to do nonrandom sample corrections with a probit or tobit model under exogenous or endogenous explanatory variables, and how to deal with other nonrandom sample issues.

Wooldridge (2002, p. 590) explains **stratified samples** are a form of nonrandom samples. In stratified samples different subsets of the population are sampled with different frequencies. Stratification can be based on exogenous variables or endogenous variables or a combination of these. Wooldridge (2002, p. 596) explains that when $\mathbf{x}$ is exogenous (see Wooldridge 2002, p. 596 for the sense in which $\mathbf{x}$ must be exogenous) and stratification is based entirely on $\mathbf{x}$, the standard unweighted estimator on the stratified sample is consistent and asymptotically normal. In addition, **Wooldridge (1999)** shows that the usual asymptotic variance estimators are valid when stratification is based on $\mathbf{x}$ and we ignore the stratification problem. In this case the usual conditional maximum likelihood analysis holds, and in the case of regression the usual heteroskedasticity robust variance matrix estimator can be used

(Wooldridge, 1999, p. 597).

Two common kinds of stratification are discussed by Wooldridge (2002, pp. 590–591): standard stratified sampling (SS sampling) and variable probability sampling (VP sampling).

> In **SS sampling**, the population is first partitioned into J groups, $\mathcal{W}_1$, $\mathcal{W}_2$, ..., $\mathcal{W}_J$, which are assumed to be nonoverlapping and exhaustive. We let $\mathbf{w}$ denote the random variable representing the population of interest... For $j = 1, \ldots, J$, draw a random sample of size N_j from stratum j. For each j, denote this random sample by $\{\mathbf{w}_{ij}$: $i = 1, 2, \ldots, N_j\}$. The strata samples sizes N_j are nonrandom. Therefore, the total sample size, $N = N_1 + \ldots + N_J$, is also nonrandom. A randomly drawn observation from stratum j, $\mathbf{w}_{ij}$, has distribution $\mathrm{D}(\mathbf{w}|\mathbf{w} \in \mathcal{W}_j)$. Hence, the observations within a stratum are identically distributed but observations across strata are not.

Notice that Wooldridge's (2002) definition of SS sampling is the same as Lohr (1999) definition of stratified random sampling. Now, consider Wooldridge's (2002, p. 591) explanation of variable probability sampling (VP sampling).

> [In **VP sampling**,] an observation is drawn at random from the population. If the observation falls into stratum j, it is kept with probability p_j. Therefore, random draws from the population are discarded with varying frequencies depending on which stratum they fall into. This kind of sampling is appropriate when information on the variable or variables that determine the strata is relatively easy to obtain compared with the rest of the information. Survey data sets, including interviews to collect panel or longitudinal data, are good examples. Suppose we want to oversample individuals from, say, lower income classes. We can first ask an individual her or his income. If the response is in income class j, this person is kept in the sample with probability p_j, and then the remaining information, such as education, work history, family background, and so on can be collected; otherwise, the person is dropped without further interviewing.

It is important to notice that in VP sampling the observations within a stratum are discarded randomly. Wooldridge (1999) discusses why VP sampling is equivalent to the procedure in Table 3.1.

The number of observations falling into stratum j is denoted by N_j, the number of data points we actually have for estimation is $N_0 = N_1 + N_2 + \ldots + N_J$, and N is

Table 3.1: Variable Probability Sampling (VP Sampling)

Repeat the following steps N times
1. Draw an observation $\mathbf{w}_i$ at random from the population.
2. If $\mathbf{w}_i$ is in stratum j, toss (a biased) coin with probability p_j of turning up heads. Let $h_{ij} = 1$ if the coin turns up heads and zero otherwise.
3. Keep observation i if $h_{ij} = 1$; otherwise, omit it from the sample.

Source: Wooldridge (2002, p. 591).

the number of times the population is sampled. Wooldridge (2002, p. 592) explains that if N is fixed, then N_0 is a random variable. It is not known what each N_j would be prior to sampling.

In VP sampling, Wooldridge (2002, p. 594) shows that in estimating the following linear model by weighted least squares (WLS),

$$y = \mathbf{x}\boldsymbol{\beta}_0 + u, \quad E(\mathbf{x}'u) = \mathbf{0}, \tag{3.1}$$

where $\mathbf{x}$ is a $(1 \times K)$ vector of explanatory variables, y is a scalar response variable, and u is a scalar disturbance variable; the asymptotic variance estimator is

$$\left(\sum_{i=1}^{N_0} p_{j_i}^{-1}\mathbf{x}_i'\mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^{N_0} p_{j_i}^{-2}\hat{u}_i^2\mathbf{x}_i'\mathbf{x}_i\right) \left(\sum_{i=1}^{N_0} p_{j_i}^{-1}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}, \tag{3.2}$$

where $\hat{u}_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}_w$ is the residual after WLS estimation, $p_{j_i}^{-1}$ the weight attached to observation i in the estimation, and j_i the stratum for observation i. Wooldridge (2002, p. 593) explains that in practice, the $p_{j_i}^{-1}$ are the sampling weights reported with other variables in stratified samples. Additionally, Wooldridge (2002, p. 594) explains that this asymptotic variance matrix estimator is simply White's (1980) heteroskedastic-consistent covariance matrix estimator applied to the stratified sample, where all variables for observation i are weighted by $p_{j_i}^{-1/2}$ before performing the regression. This estimator has also been suggested by Hausman and Wise (1981). Additionally, Wooldridge (2002, p. 54) remarks that it is important to remember that the asymptotic variance matrix estimator above is not due to potential heteroskedasticity in the underlying population model. Even if $E(u^2|\mathbf{x}) = \sigma_0^2$, the estimator in equation (3.1) is generally needed because of the stratified sampling. Wooldridge (2002, p. 54) explains this estimator works in the presence of heteroskedasticity of arbitrary and unknown form in the population, and it is routinely computed by many regression packages.

The weights in SS sampling are different from those in the VP sampling. In SS sampling the weights are (Q_{j_i}/H_{j_i}) rather than $p_{j_i}^{-1}$, where j_i denotes the stratum for observation i, $Q_j = \mathrm{P}(w \in \mathcal{W}_j)$ denotes the population frequency for stratum j (it is assumed that Q_j are known), and $H_j = N_j/N$ denotes the fraction of observations in stratum j. Additionally, the formula for the asymptotic variance is different.

In SS sampling, **Wooldridge (2001, p. 464)** shows that in estimating the linear model in equation (3.1) above, the weighted estimator is consistent for $\boldsymbol{\beta}_0$. Additionally, if the stratification is exogenous and $\mathrm{E}(u|\mathbf{x}) = 0$, the asymptotic variance matrix estimator of $\hat{\boldsymbol{\beta}}_w$ can be written as

$$\left(\sum_{i=1}^{N}(Q_{j_i}/H_{j_i})\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(\sum_{i=1}^{N}(Q_{j_i}/H_{j_i})^2\hat{u}_i^2\mathbf{x}_i'\mathbf{x}_i\right)\left(\sum_{i=1}^{N}(Q_{j_i}/H_{j_i})\mathbf{x}_i'\mathbf{x}_i\right)^{-1}, \tag{3.3}$$

which is again simply White's (1980) heteroskedasticity-consistent covariance matrix estimator applied to the stratified sample, where all variables for observation j are weighted by $(Q_{j_i}/H_{j_i})^{-1/2}$ before performing the regression.

Wooldridge (2002, pp. 595–596) comments that if the population frequencies Q_j are known in VP sampling, he recommends using as weights $Q_j/(N_j/N_0)$ rather than p_j^{-1}. His recommendation is based on his findings in Wooldridge (1999). Additionally, Wooldridge (2002, p. 596) explains that when the sampling weights Q_{j_i}/H_{j_i} or $p_{j_i}^{-1}$ and the stratum are given, the weighted M-estimator under SS or VP sampling is fairly straightforward, but it is not likely to be efficient. It is possible to do better with conditional maximum likelihood (Imbens and Lancaster, 1996).

Summarizing, when dealing with stratified sampling, the weighted estimator is consistent (Wooldridge, 2001, p. 464). "If [we] use weights w_i in the weighted least squares estimation, [we] will obtain the same point estimates...; however, in complex surveys, the standard errors and hypothesis tests the software provides will be incorrect and should be ignored" (Lohr, 1999, p. 355). Hence, we briefly mentioned procedures that can be used to estimate standard errors and hypothesis tests. Of particular interest, Lohr (1999, pp. 298–308) points to the use of Bootstrap or Jacknife in complex survey designs. We also provided Wooldridge's (2002; 2001) estimators of asymptotic variances. However, since Wooldridge's (2001) SS sampling estimator of asymptotic variances is not in the context of seemingly unrelated regressions (i.e., do not deal with the estimation of a system of equations), Lohr's (1999, pp. 306–307) bootstrap procedure is more general and appropriate to the specific objective of this study of providing an empirical application of a seemingly unrelated regression model. Therefore, the bootstrap procedure will be adopted in this study.

3.6 The Bootstrap

The bootstrap was first proposed by Efron (1979). Then, further theory was presented by Singh (1981), Bickel and Freedman (1981), and Efron (1982). Efron and Tibshirani (1993) provided a good introductory statistics treatment. Other studies, mentioned in the literature bellow, include Freedman (1984), Sitne (1990), Hall (1992), Dixon (1993), Hjorth (1994), Brownstone and Kazimi, and Mackinnon (2002).

Cameron and Trivedi (2005, p. 355) explain that "bootstrap methods for statistical inference... have the attraction of providing a simple way to obtain standard errors when the formulae from asymptotic theory are complex." There is a wide range of bootstrap methods. Cameron and Trivedi (2005, p. 357) classify the wide range of bootstrap method into two broad approaches. "First, the simplest bootstrap methods can permit statistical inference when conventional methods such as standard error computation are difficult to implement. Second, more complicated bootstraps can have the additional advantage of providing asymptotic refinements that can lead to a better approximation in finite samples."

Lohr (1999, p. 306) explains the bootstrap for an simple random sample (SRS) with replacement. When applying the bootstrap for an SRS with replacement, we hope that it will reproduce properties of the whole population. Lohr (1999, p. 306) provides the following example. Suppose $\mathcal{S}$ is an SRS of size n. The sample $\mathcal{S}$ is treated as if it were a population, and resamples from $\mathcal{S}$ are taken. If the sample really is similar to the population—if the empirical probability mass function (epmf) of the sample is similar to the probability mass function of the population—then samples generated from the epmf should behave like samples taken from the population.

Lohr (1999, p. 307) further explains that after a total of B SRSs with replacement are taken from $\mathcal{S}$ (i.e., B resamples), the bootstrap distribution of the parameter of interest is calculated. Then, this distribution may be used to calculate a confidence interval directly. A 95% confidence interval is calculated by finding the 2.5 percentile and 97.5 percentile of the bootstrap distribution of the parameter of interest.

The bootstrap for an SRS can also be without replacement (Lohr, 1999, p. 307). Gross (1980) discusses some properties of with-replacement and without-replacement bootstrap distributions. When the original SRS is without replacement, Gross (1980) proposes creating N/n copies of the sample to form a "pseudopopulation," where N denotes the population size, and then drawing B SRSs without replacement from the pseudopopulation. When n/N is small, the with-replacement and without-replacement bootstrap distribution should be similar (Lohr, 1999, p. 307).

Bootstrap methods for statistical inference in the context of stratified samples have also been studied. For example, Rao and Wu (1988) explain rescaling bootstrap methods for a stratified random sample, Sitter (1992) describes and compares three bootstrap methods for complex surveys, and Shao and Tu (1995) summarize theoretical results for the bootstrap in complex survey samples.

Cameron and Trivedi (2005, p. 358) summarize key bootstrap methods for an estimator $\hat{\boldsymbol{\theta}}$ and associated statistics based on an iid sample $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$, where usually $\mathbf{w}_i = (y_i, \mathbf{x}_i)$ and $\hat{\boldsymbol{\theta}}$ is a smooth estimator that is $\sqrt{N}$ consistent and asymptotically normally distributed.[8] For notational simplicity they generally presented

[8]Cameron and Trivedi (2005, p. 358) use N to denote the bootstrap sample size. If N denotes the population size, and a bootstrap sample size n is desired, then replace N by n.

results for scalar θ. For vector $\boldsymbol{\theta}$ in most instances the replacement of θ by θ_j, the j^{th} component of $\boldsymbol{\theta}$ is required. Statistics of interest include the usual regression output: the estimate $\hat{\theta}$; standard errors $s_{\hat{\theta}}$; t-statistic $t = \frac{(\hat{\theta}-\theta_0)}{s_{\hat{\theta}}}$, where θ_0 is the null hypothesis value; the associated critical value or p-value for this statistic; and confidence interval.

A general **bootstrap algorithm** is presented by Cameron and Trivedi (2005, p. 360):

1. Given data $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N$ draw a bootstrap sample [of] size N using a method given [below] and denote this new sample $\mathbf{w}_1^*, \mathbf{w}_2^*, \ldots, \mathbf{w}_N^*$.
2. Calculate an appropriate statistic using the bootstrap sample. Examples include (a) the estimate $\hat{\theta}^*$ of θ, (b) the standard error $s_{\hat{\theta}^*}$ of the estimate $\hat{\theta}^*$, and (c) a t-statistic $t^* = \frac{(\hat{\theta}^*-\hat{\theta})}{s_{\hat{\theta}^*}}$ centered at the original estimate $\hat{\theta}$. Here $\hat{\theta}^*$ and $s_{\hat{\theta}^*}$ are calculated in the usual way but using the new bootstrap sample rather than the original sample.
3. Repeat steps 1 and 2 B independent times, where B is a large number, obtaining B bootstrap replications of the statistic of interest, such as $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$ or $t_1^*, t_2^*, \ldots, t_B^*$.
4. Use these B bootstrap replications to obtain a bootstrapped version of the statistic.

The following **bootstrap sampling methods** are explained by Cameron and Trivedi (2005, p. 360):

- **Empirical distribution function (EDF) bootstrap or nonparametric bootstrap.**

  The simplest bootstrapping method is to use the empirical distribution of the data, which treats the sample as being the population. The $\mathbf{w}_1^*, \mathbf{w}_2^*, \ldots, \mathbf{w}_N^*$ are obtained by sampling with replacement from $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N$. In each bootstrap sample so obtained, some of the original data points will appear multiple times whereas others will not appear at all... [This method] is also called a **paired bootstrap** since in single equation regression models $\mathbf{w}_i = (y_i, \mathbf{x}_i)$, so here both y_i and $\mathbf{x}_i$ are resampled.

- **Parametric bootstrap.**

Suppose the conditional distribution of the data is specified, say $y|\mathbf{x} \sim F(\mathbf{x}, \boldsymbol{\theta}_0)$, and an estimate $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ is available. Then in step 1 we can instead form a bootstrap sample by using the original $\mathbf{x}_i$ while generating y_i by random draws from $F(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$. This corresponds to regressors fixed in repeated samples [see Cameron and Trivedi (2005, Section 4.4.5)]. Alternatively, we may first resample $\mathbf{x}_i^*$ from $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ and then generate y_i from $F(\mathbf{x}_i^*, \hat{\boldsymbol{\theta}})$, $i = 1, 2, \ldots, N$. Both... examples... can be applied in fully parametric models.

- **Residual bootstrap.**

For regression model with additive iid error, say $y_i = g(\mathbf{x}_i, \boldsymbol{\beta}) + u_i$, we can form fitted residuals $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_N$, where $\hat{u}_i = y_i - g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})$. Then in step 1 bootstrap from these residuals to get a new draw of residuals, say $(\hat{u}_1^*, \hat{u}_2^*, \ldots, \hat{u}_N^*)$, leading to a bootstrap sample $(y_1^*, \mathbf{x}_1), (y_2^*, \mathbf{x}_2), \ldots, (y_N^*, \mathbf{x}_N)$, where $y_i^* = g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}) + u_i^*$. [The residual bootstrap] uses information intermediate between the nonparametric and parametric bootstrap. It can be applied if the error term has distribution that does not depend on unknown parameters.

In this study, the first bootstrap sampling method is used. According to Cameron and Trivedi (2005, p. 361), "the paired bootstrap... appli[es] to a wide range of nonlinear models, and reli[es] on weak distributional assumptions." However, according to Cameron and Trivedi (2005, p. 361), the other bootstraps generally provide better approximations (see Horowitz, 2001, p. 3185).

Particularly, this study uses the %BOOT macro developed by SAS Online Support (Accessed July 1, 2008). "The %BOOT macro does elementary nonparametric bootstrap analyses for simple random samples, computing approximate standard errors, bias-corrected estimates, and confidence intervals assuming a normal sampling distribution" (SAS Institute Inc., p. 1). Additionally, this study resamples observations and the %BOOT macro executes a macro loop that generates and analyzes the resamples one at time. Moreover, with the %BOOT macro "[e]ither method of resampling for regression models (observations or residuals) can be used regardless of the form of the error distribution. However, residuals should be resampled only if the errors are independent and identically distributed and if the functional form of the model is correct within a reasonable approximation. If these assumptions are questionable, it is safer to resample observations" (SAS Institute Inc., p. 8). Finally, the default size of each resample used by the %BOOT macro is equal to the size of the input dataset from which the rample is being taken. For detailed information about the %BOOT macro refer to SAS Online Support.

CHAPTER IV
METHODS AND PROCEDURES

This chapter starts by explaining the Mexican database on household income and expenditures that is used in the study. In particular, Section 4.1 explains what type of information is contained in the database, the sampling methods used to collect the data, how the data is collected, and the activities performed to preserve the quality of the data. Additionally, Section 4.1 explains how the Mexican database is divided into seven datasets. Then, Section 4.2 begins by explaining the variables from the seven datasets that are used in the study. It continues to give details about how new variables are created or transformed from the variables provided. In addition, it reports the difficulties that emerge as the data is organized in the desired manner. Further, it describes the procedure adopted to reduce the number of missing observations and how this study stayed away from price imputations. Finally, given the outcome of Section 4.2, Section 4.3 specifies the SUR models that will be estimated. In particular, Section 4.3.1 provides one general model while Section 4.3.2 explains how individual models will be estimated for each urbanization level within each Mexican region.

4.1 Data

Mexican data on household income and expenditures was obtained from *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)*. This nation-wide survey is published since 1984 by *Instituto Nacional de Estadística, Geografía e Informática* (INEGI). Even though ENIGH is available for the years 1984, 1989, 1992, 1994, 1998, 2000, 2002, 2004 and 2006, this study only uses data for the year 2006, which was collected between August and November 2006.

ENIGH nation-wide Mexican household survey encompasses Mexico's 31 states and the Federal District (a territory which belongs to all states), and contains information about house infrastructure, appliances and services as well as household members demographic and socio-demographic characteristics and occupational activities. Particularly, ENIGH contains information about household incomes, and quantities and prices of goods purchased.

According to ENIGH—*Síntesis Metodológica*, ENIGH's sampling methods are probabilistic, multi-staged, stratified, and conglomerated. According to *Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH)— Síntesis Methodológica (2006)*, the sampling method is *probabilistic* because the sampling units have a probability of being selected, which is known and different from zero. Additionally, the sampling method is *multi-staged* because the sampling units are selected in multiple stages. It is *stratified* because the target population is divided into groups with similar characteristics, which form the strata. Finally, it is *conglomerated* because the sampling units (households) are made up from the observation units (household members). However, for some data the observation unit is the household. For example, each household contains information on its members about age, gender, marital status, etc., but information on food expenditures is recorded for the household unit only.

Results obtained from the survey can be generalized to the entire population (ENIGH—*Síntesis Metodológica, 2006*). ENIGH chooses households for interview and reports information mainly for the household unit. Excluded from the analysis are diplomatic foreign homes and homes maintained by companies for business-related purposes. Additionally, ENIGH is based on the international recommendations of the United Nations (UN) and the International Labour Organization (ILO). Furthermore, it is articulated to the Mexican governmental institutions and surveys accomplished by INEGI.

In order to collect the data, ENIGH performs direct interviews to each household during one week, usually from August to November. The workforce is organized into interviewers, supervisors, and state project managers. Two instruments are used to collect the data: a questionnaire and a journal. The questionnaire is designed to collect the data concerning the house infrastructure, the members and their household identification, and members' socio-demographic characteristics. In addition, for household members older than 12 years old, the questionnaire will capture occupational activities and related characteristics as well as income and expenditures. On the other hand, the journal is designed to collect at-home and away-from-home expenditures on food, drinks, cigarettes and public transportation. During the first day of interview, these latter expenditures are recorded in the journal by the interviewer in order to train the interviewee. The journal remains with, and is filled by, the interviewee for the next six days of the week (INEGI, personal contact). Hence, data on food, drinks, cigarettes and public transportation is recorded in the Expenditure dataset (see Table 4.1) only when the household makes a purchase.[1] However, the interviewer will visit the household each day until the end of the week of the interview in order to continue training the interviewee and make sure expenditures on food, drinks, cigarettes and public transportation are correctly being recorded by the interviewee in the journal (INEGI, personal contact). In the first day of interview, food that already belonged to the household, before the interviewer arrived, is recorded in the journal only if the food was acquired the day before the interviewer arrived (INEGI, personal contact).

To assure the quality of the data during the collection period, the following supervising activities are performed: a) registering the questionnaire and journal by an id number, which contains the year, state, stage, consecutive number and type of home; b) controlling the number of homes in the framework; c) verifying the nonresponse; d) observing directly the interview and supervisor; and e) applying a re-interview ques-

[1] In Section 3.1, this problem was referred as censored data. Additionally, although ENIGH will not record meat cuts that the household did not buy during the week of the interview, if we consider Section 3.4's terminology, there will be item nonresponse in some variables (e.g. place of purchase, price, quantity, expenditure, etc.), but we can still recover other variables (e.g. the "expansion factor", stratum, household size, etc.). Now, if we look at the Expenditure dataset as it is reported by ENIGH and consider the demand of certain items as equations (i.e., quantity as a function of prices and income), there will be equations with unequal number of observations as discussed in Section 3.3.

tionnaire to completed interviews. After the data is collected, it is carefully entered into the database, which is then electronically validated. In case of omitted item observations, incomplete observations, errors or inconsistent information, the data is verified via phone or by returning to the collection field. When it is not possible to have a 100% response rate, a nonreponse rate is reported. In ENIGH 2006, there was a nonresponse rate of 10.55%.

The ENIGH database is divided into seven datasets as described in Table 4.1. The observation unit for the Concentrated, Household, Expenditures, and Financial Transactions datasets is the household, while the observation unit for the Members and Incomes datasets is the household member. For the No Monetary Transactions dataset the observation unit is the household or the household member. For more detailed information, the reader should refer to ENIGH.

Table 4.1: List of the Seven Datasets in ENIGH 2006 Database.

Dataset	Number of Records in 2006	General Description
Concentrated (concentrado.dbf)	20,875	Information about the expansion factor (number of households that a particular household represent nationally) and other variables that appear in the other six datasets.
Households (hogares.dbf)	20,875	Information about the household geographical location, household stratum, house infrastructure, utilities, home vehicles and home appliances, etc.
Members (poblacion.dbf)	83,624	Information about number of household members, relationships among household members, gender, age, city of residency, level of education, marital status, employment status, job position, if member has salary/wages, job description, weekly number of workdays, if member has social security contributions, etc.
Income (ingresos.dbf)	79,752	Information about type of employment, current income, income one, two, three, four, five and six months ago, quarterly income, etc.
Expenditures (gastos.dbf)	1,348,530	Information about items purchased, place of purchase, day of purchase, payment option, quantity, cost, price, expenditure, last month expenditure, quarterly expenditure, and frequency of purchase.
Financial Transactions (erogaciones.dbf)	18,269	Information about bank deposits, loans, credit card payments, debt with employer, interest payment, purchase of local and foreign currency, purchase of jewelries, life insurance, money inherited, purchase of houses, purchase of condominiums, purchase of land, mortgage payments, others, equipment purchases, stock investment, patent investments, etc.
No Monetary Transactions (nomonetario.dbf)	174,490	Information about the type of expenditure, reason of purchase, day of purchase, quantity, price, expenditure, and quarterly expenditure.

Source: ENIGH 2006, summarized by author.

4.2 Procedures

As explained in Chapter I, the Mexican meat consumption will be analyzed in this research. In the previous section, Table 4.1 listed the seven datasets in ENIGH 2006 database. However, in order to provide an empirical application of a SUR model using ENIGH 2006 database, the variables of interest need to organized in one dataset first. Table 4.2 list the variables, from the seven ENIGH 2006 datasets, used in this study.

However, before putting all variables of interest together in one dataset, a new variable (the number of adult equivalents per household) needs to be computed from the "edad" variable in the Members dataset. As explained in Section 3.2, adult equivalence scales are used to compute the number of adult equivalents per household by taking into account how much an individual household member of a given age and sex contributes to household expenditures or consumption of goods relative to a standard household member. Adult equivalents are computed to be able to compare household consumption. For instance, meat consumption in different households cannot be directly compared without computing per capita meat consumption because a bigger households will naturally have a tendency to consume more meat than smaller households. Not adjusting meat consumption and expenditures by adult equivalents presents a problem when estimating quantity consumed (quantity demand) as a function of prices and total expenditure. For example, suppose one household demands q amount of beef and suppose a bigger household who pays a higher price demands more beef. If we compare these two households without adjusting by adult equivalents, price increases but does quantity decrease? On the other hand, adjusting by adult equivalents (i.e., computing per capita beef consumption) in our example, price will always increase but this time, quantity will decrease. Hence, this study used the 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) to compute the number of adult equivalents.

After computing the number of adult equivalents per household, all nominal variables[2] (nominal prices and nominal expenditures) are transformed into real variables[3] (real prices and real expenditures). Real prices and real expenditures are computed as follows:

$$p_i^{real} = \frac{p_i^{nominal}}{\left(\frac{CPI_{2006}}{100}\right)}, \qquad m_i^{real} = \frac{m_i^{nominal}}{\left(\frac{CPI_{2006}}{100}\right)}, \tag{4.1}$$

where $i = A025, A026, \ldots, A074$ and CPI_{2006} is the simple average of the consumer price index (CPI) for the months of August, September, October, and November 2006[4] as reported by Banco de Mexico. The base period for the CPI of Banco de Mexico is the second half of June 2002. Hence, the new prices of each meat cut[5] and the new expenditures on each meat cut become the real price of each meat cut in 2002 Mexican pesos per kilogram and the real expenditure of each meat cut in 2002 Mexican pesos per household respectively.

Then, the meat consumption variables in kilograms per household are divided by the number of adult equivalents to compute per capita meat consumption variables in kilograms. Similarly, the new real expenditure variables are divided by the number of adult equivalents to obtain per capita real expenditure variables in 2002 Mexican pesos.

Descriptive statistics for each meat cut with the original number of observations as reported by ENIGH 2006 are provided in Table 4.4 through 4.53. In Tables 4.4 through 4.53, p_i is the real price of meat cut i in 2002 Mexican pesos per kilogram, q_i is the per capita consumption of meat cut i in kilograms, m_i is the per capita real

[2]A nominal variable is a variable whose unit of measurement is in nominal economic value. In economics, nominal value is the value of anything expressed in money of the day. A nominal variable does not adjust for inflation. For example, nominal price does not adjust for inflation.

[3]In economics, real value is the value of anything expressed in the nominal value of that anything in the base period. For example, real price adjusts for inflation; therefore, it is expressed in the nominal price of the base period.

[4]The survey period of ENIGH 2006 was from August to November 2006. The simple average CPI of these months was used in order to use the same months when ENIGH 2006 collected the data.

[5]As it can be observed from Table 4.3, sometimes a specific code, for instance A025, may refer to more than one cut of meat. However, to facilitate the flow of the discussion in this study, a code such as A025 will be referred as if it were only one meat cut.

expenditure on meat cut i in 2002 Mexican pesos, and $i = 025, 026, \ldots, 074$ stands for the corresponding meat cuts $A025, A026, \ldots, A074$ provided in Table 4.3. The number of observations for each meat cut varies because ENIGH interviewers only recorded a transaction when a household consumed a meat cut. Hence, the meat cuts that are consumed most often by households are those who have the largest number of observations. Additionally, as mentioned in Section 4.1 at-home and away-from-home expenditure on food is collected by ENIGH 2006. In particular, expenditures on food at home and away from home are identified by the variable "place of purchase" in the Expenditures dataset (see Table 4.1). Even though this variable was not included in this study, it is important to mention that the descriptive statistic in Table 4.4 through 4.53 include as different observations purchases made at different places by the same household of the same meat cut. That is, if during the week of the interview a household purchased the same meat cut twice but at different places, then two transactions will be recorded appearing as two observations. However, this method of recording transactions has no distorting effect on the descriptive statistics provided in Table IV.4 through IV.53. Finally, the "N Miss" column reports the number of missing observations for each of the three variables reported. Only for meat cuts $A057$, $A068$, and $A070$, households failed to report both price and quantity, but yet reported meat expenditure. Hence, the number of missing observations due to household not reporting prices and quantities occurs very rarely. However, once again, the number of observations in each table reflects the number of times each meat cut was reported by all households, including more than one record per household.

Now, if we would like to put all meat cut datasets into one dataset where the columns of this new dataset are the prices, quantities, and expenditures of each meat cut; then, only one transaction per meat cut per household has to be allowed. To do this, when a household purchased the same meat cut during the week of the interview more than once but in different places, a simple average of the same meat cut is computed, but the sum of the quantity is computed, and expenditure is computed as price times quantity. Once again, doing this operation will only allow one transaction per meat cut per household. This is required in order to combine all datasets using a one-to-one match merge by household id. Since each meat cut dataset has four columns (household id, price, quantity, and expenditure), a one-to-one match merge by household id will produce a dataset with 50(3)+1=151 columns. Assuming all households purchased at least one meat cut, then the number of rows of this dataset equals the number of households. Additionally, when a household did not consume a particular meat cut, for instance $A025$, but consumed all others meat cuts, then a missing value appears in that row for the columns corresponding to the real price of meat cut $A025$ ($p025$), the per capita consumption of meat cut $A025$ ($q025$), and the per capita real expenditure on meat cut $A025$ ($m025$); but the corresponding numeric value for all other columns. However, some households will not consume any meat cut at all during the week of the interview and several households will only consume few (in some cases only one) meat cut during the week of the interview. Hence, the dataset will have a lot of missing observations for the corresponding columns of meat cuts that are rarely consumed; but a moderate amount of missing observations for the corresponding columns of the most frequently consumed meat cuts. Table 4.54 shows the descriptive statistics of this dataset.

Once again, Table 4.54 was generated by allowing only one transaction per meat cut per household[6] and then by performing a one-to-one match merge by household id to merge all meat cut datasets. Since we know that $20,875$ households participated in the survey (Table 4.1), this means that $20,875 - 16,909 = 3,966$ households of the total number of households that participated in the survey did not consume any meat cut at all during the week of the interview. In addition to this information, Table 4.54 also shows the new number of missing observations (column "N Miss") of the price, quantity, and expenditure of meat cut i, $i = 025, 026, \ldots, 074$, resulting from the merge of all meat cut datasets. Clearly, the number of missing observations is extremely high compared to the total number of observations, which is 16,909.

[6]That is, by recalculating the corresponding datasets of the descriptive statistics in Tables 4.4 through 4.53, but this time only allowing one transaction per meat cut per household.

However, a missing quantity in Table 4.54 is simply a decision of a household of not to purchase that particular meat cut during the week of the interview. Hence, missing quantities in Table 4.54 are transformed to zero quantities. Finally, it is very important to notice that the sum of weights in Table 4.54 is an estimate of the total number of households in Mexico that consumed meat during the week of the interview. That is, 22.1 million households eat at least one meat cut during the week of the interview.

To reduce this high number of missing price observations, the meat cuts can be aggregated according to the meat categories reported in Table 4.3. That is, instead of having 50 meat cuts, only 6 meat commodities can be considered: beef, pork, processed meat, chicken, other meat, and seafood. Beef including section (a) of Table 4.3; pork including section (b); processed meat including section (c) and (e); chicken including section (d); including sections (g), (h), (i), (j), and (k); and other meat including section (f). In order to aggregate the corresponding meat cuts in these six new categories, the corresponding quantities of each new category are obtained by summing all corresponding meat-cut quantities that belong to that category, while the corresponding prices of each new category are computed by diving total expenditure by total quantity of each category. Finally, total meat expenditure is computed by $\sum_{i=1}^{6} p_i q_i$, where $1 =$ beef, $2 =$ pork, $3 =$ processed meat, $4 =$ chicken, $5 =$ other meat, and $6 =$ seafood.

Table 4.55 reports the new number of missing and non-missing observations per stratum when six commodities are considered. Other meat category has the largest number of missing observations. This is not surprising because only three meat cuts are in this category ($A063$, $A064$, and $A065$) and mainly because these three meat cuts represent exotic meats (lamb, goat, horses, iguana, etc.). Excluding other meat from the analysis, notice that stratum one, two and three have in three occasions (p_{beef}, $p_{process}$, and $p_{chicken}$) more non-missing than missing observations. However, stratum four has only one occasion ($p_{chicken}$) where there are more non-missing than missing observations. Additionally, the reader may be surprised at this point that pork has more missing than non-missing observations in all four strata given that we are talking about Mexicans. Going back to Table 4.3, the reader will realize that this is because several pork cuts ($A048$, $A049$, $A052$, and $A054$) are included in the processed meat category. The total intersection of all non-missing price observations of the meat categories in Table 4.55 (excluding p_{other}) is only 306 non-missing price observations. However, if in addition we drop seafood from this total intersection of all non-missing price observations, the number of non-missing price observations increased to 1,008.

The number of missing observations reported in Table 4.55 can be reduced even further by redefining the meat categories and then excluding non-relevant meat categories. Table 4.56 reports the meat categories and meat cuts used in this study. Table 4.57 reports the number of missing and nonmissing observation per stratum when this new meat categories are used. Comparing Table 4.55 and Table 4.57, notice that p_{beef} and $p_{seafood}$ missing, non-missing and total observations, and their means remained the same because the beef and seafood category were not modified. However, meat cuts $A048$, $A049$, $A052$, and $A054$ were moved from the processed meat category to the pork category; and meat cut $A062$ was moved from the processed meat category to the chicken category. Consequently, p_{pork}, $p_{process}$, and $p_{chicken}$ changed. Now, if we consider the total intersection of all non-missing price observations of beef, pork and chicken, there are 3,707 non-missing observations. Additionally, in this new dataset, the means of p_{beef}, p_{pork} and $p_{chicken}$ are 47.9163, 44.4189, and 27.5099 pesos/kg respectively. Table 4.58 reports the descriptive statistics of this new dataset obtained by computing the total intersection of all non-missing price observations of beef, pork and chicken reported in Table 4.57. It is worthwhile mentioning that meat expenditure (m) in Table 4.58 is meat expenditure on all meats (beef, pork, processed meat, chicken, other meat, and seafood) rather than meat expenditure on only the three meats reported in the table (beef, pork and chicken).

In order to keep it simple, this study will work with this latter dataset, which resulted from the total intersection of the non-missing price observations of beef, pork and chicken. Working with this dataset has the advantage that it avoids having to do price imputation. An alternative procedure is to impute prices in either Table 4.55 or 4.57. The author preferred the former procedure, given the large amount of observations that would have to be imputed if price imputation were selected.

Since the dataset corresponding to Table 4.58 has 3,707 observations, we could further analyze the information contained in this dataset by subsetting it by region (i.e., Northeast, Northwest, Central-West, Central, and Southeast regions)[7] and urbanization level (i.e., urban or rural).[8] Figure 4.1 provides a map of the Mexican states and the Federal District. Figure 4.2 shows the Mexican geographical regions used in this study. The Northeast region of Mexico consists of the states of Chihuahua, Cohahuila de Zaragoza, Durango, Nuevo León, and Tamaulipas. The Northwest region of Mexico consists of the states of Baja California, Sonora, Baja California Sur, and Sinaloa. The Central-West region of Mexico consists of the states of Zacatecas, Mayarit, Aguascalientes, San Luis Potosí, Jalisco, Guanajuato, Querétaro Arteaga, Colima, and Michoacán de Ocampo. The Central region of Mexico consists of the states of Hidalgo, Estado de México, Tlaxcala, Morelos, Puebla, and Distrito Federal. Finally, the Southeast region of Mexico consists of the states of Veracruz de Ignacio de la Llave, Yucatán, Quintana Roo, Campeche, Tabasco, Guerrero, Oxaca, and Chiapas. Table 4.59 through Table 4.68 provides the descriptive statistics of Table 4.58 when the analysis is performed by region and urbanization level.

In summary, the following outline was applied in this section. First, the variables from the seven ENIGH datasets that are used in the study were explained (Table 4.2). Second, details about how new variables are created or transformed from the variables provided were given. In particular, the variable adult equivalents was created, nom-

[7]This study used the same five-region definitions provided by SIACON-SIAP-SAGARPA (2006), which used ENIGH 2000, 2002 and 2004 databases. Other studies or sources such as Barrera et al. (2008), Zepeda (2007), Arroyo (2002), and Wikipedia (2008) consider eight regions.

[8]Once again, following SIACON-SIAP-SAGARPA (2006), this study also considers stratum 1 and 2 as the urban sector, and stratum 3 and 4 as the rural sector. See Table 4.2 for the definitions of stratum 1, 2, 3 and 4.

inal prices and expenditures were transformed to real variables (equation 4.1), meat consumption was transformed to per capita consumption, and real expenditure was transformed to per capita real expenditure. Third, all meat cut datasets were merged into one dataset (Table 4.54). Fourth, the number of missing price observations was reduced by aggregating meat cuts into meat categories (Table 4.55). In particular, we created new variables for prices, quantities, and total meat expenditure. Fifth, the number of missing observations was reduced even further by redefining the meat categories (Table 4.56) and excluding non-relevant meat categories (processed meat and seafood in Table 4.57). Sixth, the total intersection of all non-missing price observations of the beef, pork and chicken datasets from the previous step was considered. Consequently, in this manner price imputation was avoided. Seventh, the resulting dataset is the dataset used in this study (Table 4.58). This dataset can be analyzed by subsetting it by region (Figure 4.1) and urbanization level (urban = stratum 1 and 2, rural = stratum 3 and 4). In addition, in each step we reported the difficulties that emerged as the data was being organized.

Table 4.2: List of Variables Used in this Study from the seven ENIGH 2006 Datasets.

Dataset	**Variable Used**	**Variable Description**
Concentrated (concentrado.dbf)	*hog*	This is the sampling weight variable. That is, the number of households that the interviewed household represents nationally.
Households (hogares.dbf)	*estrato*	This is the stratum variable. This variable equals "1" if household location is within a population of 100,000 people or more, "2" if household location is within a population between 15,000 and 99,999 people, "3" if household location is within a population between 2,500 people and 14,999 people, and "4" if household location is within a population of less than 2,500 people.
Members (poblacion.dbf)	*folio*	This variable is the household id number. It is a categorical variable of 11 digits that identifies the households. From left to right digits 1 to 4 read the year, digits 5 and 6 read the code for the Mexican state, digit 7 reads the code of the time period in which households were interviewed, digits 8 to 10 read the consecutive order of household interviews. Finally, digit 11 codifies a character variable (type of household) taking values from 0 to 9.
	edad	This variable is the age of each household member in years.
Expenditures (gastos.dbf)	*folio*	This variable is the household id number.
	clave	This variable takes the values of A025, A026, ..., A074 which are codes for the different cuts or group of cuts of meat. Refer to Table 4.3.
	precio	This variable is the nominal price of "clave" in Mexican pesos per kilogram (nominal pesos/kg)
	cantidad	This variable is the quantity consumed of "clave" in kilograms per household (kg).
	gasto	This variable is the nominal expenditure on "clave" in Mexican pesos per household (nominal pesos).

Source: ENIGH 2006, summarized by author.

Table 4.3: Meat Cuts Reported by ENIGH 2006.

Code	**Description**
Beef, Pork Chicken and Other Meats	
(a) Beef and Veal	
A025	Beefsteak: boneless rump, bottom round, top round, etc.
A026	Brisket and fillet steak
A027	Milanesa
A028	Tore shank
A029	Rib cutlet
A030	Chuck, strips for grilling and sirloin steak
A031	Meat for stewing/boiling or meat cut with bone
A032	Special cuts: t-bone, roast beef, etc.
A033	Hamburger patty
A034	Ground beef
A035	Chopped loin, chopped top and bottom round
A036	Other beef cuts: head, udder, etc.
A037	Guts/innards/viscera: heart, liver, marrow, rumen/belly, etc.
(b) Pork	
A038	Pork steak
A039	(Chopped) leg
A040	Chopped loin
A041	Ground pork
A042	Ribs and cutlet
A043	Shoulder blade
A044	Elbow
A045	Other pork cuts: head, ridge/backbone, belly, breast, etc.

continued on next page

Table 4.3: *continued*

Code	Description
A046	Guts/innards/viscera: heart, liver, kidney, etc.
	(c) Processed Beef and Pork
A047	Shredded meat
A048	Pork skin/chicharron
A049	Pork sausage
A050	Smoked cutlet
A051	Crusher and dried meats
A052	Ham
A053	Bologna, embedded pork and salami
A054	Bacon
A055	Sausages
A056	Other processed meats from beef and pork: stuffing, smoked meat/dried meat, etc.
	(d) Chicken
A057	Leg, thigh and breast with bone
A058	Boneless leg, boneless thigh and boneless breast
A059	Whole chicken or in parts (except legs, thigh and breast)
A060	Guts/innards/viscera and other chicken parts: wings, head, neck, gizzard, liver, etc.
A061	Other poultry meat: hen/fowl, turkey, duck, etc.
	(e) Processed Poultry Meat
A062	Chicken sausage, ham & nuggets, bologna, etc.
	(f) Other Meats
A063	Lamb: sheep and ram

continued on next page

Table 4.3: *continued*

Code	Description
A064	Goat and goatling
A065	Other meats: horses, iguana, rabbit, frog, deer, etc.
	Seafood
	(g) Fresh Fish
A066	Whole fish, clean and not clean (catfish, carp, tilapia, etc.)
A067	Fish fillet
	(h) Processed Fish
A068	Tuna
A069	Salmon and codfish
A070	Smoked fish, dried fish, fish nuggets and sardines
	(i) Other Fish
A071	Young eel, manta ray, eel, fish/crustaceous eggs, etc.
	(j) Shellfish
A072	Fresh shrimp
A073	Other fresh shellfish: clam, crab, oyster, octopus
	(k) Processed Shellfish
A074	Processed: smoked, packaged, breaded, dried shrimp

Source: ENIGH 2006—Clasificación de Variables, translated into English by author.

Table 4.4: Descriptive Statistics of Meat Cut A025 (Beefsteak).

Number of Strata	4
Number of Observations	7395
Sum of Weights	10163161

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p025	0	2.892118	790.512177	51.363693	0.203991
q025	0	0.014388	5.056180	0.215600	0.002818
m025	0	0.745151	188.380897	10.737634	0.138807

Source: ENIGH 2006, computed by author.

Table 4.5: Descriptive Statistics of Meat Cut A026 (Beef Brisket and Fillet Steak).

Number of Strata	4
Number of Observations	178
Sum of Weights	219141

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p026	0	20.789716	251.488498	60.630383	2.321882
q026	0	0.040984	3.636364	0.277609	0.022701
m026	0	2.832681	304.834542	17.068915	1.758529

Source: ENIGH 2006, computed by author.

Table 4.6: Descriptive Statistics of Meat Cut A027 (Milanesa).

Number of Strata	4
Number of Observations	545
Sum of Weights	877336

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p027	0	10.059540	125.744249	51.627196	0.657835
q027	0	0.028571	1.685393	0.218486	0.008534
m027	0	1.437077	81.945690	11.087110	0.458728

Source: ENIGH 2006, computed by author.

Table 4.7: Descriptive Statistics of Meat Cut A028 (Beef Tore Shank).

Number of Strata	4
Number of Observations	51
Sum of Weights	58802

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p028	0	14.670162	186.939783	40.200860	3.776058
q028	0	0.030211	1.752809	0.371525	0.074389
m028	0	3.717494	64.049505	12.665383	2.471559

Source: ENIGH 2006, computed by author.

Table 4.8: Descriptive Statistics of Meat Cut A029 (Beef Rib Cutlet).

Number of Strata	4
Number of Observations	442
Sum of Weights	547703

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p029	0	5.591428	652.009079	47.302999	2.079475
q029	0	0.016949	4.454343	0.285171	0.040725
m029	0	1.169714	74.681068	11.164584	0.714768

Source: ENIGH 2006, computed by author.

Table 4.9: Descriptive Statistics of Meat Cut A030 (Beef Chuck, Strips for Grilling and Sirloin Steak).

Number of Strata	4
Number of Observations	293
Sum of Weights	313297

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p030	0	11.174472	150.893099	49.577944	1.046987
q030	0	0.032808	3.370787	0.327055	0.024669
m030	0	1.100125	176.651617	15.626541	1.182929

Source: ENIGH 2006, computed by author.

Table 4.10: Descriptive Statistics of Meat Cut A031 (Beef Meat for Stewing/Boiling or Meat Cut with Bone).

Number of Strata	4
Number of Observations	1418
Sum of Weights	1946528

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p031	0	0.419147	140.833559	37.408912	0.429271
q031	0	0.033025	8.620690	0.282494	0.007870
m031	0	1.053134	84.771404	9.764357	0.240351

Source: ENIGH 2006, computed by author.

Table 4.11: Descriptive Statistics of Meat Cut A032 (Special Beef Cuts).

Number of Strata	4
Number of Observations	41
Sum of Weights	85609

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p032	0	25.148850	314.360622	60.941159	3.586207
q032	0	0.042857	1.363636	0.329285	0.069094
m032	0	2.785535	102.767345	19.237933	3.706066

Source: ENIGH 2006, computed by author.

Table 4.12: Descriptive Statistics of Meat Cut A033 (Beef Hamburger Patty).

Number of Strata	4
Number of Observations	64
Sum of Weights	104568

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p033	0	8.382950	139.718626	42.582302	2.928313
q033	0	0.043436	1.123596	0.313114	0.034243
m033	0	1.212129	28.257135	11.162960	1.105516

Source: ENIGH 2006, computed by author.

Table 4.13: Descriptive Statistics of Meat Cut A034 (Ground Beef).

Number of Strata	4
Number of Observations	3209
Sum of Weights	4153054

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p034	0	0.838295	127.420839	46.754757	0.324936
q034	0	0.020509	51.923077	0.188570	0.004623
m034	0	0.503985	65.933314	8.233624	0.134517

Source: ENIGH 2006, computed by author.

Table 4.14: Descriptive Statistics of Meat Cut A035 (Beef Chopped Loin, Chopped Top & Bottom Round).

Number of Strata	4
Number of Observations	1097
Sum of Weights	1281198

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p035	0	2.514885	251.488498	50.038120	0.850352
q035	0	0.019763	3.370787	0.251376	0.010136
m035	0	0.703268	146.937100	11.873944	0.492596

Source: ENIGH 2006, computed by author.

Table 4.15: Descriptive Statistics of Meat Cut A036 (Other Beef Cuts).

Number of Strata	4
Number of Observations	262
Sum of Weights	352053

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p036	0	4.116028	128.971684	46.495205	1.857550
q036	0	0.043178	3.370787	0.224358	0.012651
m036	0	1.679389	47.095224	9.263533	0.519137

Source: ENIGH 2006, computed by author.

Table 4.16: Descriptive Statistics of Meat Cut A037 (Beef Guts/Innards/Viscera).

Number of Strata	4
Number of Observations	545
Sum of Weights	784509

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p037	0	2.791522	83.829499	23.887400	0.673344
q037	0	0.023502	9.049774	0.324463	0.019627
m037	0	0.742160	113.795700	7.086253	0.401089

Source: ENIGH 2006, computed by author.

Table 4.17: Descriptive Statistics of Meat Cut A038 (Pork Steak).

Number of Strata	4
Number of Observations	977
Sum of Weights	1174488

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p038	0	3.353180	104.661130	41.945047	0.484168
q038	0	0.028835	3.205128	0.204533	0.008302
m038	0	1.318318	56.514269	8.165345	0.325789

Source: ENIGH 2006, computed by author.

Table 4.18: Descriptive Statistics of Meat Cut A039 (Pork (Chopped) Leg).

Number of Strata	4
Number of Observations	742
Sum of Weights	836817

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p039	0	7.335081	117.361299	37.368560	0.559719
q039	0	0.029762	11.428571	0.238185	0.009360
m039	0	1.305755	153.288227	8.350581	0.266908

Source: ENIGH 2006, computed by author.

Table 4.19: Descriptive Statistics of Meat Cut A040 (Pork Chopped Loin).

Number of Strata	4
Number of Observations	867
Sum of Weights	897714

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p040	0	10.478687	419.147496	41.887856	0.562904
q040	0	0.008475	2.272727	0.248337	0.011504
m040	0	0.710419	114.312953	10.197952	0.565427

Source: ENIGH 2006, computed by author.

Table 4.20: Descriptive Statistics of Meat Cut A041 (Ground Pork).

Number of Strata	4
Number of Observations	382
Sum of Weights	499110

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p041	0	2.934032	100.595399	40.865964	0.806886
q041	0	0.030581	1.910828	0.167421	0.008143
m041	0	0.859790	37.676179	6.375383	0.264291

Source: ENIGH 2006, computed by author.

Table 4.21: Descriptive Statistics of Meat Cut A042 (Pork Ribs and Cutlet).

Number of Strata	4
Number of Observations	1519
Sum of Weights	2092730

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p042	0	0.335318	335.317997	40.365124	0.430226
q042	0	0.032468	26.785714	0.259162	0.033903
m042	0	0.317536	111.772666	8.750041	0.311599

Source: ENIGH 2006, computed by author.

Table 4.22: Descriptive Statistics of Meat Cut A043 (Pork Shoulder Blade).

Number of Strata	4
Number of Observations	31
Sum of Weights	36985

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p043	0	20.957375	3912.046089	102.729298	69.429087
q043	0	0.007160	0.454545	0.246436	0.025322
m043	0	3.488354	28.009857	9.443449	0.842189

Source: ENIGH 2006, computed by author.

Table 4.23: Descriptive Statistics of Meat Cut A044 (Pork Elbow).

Number of Strata	4
Number of Observations	68
Sum of Weights	92050

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p044	0	6.622530	117.361299	31.875297	2.263342
q044	0	0.047170	1.685393	0.245184	0.031958
m044	0	1.724493	32.966657	6.687653	0.682873

Source: ENIGH 2006, computed by author.

Table 4.24: Descriptive Statistics of Meat Cut A045 (Other Pork Cuts).

Number of Strata	4
Number of Observations	487
Sum of Weights	702646

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p045	0	5.029770	177.718538	30.949090	0.914002
q045	0	0.019493	5.454545	0.280188	0.011082
m045	0	1.116867	76.208636	7.806084	0.313442

Source: ENIGH 2006, computed by author.

Table 4.25: Descriptive Statistics of Meat Cut A046 (Pork Guts/Innards/Viscera).

Number of Strata	4
Number of Observations	25
Sum of Weights	22212

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p046	0	11.174472	41.914750	22.277458	2.272939
q046	0	0.038071	1.123596	0.238980	0.055009
m046	0	1.063826	32.966657	5.229487	1.660989

Source: ENIGH 2006, computed by author.

Table 4.26: Descriptive Statistics of Meat Cut A047 (Shredded Meat).

Number of Strata	4
Number of Observations	201
Sum of Weights	241850

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p047	0	5.172280	201.190798	56.161330	2.278915
q047	0	0.023170	1.388889	0.157676	0.013210
m047	0	1.311544	37.912839	7.688277	0.501253

Source: ENIGH 2006, computed by author.

Table 4.27: Descriptive Statistics of Meat Cut A048 (Pork Skin/Chicharron).

Number of Strata	4
Number of Observations	1927
Sum of Weights	2749824

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p048	0	0.838295	402.381596	58.791968	0.695232
q048	0	0.004735	1.818182	0.104012	0.003119
m048	0	0.284168	70.642836	5.356848	0.153523

Source: ENIGH 2006, computed by author.

Table 4.28: Descriptive Statistics of Meat Cut A049 (Pork Sausage).

Number of Strata	4
Number of Observations	3559
Sum of Weights	4680713

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p049	0	0.108978	898.174403	42.566750	0.716380
q049	0	0.000902	11.782477	0.112814	0.003326
m049	0	0.241637	75.352359	4.076692	0.079598

Source: ENIGH 2006, computed by author.

Table 4.29: Descriptive Statistics of Meat Cut A050 (Smoked Cutlet).

Number of Strata	4
Number of Observations	389
Sum of Weights	550048

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p050	0	4.191475	382.430175	45.445364	1.387205
q050	0	0.018685	1.984127	0.216055	0.020258
m050	0	0.751834	56.318182	8.393312	0.549958

Source: ENIGH 2006, computed by author.

Table 4.30: Descriptive Statistics of Meat Cut A051 (Crushed and Dried Meats).

Number of Strata	4
Number of Observations	178
Sum of Weights	210450

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p051	0	16.765900	1341.271987	133.167069	11.330517
q051	0	0.003077	2.727273	0.131095	0.045281
m051	0	1.053513	457.251814	16.914164	7.596779

Source: ENIGH 2006, computed by author.

Table 4.31: Descriptive Statistics of Meat Cut A052 (Ham).

Number of Strata	4
Number of Observations	4549
Sum of Weights	6325606

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p052	0	2.791522	871.826791	42.140284	0.353458
q052	0	0.001859	1.685393	0.104221	0.002557
m052	0	0.109295	56.514269	4.193777	0.098111

Source: ENIGH 2006, computed by author.

Table 4.32: Descriptive Statistics of Meat Cut A053 (Bologna, Embedded Pork and Salami).

Number of Strata	4
Number of Observations	422
Sum of Weights	598727

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p053	0	5.029770	394.836941	32.949710	1.335646
q053	0	0.011765	0.909091	0.117336	0.005908
m053	0	0.408261	45.725181	3.312677	0.171724

Source: ENIGH 2006, computed by author.

Table 4.33: Descriptive Statistics of Meat Cut A054 (Bacon).

Number of Strata	4
Number of Observations	380
Sum of Weights	536795

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p054	0	8.382950	131.612314	52.476894	1.534118
q054	0	0.006748	0.727273	0.088734	0.007521
m054	0	0.339391	33.150756	4.171160	0.310386

Source: ENIGH 2006, computed by author.

Table 4.34: Descriptive Statistics of Meat Cut A055 (Sausages).

Number of Strata	4
Number of Observations	2670
Sum of Weights	3624571

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p055	0	0.838295	251.488498	26.232814	0.416556
q055	0	0.006042	3.333333	0.161268	0.004446
m055	0	0.117244	48.979033	3.705498	0.103433

Source: ENIGH 2006, computed by author.

Table 4.35: Descriptive Statistics of Meat Cut A056 (Other Processed Meats).

Number of Strata	4
Number of Observations	671
Sum of Weights	822094

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p056	0	3.462158	153.684621	52.256246	1.401715
q056	0	0.006519	1.376147	0.166977	0.006797
m056	0	0.218591	103.825527	7.871881	0.357426

Source: ENIGH 2006, computed by author.

Table 4.36: Descriptive Statistics of Meat Cut A057 (Chicken Leg, Thigh and Breast with Bone).

Number of Strata	4
Number of Observations	5214
Sum of Weights	8166414

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p057	1	1.399953	217.956698	27.728852	0.199535
q057	1	0.018416	22.471910	0.307679	0.004576
m057	0	0.273357	84.250753	7.716511	0.099213

Source: ENIGH 2006, computed by author.

Table 4.37: Descriptive Statistics of Meat Cut A058 (Chicken Boneless Leg, Thigh and Breast).

Number of Strata	4
Number of Observations	1534
Sum of Weights	2519080

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p058	0	1.131698	385.615696	35.323948	0.468309
q058	0	0.013587	5.909091	0.281889	0.006010
m058	0	1.076709	117.738061	9.441513	0.228653

Source: ENIGH 2006, computed by author.

Table 4.38: Descriptive Statistics of Meat Cut A059 (Whole Chicken).

Number of Strata	4
Number of Observations	7497
Sum of Weights	8917172

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p059	0	0.838295	301.786197	23.840577	0.203954
q059	0	0.017301	11.235955	0.347922	0.004773
m059	0	0.231939	121.687983	7.587128	0.099681

Source: ENIGH 2006, computed by author.

Table 4.39: Descriptive Statistics of Meat Cut A060 (Chicken Guts/Innards/Viscera).

Number of Strata	4
Number of Observations	920
Sum of Weights	1491449

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p060	0	0.838295	356.275372	17.878599	0.653485
q060	0	0.004695	5.617978	0.372337	0.038521
m060	0	0.210099	33.908561	4.370160	0.179723

Source: ENIGH 2006, computed by author.

Table 4.40: Descriptive Statistics of Meat Cut A061 (Other Poultry Meat).

Number of Strata	4
Number of Observations	138
Sum of Weights	167852

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p061	0	4.191475	167.658998	37.160563	3.667566
q061	0	0.030035	4.494382	0.642142	0.064259
m061	0	1.194375	131.866628	15.650652	1.397909

Source: ENIGH 2006, computed by author.

Table 4.41: Descriptive Statistics of Meat Cut A062 (Chicken Sausage, Ham & Nuggets, Bologna, etc.).

Number of Strata	4
Number of Observations	3190
Sum of Weights	4002486

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p062	0	0.368850	352.083897	39.072679	0.427191
q062	0	0.003964	8.532423	0.161968	0.004351
m062	0	0.257146	132.431771	5.669607	0.149021

Source: ENIGH 2006, computed by author.

Table 4.42: Descriptive Statistics of Meat Cut A063 (Lamb).

Number of Strata	4
Number of Observations	12
Sum of Weights	15380

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p063	0	31.016915	100.595399	52.515396	1.489917
q063	0	0.057870	2.352941	0.754580	0.094510
m063	0	2.636148	94.190449	34.988288	4.109889

Source: ENIGH 2006, computed by author.

Table 4.43: Descriptive Statistics of Meat Cut A064 (Goat and Goatling).

Number of Strata	4
Number of Observations	12
Sum of Weights	12056

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p064	0	16.765900	117.361299	45.751943	11.246870
q064	0	0.083963	0.807754	0.403425	0.102407
m064	0	5.425858	28.418200	13.168530	1.560181

Source: ENIGH 2006, computed by author.

Table 4.44: Descriptive Statistics of Meat Cut A065 (Other Meats).

Number of Strata	4
Number of Observations	17
Sum of Weights	22697

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p065	0	13.974378	75.446549	55.240181	6.187467
q065	0	0.031726	1.123596	0.365664	0.083044
m065	0	2.127652	52.393437	21.541523	7.270937

Source: ENIGH 2006, computed by author.

Table 4.45: Descriptive Statistics of Meat Cut A066 (Whole Fish).

Number of Strata	4
Number of Observations	1447
Sum of Weights	2100444

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p066	0	2.095737	251.488498	31.710482	0.711140
q066	0	0.023866	4.494382	0.342140	0.010288
m066	0	0.781263	176.607091	9.879831	0.477757

Source: ENIGH 2006, computed by author.

Table 4.46: Descriptive Statistics of Meat Cut A067 (Fish Fillet).

Number of Strata	4
Number of Observations	736
Sum of Weights	1066212

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p067	0	3.353180	733.508118	59.725915	1.383307
q067	0	0.013441	2.247191	0.273767	0.011414
m067	0	0.901392	122.447583	15.043534	0.703100

Source: ENIGH 2006, computed by author.

Table 4.47: Descriptive Statistics of Meat Cut A068 (Tuna).

Number of Strata	4
Number of Observations	2104
Sum of Weights	2807024

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p068	5	0.058681	528.125845	39.957431	0.482137
q068	5	0.011225	52.959502	0.137917	0.004624
m068	0	0.290067	67.817123	4.945517	0.154679

Source: ENIGH 2006, computed by author.

Table 4.48: Descriptive Statistics of Meat Cut A069 (Salmon and Codfish).

Number of Strata	4
Number of Observations	20
Sum of Weights	21157

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p069	0	16.514411	188.616373	75.987109	11.228395
q069	0	0.010627	1.123596	0.234317	0.066629
m069	0	0.534513	169.542807	20.614232	5.967604

Source: ENIGH 2006, computed by author.

Table 4.49: Descriptive Statistics of Meat Cut A070 (Smoked Fish, Dried Fish, Fish Nuggets and Sardines).

Number of Strata	4
Number of Observations	516
Sum of Weights	626056

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p070	7	3.353180	356.275372	35.435446	1.783908
q070	7	0.004545	1.471910	0.141836	0.009455
m070	0	0.280366	114.312953	4.151982	0.306336

Source: ENIGH 2006, computed by author.

Table 4.50: Descriptive Statistics of Meat Cut A071 (Other Fish).

Number of Strata	3
Number of Observations	11
Sum of Weights	12971

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p071	0	25.148850	117.361299	53.251194	11.726920
q071	0	0.135747	1.235955	0.411442	0.104901
m071	0	3.929508	53.346045	21.220646	6.232627

Source: ENIGH 2006, computed by author.

Table 4.51: Descriptive Statistics of Meat Cut A072 (Fresh Shrimp).

Number of Strata	4
Number of Observations	590
Sum of Weights	745439

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p072	0	4.191475	251.488498	66.756215	1.837216
q072	0	0.007541	3.370787	0.265041	0.016811
m072	0	0.632198	121.933817	15.399127	0.749834

Source: ENIGH 2006, computed by author.

Table 4.52: Descriptive Statistics of Meat Cut A073 (Other Fresh Shellfish).

Number of Strata	4
Number of Observations	91
Sum of Weights	108832

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p073	0	1.131698	393.353159	40.204032	4.908225
q073	0	0.028235	4.938272	0.304055	0.044016
m073	0	1.123971	62.165696	8.049098	0.958040

Source: ENIGH 2006, computed by author.

Table 4.53: Descriptive Statistics of Meat Cut A074 (Processed Shellfish).

Number of Strata	4
Number of Observations	149
Sum of Weights	206475

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p074	0	12.574425	279.428870	80.977790	5.185289
q074	0	0.009251	0.909091	0.088319	0.007782
m074	0	0.370381	76.208636	5.831758	0.529116

Source: ENIGH 2006, computed by author.

Table 4.54: Descriptive Statistics of all Meat Cuts in one Dataset

Number of Strata	4
Number of Observations	16909
Sum of Weights	22106253

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p025	10936	2.892118	790.512177	51.451538	0.224300
q025	10936	0.014388	5.056180	0.265041	0.004033
m025	10936	0.745151	188.362059	13.256654	0.202384
p026	16738	20.789716	251.488498	60.816440	2.400531
q026	16738	0.040984	3.636364	0.291295	0.024390
m026	16738	2.832681	304.834542	17.932886	1.858759
p027	16396	10.059540	100.595399	51.506331	0.671051
q027	16396	0.028571	1.685393	0.234351	0.009527
m027	16396	1.437077	81.945690	11.941097	0.519943
p028	16860	14.670162	186.939783	40.097447	3.834719
q028	16860	0.030211	1.752809	0.378031	0.075376
m028	16860	3.717309	64.049882	12.896655	2.508630
p029	16488	5.591428	652.009079	47.412777	2.177495
q029	16488	0.016949	4.454343	0.299046	0.042560
m029	16488	1.169714	74.681068	11.694980	0.743196
p030	16634	11.174472	150.893099	49.922814	1.092015
q030	16634	0.032808	3.370787	0.351865	0.027616
m030	16634	1.100125	176.651617	16.799607	1.325401
p031	15548	3.478924	140.833559	37.483699	0.438344
q031	15548	0.033025	13.505747	0.291544	0.008762
m031	15548	1.281470	84.771404	10.090048	0.262159
p032	16869	25.148850	314.360622	61.282144	3.679203
q032	16869	0.042857	1.363636	0.337307	0.070861
m032	16869	2.785415	102.767345	19.706563	3.795864
p033	16848	9.640392	139.718626	43.532503	3.035658
q033	16848	0.043436	1.935484	0.336642	0.049957
m033	16848	1.212170	40.019108	12.001753	1.201876
p034	14011	0.838295	127.420839	46.723015	0.340698
q034	14011	0.020509	51.923077	0.203759	0.005099
m034	14011	0.503985	96.777578	8.915496	0.153208
p035	15919	2.514885	251.488498	50.049617	0.908143
q035	15919	0.019763	3.370787	0.272338	0.011790
m035	15919	0.703268	146.937100	12.952992	0.600089
p036	16668	4.116028	128.971684	45.941418	1.955225
q036	16668	0.043178	3.370787	0.244480	0.016317
m036	16668	1.679389	50.297700	10.120805	0.719266
p037	16394	4.191475	83.829499	24.165265	0.710731

continued on next page

Table 4.54: *continued*

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
q037	16394	0.023502	9.049774	0.345978	0.021366
m037	16394	0.742256	113.795700	7.564246	0.431791
p038	16017	3.353180	104.661130	42.192289	0.506573
q038	16017	0.028835	3.205128	0.223072	0.009455
m038	16017	1.450337	56.508618	8.921678	0.368936
p039	16247	7.335081	117.361299	37.554367	0.607190
q039	16247	0.033113	11.428571	0.264794	0.010969
m039	16247	1.305755	153.288227	9.298677	0.319133
p040	16112	10.478687	419.147496	42.031636	0.584433
q040	16112	0.016129	2.272727	0.261178	0.012641
m040	16112	1.566321	114.312953	10.737146	0.623890
p041	16543	2.934032	100.595399	40.773894	0.812119
q041	16543	0.030581	1.910828	0.175451	0.008958
m041	16543	0.859790	45.211415	6.681044	0.302124
p042	15486	0.335318	335.317997	40.456257	0.438093
q042	15486	0.032468	26.785714	0.274093	0.035867
m042	15486	0.317536	111.772666	9.284232	0.335517
p043	16880	20.957375	3912.046089	106.018114	72.463914
q043	16880	0.007160	0.590551	0.256348	0.026089
m043	16880	3.488354	28.009877	9.823237	0.844241
p044	16842	6.622530	117.361299	31.981647	2.280368
q044	16842	0.047170	1.685393	0.247573	0.032339
m044	16842	1.724186	32.961947	6.752669	0.688960
p045	16436	5.029770	177.718538	30.948206	0.928518
q045	16436	0.019493	5.454545	0.285814	0.011503
m045	16436	1.116764	76.223877	7.964579	0.326662
p046	16884	11.174472	41.914750	22.277458	2.272939
q046	16884	0.038071	1.123596	0.238980	0.055009
m046	16884	1.063720	32.966657	5.229489	1.660987
p047	16714	5.172280	201.190798	56.447267	2.325552
q047	16714	0.023170	1.388889	0.161947	0.013656
m047	16714	1.311544	37.912839	7.888930	0.515825
p048	15096	0.838295	402.381596	58.962812	0.711739
q048	15096	0.005252	5.454545	0.111308	0.003550
m048	15096	0.284168	70.642836	5.766887	0.169085
p049	13734	0.108978	898.174403	42.574408	0.760507
q049	13734	0.001357	11.782477	0.126517	0.003822
m049	13734	0.241637	170.775845	4.625379	0.098248
p050	16529	4.191475	382.430175	45.565037	1.408615
q050	16529	0.018685	1.984127	0.219913	0.020662
m050	16529	0.751834	56.318182	8.541251	0.560397
p051	16749	16.765900	1341.271987	136.437581	12.649351
q051	16749	0.003077	2.727273	0.148293	0.051024
m051	16749	1.053513	457.251814	19.132329	8.563130

continued on next page

Table 4.54: *continued*

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
p052	13202	2.791522	871.826791	42.544479	0.389370
q052	13202	0.003326	1.685393	0.127683	0.003153
m052	13202	0.109295	66.633709	5.170259	0.123763
p053	16515	5.029770	394.836941	33.156370	1.426445
q053	16515	0.011765	0.909091	0.126304	0.006473
m053	16515	0.408261	45.725181	3.570693	0.188994
p054	16542	8.382950	131.612314	52.608900	1.546463
q054	16542	0.006748	0.727273	0.090407	0.007656
m054	16542	0.339391	33.150756	4.251843	0.315814
p055	14525	0.838295	251.488498	26.211777	0.446549
q055	14525	0.006042	3.333333	0.178694	0.004971
m055	14525	0.117244	48.979033	4.129285	0.117187
p056	16284	3.462158	153.684621	51.997221	1.456535
q056	16284	0.006519	1.376147	0.174716	0.007205
m056	16284	0.218591	103.825527	8.245754	0.383986
p057	12907	1.492165	217.956698	27.770706	0.230699
q057	12906	0	22.471910	0.405444	0.007246
m057	12907	0.441872	141.100522	10.278208	0.181287
p058	15529	1.131698	385.615696	35.444036	0.494315
q058	15529	0.013587	5.909091	0.314146	0.007675
m058	15529	1.076540	117.738061	10.522074	0.279984
p059	11193	0.838295	301.786197	23.973716	0.241093
q059	11193	0.018625	11.423221	0.447971	0.007264
m059	11193	0.240127	166.000462	9.850781	0.153683
p060	16149	0.838295	356.275372	18.804748	0.750220
q060	16149	0.004695	6.532663	0.471931	0.056344
m060	16149	0.363423	80.019067	5.758415	0.374949
p061	16780	4.191475	167.658998	38.431811	3.840393
q061	16780	0.030035	5.617978	0.678946	0.094135
m061	16780	1.194085	131.866628	16.621514	1.606761
p062	14316	0.368850	243.105548	39.184385	0.467858
q062	14316	0.005298	8.532423	0.196946	0.005644
m062	14316	0.309334	132.441190	6.990777	0.197870
p063	16898	31.016915	100.595399	53.747958	1.688555
q063	16898	0.057870	2.352941	0.842317	0.104632
m063	16898	2.636148	94.195158	39.056752	4.287618
p064	16897	16.765900	117.361299	45.751943	11.246870
q064	16897	0.083963	0.807754	0.403425	0.102407
m064	16897	5.425858	28.418200	13.168584	1.560188
p065	16892	13.974378	75.446549	55.240181	6.187467
q065	16892	0.031726	1.123596	0.365664	0.083044
m065	16892	2.127652	52.393437	21.541737	7.270924
p066	15648	2.095737	251.488498	32.772028	0.756936
q066	15648	0.023866	4.494382	0.405146	0.015770

continued on next page

Table 4.54: *continued*

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
m066	15648	1.408110	176.607091	11.728725	0.583385
p067	16216	3.353180	733.508118	60.270314	1.424405
q067	16216	0.013441	2.808989	0.287611	0.012696
m067	16216	0.901392	122.447583	15.819079	0.756468
p068	14957	0.058681	528.125845	39.980417	0.506410
q068	14953	0	52.959502	0.146413	0.005929
m068	14957	0.364476	103.603277	5.276339	0.196456
p069	16890	16.514411	188.616373	75.109529	11.342438
q069	16890	0.010627	2.247191	0.237062	0.070597
m069	16890	0.534513	339.085615	20.855725	6.788215
p070	16440	3.353180	356.275372	35.675779	1.784772
q070	16433	0	2.943820	0.148417	0.012307
m070	16440	0.485499	114.312953	4.431164	0.355217
p071	16898	25.148850	117.361299	53.251194	11.726920
q071	16898	0.135747	1.235955	0.411442	0.104901
m071	16898	3.929508	53.346045	21.220618	6.232634
p072	16372	6.287212	251.488498	68.473365	1.833052
q072	16372	0.007541	3.370787	0.295148	0.019597
m072	16372	0.632198	121.933817	17.196789	0.828532
p073	16821	1.131698	393.353159	38.628169	4.534343
q073	16821	0.030211	4.938272	0.311640	0.045721
m073	16821	1.123971	87.639931	8.331223	1.234919
p074	16766	12.574425	279.428870	84.023077	5.056411
q074	16766	0.009251	0.909091	0.093814	0.010950
m074	16766	0.370381	76.208636	6.200018	0.545316

Source: ENIGH 2006, computed by author.

Table 4.55: Missing and Non-missing Observations per Stratum with Six Meat Categories (Meat Categories Not Modified).

		Number Missing	**Number Non-missing**	**Total Observations**	**Mean (pesos/kg)**
p_{beef}	str1	2,471	4,814	7,285	48.7186
	str2	1,524	2,418	3,942	46.8417
	str3	766	808	1,574	45.0591
	str4	2,373	1,735	4,108	44.9205
	Total	**7,134**	**9,775**	**16,909**	**47.4047**
p_{pork}	str1	5,382	1,603	7,285	40.2338
	str2	2,891	1,051	3,942	39.8023
	str3	1,161	413	1,574	38.3144
	str4	3,195	913	4,108	36.8476
	Total	**12,929**	**3,980**	**16,909**	**39.2397**
$p_{process}$	str1	2,378	4,907	7,285	42.8529
	str2	1,482	2,460	3,942	40.4771
	str3	756	818	1,574	41.0264
	str4	2,168	1,940	4,108	40.9420
	Total	**6,784**	**10,125**	**16,909**	**41.9739**
$p_{chicken}$	str1	2,807	4,478	7,285	27.3432
	str2	1,512	2,430	3,942	25.8124
	str3	575	999	1,574	24.3402
	str4	1,667	2,441	4,108	25.8664
	Total	**6,561**	**10,348**	**16,909**	**26.4170**
p_{other}	str1	7,274	11	7,285	51.2946
	str2	3,930	12	3,942	41.6323
	str3	1,570	4	1,574	89.6085
	str4	4,095	13	4,108	45.4641
	Total	**16,869**	**40**	**16,909**	**52.4595**
$p_{seafood}$	str1	5,351	1,934	7,285	47.6885
	str2	2,893	1,049	3,942	45.2982
	str3	1,167	407	1,574	36.0679
	str4	3,124	984	4,108	36.4365
	Total	**12,535**	**4,374**	**16,909**	**43.5943**

Source: ENIGH 2006 Database, computed by author.

Table 4.56: Meat Categories and Cuts Used in this Study.

Code	Description
(1) Beef	
A025	Beefsteak: boneless rump, bottom round, top round, etc.
A026	Brisket and fillet steak
A027	Milanesa
A028	Tore shank
A029	Rib cutlet
A030	Chuck, strips for grilling and sirloin steak
A031	Meat for stewing/boiling or meat cut with bone
A032	Special cuts: t-bone, roast beef, etc.
A033	Hamburger patty
A034	Ground beef
A035	Chopped loin, chopped top and bottom round
A036	Other beef cuts: head, udder, etc.
A037	Guts/innards/viscera: heart, liver, marrow, rumen/belly, etc.
(2) Pork	
A038	Pork steak
A039	(Chopped) leg
A040	Chopped loin
A041	Ground pork
A042	Ribs and cutlet
A043	Shoulder blade
A044	Elbow
A045	Other pork cuts: head, ridge/backbone, belly, breast, etc.
A046	Guts/innards/viscera: heart, liver, kidney, etc.

continued on next page

Table 4.56: *continued*

Code	Description
A048	Pork skin/chicharron
A049	Pork sausage
A052	Ham
A054	Bacon
(3) Processed Meat	
A047	Shredded meat
A050	Smoked cutlet
A051	Crusher and dried meats
A053	Bologna, embedded pork and salami
A055	Sausages
A056	Other processed meats from beef and pork: stuffing, smoked meat/dried meat, etc.
(4) Chicken	
A057	Leg, thigh and breast with bone
A058	Boneless leg, boneless thigh and boneless breast
A059	Whole chicken or in parts (except legs, thigh and breast)
A060	Guts/innards/viscera and other chicken parts: wings, head, neck, gizzard, liver, etc.
A061	Other poultry meat: hen/fowl, turkey, duck, etc.
A062	Chicken sausage, ham & nuggets, bologna, etc.
(5) Seafood	
A066	Whole fish, clean and not clean (catfish, carp, tilapia, etc.)
A067	Fish fillet
A068	Tuna

continued on next page

Table 4.56: *continued*

Code	**Description**
A069	Salmon and codfish
A070	Smoked fish, dried fish, fish nuggets and sardines
A071	Young eel, manta ray, eel, fish/crustaceous eggs, etc.
A072	Fresh shrimp
A073	Other fresh shellfish: clam, crab, oyster, octopus
A074	Processed: smoked, packaged, breaded, dried shrimp

Source: ENIGH 2006, modified by author.

Table 4.57: Missing and Non-missing Observations per Stratum with Five Meat Categories (Meat Categories Modified).

		Number Missing	**Number Non-missing**	**Total Observations**	**Mean (pesos/kg)**
p_{beef}	str1	2,471	4,814	7,285	48.7186
	str2	1,524	2,418	3,942	46.8417
	str3	766	808	1,574	45.0591
	str4	2,373	1,735	4,108	44.9205
	Total	**7,134**	**9,775**	**16,909**	**47.4047**
p_{pork}	str1	3,015	4,270	7,285	44.9448
	str2	1,656	2,286	3,942	41.7445
	str3	726	848	1,574	40.4010
	str4	2,213	1,895	4,108	41.7353
	Total	**7,610**	**9,299**	**16,909**	**43.3079**
$p_{process}$	str1	5,451	1,834	7,285	36.6032
	str2	3,039	903	3,942	37.9575
	str3	1,273	301	1,574	40.0850
	str4	3,410	698	4,108	36.2209
	Total	**13,173**	**3,736**	**16,909**	**37.1217**
$p_{chicken}$	str1	2,279	5,006	7,285	29.0105
	str2	1,232	2,710	3,942	27.0900
	str3	521	1,053	1,574	24.8532
	str4	1,487	2,621	4,108	26.5907
		5,519	**11,390**	**16,909**	**27.7195**
$p_{seafood}$	str1	5,351	1,934	7,285	47.6885
	str2	2,893	1,049	3,942	45.2982
	str3	1,167	407	1,574	36.0679
	str4	3,124	984	4,108	36.4365
	Total	**12,535**	**4,374**	**16,909**	**43.5943**

Source: ENIGH 2006 Database, computed by author.

Table 4.58: Descriptive Statistics of the Beef, Pork and Chicken Dataset.

Number of Strata	4
Number of Observations	3707
Sum of Weights	5303145

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
q_beef	0	0.015453	13.793103	0.333149	0.006350
p_beef	0	2.514885	251.488498	47.916307	0.299430
q_pork	0	0.004751	27.250000	0.249981	0.014406
p_pork	0	0.402382	871.826791	44.418884	0.402417
q_chicken	0	0.004695	22.471910	0.469278	0.009166
p_chicken	0	1.601143	301.786197	27.509891	0.274583
m	0	5.380127	759.582445	41.412522	0.743238

Source: ENIGH 2006, computed by author.

Figure 4.1: Map of Mexican States and the Federal District.

Note: 1 = Aguascalientes, 2 = Baja California, 3 = Baja California Sur, 4 = Campeche, 5 = Coahuila de Zaragoza, 6 = Colima, 7 = Chiapas, 8 = Chihuahua, 9 = Distrito Federal, 10 = Durango, 11 = Guanajuato, 12 = Guerrero, 13 = Hidalgo, 14 = Jalisco, 15 = Estado de México, 16 = Michoacán de Ocampo, 17 = Morelos, 18 = Mayarit, 19 = Nuevo León, 20 = Oaxaca, 21 = Puebla, 22 = Querétaro Arteaga, 23 = Quintana Roo, 24 = San Luis Potosí, 25 = Sinaloa, 26 = Sonora, 27 = Tabasco, 28 = Tamaulipas, 29 = Tlaxcala, 30 = Veracruz de Ignacio de la Llave, 31 = Yucatán, and 32 = Zacatecas.

Figure 4.2: Map of Mexican Geographical Regions.

Note: 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, Oxaca, and Chiapas.

Table 4.59: Descriptive Statistics of the Beef, Pork and Chicken Dataset for the Urban Sector in the Northeast Region.

Number of Strata	2
Number of Observations	216
Sum of Weights	241965

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
q_beef	0	0.040453	2.307692	0.434873	0.029777
p_beef	0	7.963802	125.744249	42.388387	1.301243
q_pork	0	0.019646	1.404494	0.235748	0.017951
p_pork	0	4.694452	203.588322	38.456993	1.394091
q_chicken	0	0.034325	2.178571	0.533158	0.028488
p_chicken	0	6.614147	70.617970	24.483795	0.856890
m	0	8.988155	171.359995	43.043123	1.954828

Source: ENIGH 2006, computed by author.

Table 4.60: Descriptive Statistics of the Beef, Pork and Chicken Dataset for the Rural Sector in the Northeast Region.

Number of Strata	2
Number of Observations	65
Sum of Weights	58040

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
q_beef	0	0.066489	2.247191	0.504193	0.089653
p_beef	0	8.299120	83.829499	41.992341	2.455289
q_pork	0	0.014970	2.247191	0.194747	0.042939
p_pork	0	13.412720	176.041948	39.528778	3.835294
q_chicken	0	0.045704	3.181818	0.344178	0.034316
p_chicken	0	7.963802	61.454148	25.417032	1.831553
m	0	11.800496	196.229559	36.810314	3.572397

Source: ENIGH 2006, computed by author.

Table 4.61: Descriptive Statistics of the Beef, Pork and Chicken Dataset for the Urban Sector in the Northwest Region.

Number of Strata	2
Number of Observations	291
Sum of Weights	465235

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
q_beef	0	0.026652	2.673797	0.399094	0.028154
p_beef	0	10.478687	134.646942	45.785982	1.241914
q_pork	0	0.009732	2.133758	0.197364	0.012922
p_pork	0	6.506170	335.317997	43.276487	1.855863
q_chicken	0	0.022727	3.630705	0.484834	0.035887
p_chicken	0	1.628687	89.261651	25.274177	0.952588
m	0	6.276350	148.125077	38.403223	1.588092

Source: ENIGH 2006, computed by author.

Table 4.62: Descriptive Statistics of the Beef, Pork and Chicken Dataset for the Rural Sector in the Northwest Region.

Number of Strata	2
Number of Observations	50
Sum of Weights	42828

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
q_beef	0	0.071327	2.247191	0.387049	0.084827
p_beef	0	2.514885	67.063599	43.395455	3.069425
q_pork	0	0.015198	1.060329	0.196513	0.035038
p_pork	0	16.765900	83.829499	42.847101	2.788116
q_chicken	0	0.051653	1.534722	0.473641	0.051745
p_chicken	0	4.191475	75.446549	24.012358	2.432507
m	0	10.090588	277.390871	36.050006	5.583017

Source: ENIGH 2006, computed by author.

Table 4.63: Descriptive Statistics of the Beef, Pork and Chicken Dataset for the Urban Sector in the Central-West Region.

Number of Strata	2
Number of Observations	887
Sum of Weights	834796

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
q_beef	0	0.015453	3.491828	0.320296	0.012393
p_beef	0	6.161468	149.694337	48.722820	0.554751
q_pork	0	0.011468	3.190909	0.234097	0.015674
p_pork	0	5.591428	125.744249	42.237818	0.783150
q_chicken	0	0.020877	7.000000	0.382793	0.020677
p_chicken	0	5.029770	150.893099	29.829943	0.632413
m	0	5.380127	759.582445	41.789481	3.341870

Source: ENIGH 2006, computed by author.

Table 4.64: Descriptive Statistics of the Beef, Pork and Chicken Dataset for the Rural Sector in the Central-West Region.

Number of Strata	2
Number of Observations	207
Sum of Weights	291773

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
q_beef	0	0.060120	2.500000	0.313575	0.048642
p_beef	0	12.574425	130.774019	46.134734	1.671735
q_pork	0	0.004751	1.590909	0.209230	0.017858
p_pork	0	2.514885	125.744249	42.923778	2.488597
q_chicken	0	0.045372	3.409091	0.388502	0.037971
p_chicken	0	6.982997	85.673748	26.690409	0.835188
m	0	7.478797	192.174512	35.223067	3.432218

Source: ENIGH 2006, computed by author.

Table 4.65: Descriptive Statistics of the Beef, Pork and Chicken Dataset for the Urban Sector in the Central Region.

Number of Strata	2
Number of Observations	918
Sum of Weights	1986442

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
q_beef	0	0.029308	2.380952	0.327425	0.008719
p_beef	0	10.776881	107.301759	49.803026	0.454592
q_pork	0	0.012300	2.506775	0.234028	0.007958
p_pork	0	6.982997	167.658998	46.571016	0.579932
q_chicken	0	0.004695	22.471910	0.486468	0.013799
p_chicken	0	1.601143	301.786197	28.143422	0.503197
m	0	8.483432	274.354137	43.017748	0.859889

Source: ENIGH 2006, computed by author.

Table 4.66: Descriptive Statistics of the Beef, Pork and Chicken Dataset for the Rural Sector in the Central Region.

Number of Strata	2
Number of Observations	198
Sum of Weights	415894

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
q_beef	0	0.033040	1.395349	0.264592	0.022585
p_beef	0	5.029770	69.075507	48.114006	1.332501
q_pork	0	0.030303	11.346633	0.223122	0.017448
p_pork	0	1.234412	113.869800	42.890489	1.876761
q_chicken	0	0.039157	4.015296	0.378698	0.031364
p_chicken	0	4.189080	67.063599	24.747150	0.991913
m	0	6.674627	124.822605	33.145215	1.962880

Source: ENIGH 2006, computed by author.

Table 4.67: Descriptive Statistics of the Beef, Pork and Chicken Dataset for the Urban Sector in the Southeast Region.

Number of Strata	2
Number of Observations	617
Sum of Weights	581798

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
q_beef	0	0.027322	13.793103	0.343734	0.015892
p_beef	0	5.365088	201.190798	47.315252	0.753877
q_pork	0	0.014085	6.944444	0.287443	0.020173
p_pork	0	0.402382	871.826791	46.468578	1.143165
q_chicken	0	0.019920	6.320225	0.569104	0.024261
p_chicken	0	2.011908	206.782226	27.344251	0.597392
m	0	8.541119	268.588144	46.380264	1.439442

Source: ENIGH 2006, computed by author.

Table 4.68: Descriptive Statistics of the Beef, Pork and Chicken Dataset for the Rural Sector in the Southeast Region.

Number of Strata	2
Number of Observations	258
Sum of Weights	384374

Statistics

Variable	N Miss	Minimum	Maximum	Mean	Std Error of Mean
q_beef	0	0.035613	3.636364	0.287976	0.019878
p_beef	0	14.670162	251.488498	45.918942	1.284695
q_pork	0	0.016949	27.250000	0.457167	0.185706
p_pork	0	0.988826	99.002639	43.769053	1.230471
q_chicken	0	0.022936	11.423221	0.535858	0.055118
p_chicken	0	7.980568	83.829499	28.375812	1.037268
m	0	8.709558	367.342749	42.330653	3.072649

Source: ENIGH 2006, computed by author.

4.3 Model Specification

4.3.1 Estimation of One General SUR Model

In order to provide an empirical application of a SUR model, this study considers the Mexican per capita meat consumption of beef, pork and chicken. Consequently, we would like to estimate a system of three equations, where $i =$ 1, 2, 3 = beef, pork, chicken. Each equation will contain $K_i = 10$ regression coefficients and a data sample of $T = 3,707$ observations for each equation.

The i^{th} equation is given by

$$\mathbf{q}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{u}_i, \quad i = 1, 2, 3, \tag{4.2}$$

where $\mathbf{q}_i$ is a (3707×1) vector of observations on the dependent variable of the i^{th} equation, $\mathbf{X}_i$ is a (3707×10) matrix containing a column of 1s and 9 columns of observations on independent variables, and $\boldsymbol{\beta_i}$ is a (10×1) vector of parameters, and $\mathbf{u}_i$ is a (3707×1) vector of disturbances.

Using the variables of interest for this study, equation (4.2) can be written as

$$\begin{aligned}
\underset{(3707\times1)}{\mathbf{q}_i} &= \begin{pmatrix} \mathbf{1}_{i1} & \mathbf{p}_{i2} & \mathbf{p}_{i3} & \mathbf{p}_{i4} & \mathbf{m}_{i5} & \mathbf{NE}_{i6} & \mathbf{NW}_{i7} & \mathbf{CW}_{i8} & \mathbf{C}_{i9} & \mathbf{urban}_{i10} \end{pmatrix}_{(3707\times10)} \\
&\quad \times \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{i10} \end{pmatrix}_{(10\times1)} + \underset{(3707\times1)}{\mathbf{u}_i}, \qquad i = 1, 2, 3 \\
&= \beta_{i1}\mathbf{1}_{i1} + \beta_{i2}\mathbf{p}_{i2} + \cdots + \beta_{i10}\mathbf{urban}_{i10} + \mathbf{u}_i, \qquad i = 1, 2, 3.
\end{aligned}$$

However, in this study $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_3$.[9] Therefore, the subscripts of the vectors

[9] In Section 2.1.1 it was shown that in this case SUR will reduce to single equation least-squares. However, for the purpose of providing an empirical application of a SUR model, we will carry on with the system of three equations even though single equation least squares will provide the same parameter estimates and parameter variances.

of the $\mathbf{X}_i$ matrix can be omitted. This implies,

$$\underset{(3707\times 1)}{\mathbf{q}_i} = \begin{pmatrix} \mathbf{1} & \mathbf{p}_{beef} & \mathbf{p}_{pork} & \mathbf{p}_{chicken} & \mathbf{m} & \mathbf{NE} & \mathbf{NW} & \mathbf{CW} & \mathbf{C} & \mathbf{urban} \end{pmatrix}_{(3707\times 10)}$$

$$\times \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{i10} \end{pmatrix}_{(10\times 1)} + \underset{(3707\times 1)}{\mathbf{u}_i}, \qquad i = 1, 2, 3$$

$$= \beta_{i1}\mathbf{1} + \beta_{i2}\mathbf{p}_{beef} + \beta_{i3}\mathbf{p}_{pork} + \beta_{i4}\mathbf{p}_{chicken} + \beta_{i5}\mathbf{m} + \beta_{i6}\mathbf{NE} + \beta_{i7}\mathbf{NW}$$
$$+\beta_{i8}\mathbf{CW} + \beta_{i9}\mathbf{C} + \beta_{i10}\mathbf{urban} + \mathbf{u}_i, \qquad i = 1, 2, 3, \tag{4.3}$$

where $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$ equal $\mathbf{q}_{beef}$, $\mathbf{q}_{pork}$, $\mathbf{q}_{chicken}$ are (3707×1) vectors of observations on the per capita consumption in kilograms (kg) of beef, pork and chicken respectively; $\mathbf{p}_{beef}$, $\mathbf{p}_{pork}$ and $\mathbf{p}_{chicken}$ are (3707×1) vectors of observations on the real price in 2002 Mexican pesos per kilogram (real pesos/kg) of beef, pork and chicken respectively; $\mathbf{m}$ is the (3707×1) vector of observations on the per capita real expenditure on all meats (beef, pork, processed meat, chicken, other meat, and seafood) in 2002 Mexican pesos (real pesos); **NE**, **NW**, **CW**, **C**, and **SE** are (3707×1) vectors formed by "dummy" (or zero-one) variables taking the value of "1" if the observation belongs to the Northeast, Northwest, Central-West, Central or Southeast region respectively, "0" otherwise; and **urban** and **rural** are (3707×1) vectors formed by "dummy" variables taking the value of "1" if the observation belong to the urban or rural sector respectively, "0" otherwise. In equation (4.3) above, notice that the baseline is the rural population of the Southeast region. In other words, we omitted the **SE** and **rural** (3707×1) vectors formed by "dummy" variables to avoid perfect multicollinearity. That is, the **SE** and **rural** vectors formed by "dummy" variables are omitted in order to avoid a perfect linear relation between the vectors **NE**, **NW**, **CW**, **C**, **SE** and the vector $\mathbf{1}_{3707}$ corresponding to the intercept. Similarly, the vector **rural** is omitted in order to avoid a perfect linear relation between vectors **urban**, **rural** and the vector $\mathbf{1}_{3707}$ corresponding to the intercept. Table 4.69 provides a description of the dependent and independent variables used in the estimation of the general SUR model.

The use of dummy variables in equation (4.3) actually estimates a different intercept for each observed region and urbanization level combination, while maintaining the same slope parameters for each of the other independent variables in the model (p_{beef}, p_{pork}, $p_{chicken}$, and m). For this reason, the dummy variables in models like the former are often called "intercept shifters".

For example, the following sub-models can be obtained from equation (4.3).

- Consumption of the i^{th} commodity by the urban population in the Northeast region of Mexico:

$$\mathbf{q}_i = (\beta_{i1} + \beta_{i6} + \beta_{i10})\mathbf{1} + \beta_{i2}\mathbf{p}_{beef} + \beta_{i3}\mathbf{p}_{pork} + \beta_{i4}\mathbf{p}_{chicken} + \beta_{i5}\mathbf{m} + \mathbf{u}_i, \; i = 1, 2, 3. \tag{4.4}$$

- Consumption of the i^{th} commodity by the rural population in the Northeast region of Mexico:

$$\mathbf{q}_i = (\beta_{i1} + \beta_{i6})\mathbf{1} + \beta_{i2}\mathbf{p}_{beef} + \beta_{i3}\mathbf{p}_{pork} + \beta_{i4}\mathbf{p}_{chicken} + \beta_{i5}\mathbf{m} + \mathbf{u}_i, \; i = 1, 2, 3. \tag{4.5}$$

- Consumption of the i^{th} commodity by the urban population in the Northwest region of Mexico:

$$\mathbf{q}_i = (\beta_{i1} + \beta_{i7} + \beta_{i10})\mathbf{1} + \beta_{i2}\mathbf{p}_{beef} + \beta_{i3}\mathbf{p}_{pork} + \beta_{i4}\mathbf{p}_{chicken} + \beta_{i5}\mathbf{m} + \mathbf{u}_i, \; i = 1, 2, 3. \tag{4.6}$$

- Consumption of the i^{th} commodity by the rural population in the Northwest region of Mexico:

$$\mathbf{q}_i = (\beta_{i1} + \beta_{i7})\mathbf{1} + \beta_{i2}\mathbf{p}_{beef} + \beta_{i3}\mathbf{p}_{pork} + \beta_{i4}\mathbf{p}_{chicken} + \beta_{i5}\mathbf{m} + \mathbf{u}_i, \; i = 1, 2, 3. \tag{4.7}$$

- Consumption of the i^{th} commodity by the urban population in the Central-West region of Mexico:

$$\mathbf{q}_i = (\beta_{i1} + \beta_{i8} + \beta_{i10})\mathbf{1} + \beta_{i2}\mathbf{p}_{beef} + \beta_{i3}\mathbf{p}_{pork} + \beta_{i4}\mathbf{p}_{chicken} + \beta_{i5}\mathbf{m} + \mathbf{u}_i, \; i = 1, 2, 3. \tag{4.8}$$

Table 4.69: List of Variables Used in the Mexican Meat Consumption Empirical Application.

Variable	**Description**
q_{beef}	Per capita beef consumption in kilograms (kg)
q_{pork}	Per capita pork consumption in kilograms (kg)
$q_{chicken}$	Per capita chicken consumption in kilograms (kg)
p_{beef}	Real price of beef in 2002 Mexican pesos per kilogram (real pesos/kg)
p_{pork}	Real price of pork in 2002 Mexican pesos per kilogram (real pesos/kg)
$p_{chicken}$	Real price of chicken in 2002 Mexican pesos per kilogram (real pesos/kg)
m	Per capita real expenditure on all meats (beef, pork, processed meat, chicken, other meat, and seafood) in 2002 Mexican pesos (real pesos)
NE	Dummy variable for the Northeast region of Mexico. This variable equals "1" if the observation belongs to the Northeast region, "0" otherwise. This region consists of the states of Chihuahua, Cohahuila de Zaragoza, Durango, Nuevo León, and Tamaulipas.
NW	Dummy variable for the Northwest region of Mexico. This variable equals "1" if the observation belongs to the Northwest region, "0" otherwise. This region consists of the states of Baja California, Sonora, Baja California Sur, and Sinaloa.
CW	Dummy variable for the Central-West region of Mexico. This variable equals "1" if the observation belongs to the Central-West region, "0" otherwise. This region consists of the states of Zacatecas, Mayarit, Aguascalientes, San Luis Potosí, Jalisco, Guanajuato, Querétaro Arteaga, Colima, and Michoacán de Ocampo.
C	Dummy variable for the Central region of Mexico. This variable equals "1" if the observation belongs to the Central region, "0" otherwise. This region consists of the states of Hidalgo, Estado de México, Tlaxcala, Morelos, Puebla, and Distrito Federal.
SE	Dummy variable for the Southeast region of Mexico. This variable equals "1" if the observation belongs to the Southeast region, "0" otherwise. This region consists of the states of Veracruz de Ignacio de la Llave, Yucatán, Quintana Roo, Campeche, Tabasco, Guerrero, Oxaca, and Chiapas.
$urban$	Dummy variable for the urban population. This variable equals "1" if household location is within a population of 15,000 people or more, "0" otherwise.
$rural$	Dummy variable for the rural population. This variable equals "1" if household location is within a population of 14,999 people or less, "0" otherwise.

- Consumption of the i^{th} commodity by the rural population in the Central-West region of Mexico:

$$\mathbf{q}_i = (\beta_{i1} + \beta_{i8})\mathbf{1} + \beta_{i2}\mathbf{p}_{beef} + \beta_{i3}\mathbf{p}_{pork} + \beta_{i4}\mathbf{p}_{chicken} + \beta_{i5}\mathbf{m} + \mathbf{u}_i,\ i = 1, 2, 3. \quad (4.9)$$

- Consumption of the i^{th} commodity by the urban population in the Central region of Mexico:

$$\mathbf{q}_i = (\beta_{i1} + \beta_{i9} + \beta_{i10})\mathbf{1} + \beta_{i2}\mathbf{p}_{beef} + \beta_{i3}\mathbf{p}_{pork} + \beta_{i4}\mathbf{p}_{chicken} + \beta_{i5}\mathbf{m} + \mathbf{u}_i,\ i = 1, 2, 3. \quad (4.10)$$

- Consumption of the i^{th} commodity by the rural population in the Central region of Mexico:

$$\mathbf{q}_i = (\beta_{i1} + \beta_{i9})\mathbf{1} + \beta_{i2}\mathbf{p}_{beef} + \beta_{i3}\mathbf{p}_{pork} + \beta_{i4}\mathbf{p}_{chicken} + \beta_{i5}\mathbf{m} + \mathbf{u}_i,\ i = 1, 2, 3. \quad (4.11)$$

- Consumption of the i^{th} commodity by the urban population in the Southeast region of Mexico:

$$\mathbf{q}_i = (\beta_{i1} + \beta_{i10})\mathbf{1} + \beta_{i2}\mathbf{p}_{beef} + \beta_{i3}\mathbf{p}_{pork} + \beta_{i4}\mathbf{p}_{chicken} + \beta_{i5}\mathbf{m} + \mathbf{u}_i,\ i = 1, 2, 3. \quad (4.12)$$

- Consumption of the i^{th} commodity by the rural population in the Southeast region of Mexico:

$$\mathbf{q}_i = (\beta_{i1})\mathbf{1} + \beta_{i2}\mathbf{p}_{beef} + \beta_{i3}\mathbf{p}_{pork} + \beta_{i4}\mathbf{p}_{chicken} + \beta_{i5}\mathbf{m} + \mathbf{u}_i,\ i = 1, 2, 3. \quad (4.13)$$

Therefore, a model like the one provided in equation (4.3) assumes that regional or urbanization factors shift the consumption of the i^{th} commodity in a parallel fashion as shown in equations (4.4) through (4.13). Hence, the underlying assumption of the model in equation (4.3) is that regional and urbanization-level differences in consumption of the i^{th} commodity can be appropriately modeled by parallel shifting the sub-models.

Now, writing all 3 equations in equation (4.2) into one model gives

$$\begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \\ \boldsymbol{\beta_3} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \end{bmatrix} \tag{4.14}$$

or

$$\underset{(3(3707)\times 1)}{\mathbf{q}} = \underset{(3(3707)\times 30)}{\mathbf{X}} \underset{(30\times 1)}{\boldsymbol{\beta}} + \underset{(3(3707)\times 1)}{\mathbf{u}}. \tag{4.15}$$

As we explained in Section 3.5, since we are dealing with a stratified sample (see Section 4.1), before estimating the parameters of the model in equations (4.14) or (4.15) by equation (2.12), we need to weight all the observations by the weight variable (see Table 4.2) as it is done in weighted least squares.[10] In fact, the weighted estimator is consistent for $\boldsymbol{\beta}$ (Wooldridge, 2001, p. 464).[11] SAS software allows to estimate the system of equations in equation (4.14) using Seemingly Unrelated Regressions as well as weighting each observation by a specified weight. However, as it was explained in Section 3.5, "[if we] use weights w_i in the weighted least squares estimation, [we] will obtain the same point estimates...; however, in complex surveys, the standard errors and hypothesis tests the software provides will be incorrect and should be ignored" (Lohr, 1999, p. 355). Hence, to calculate better estimates of the standard errors and hypothesis tests for the parameter estimates, this study applies the bootstrap by using SAS software. As explained in Section 3.5, the Bootstrap is a resampling method that can be used to estimate standard errors where other estimation methods are inappropriate. Shao and Tu (1995) summarize theoretical results for the bootstrap in complex survey samples. Wooldridge (2001, p. 464) provides an alternative procedure

[10] Weighted least squares is a special case of generalized least squares. Assuming $var[u_i(t)] = \sigma^2 w_i(t)$, where $w_i(t)$ is the weight of observation t in the i^{th} commodity equation, then $var\left(\frac{u_i(t)}{\sqrt{w_i(t)}}\right) = \sigma^2$.

[11] An estimator is consistent if the probability that the estimator and the true parameter differ by any arbitrary small positive number approaches zero as the sample size approaches infinity. That is, $\underset{T\to\infty}{lim}\, Pr\left(\left|\hat{\beta}_{ik} - \beta_{ik}\right| > \varepsilon\right) = 0$ for any $\varepsilon > 0$. Or equivalently, $\hat{\beta}_{ik} - \beta_{ik} = o_p(1)$ or $\hat{\beta}_{ik} \xrightarrow{P} \beta_{ik}$.

to the bootstrap to calculate asymptotic variances of the parameter estimates when we use one equation instead of a system of equations (see equation (3.3)).

4.3.2 Estimation of Individual SUR Models

The SUR model in equation (4.15) will also be estimated without the vectors **NE**, **NW**, **CW**, **C**, and **urban** in equation (4.3) for each corresponding dataset of Table 4.59 through Table 4.68. That is, subsets of the dataset containing all regions and urbanization levels (dataset corresponding to Table 4.58) are obtained for each urbanization level within each region (datasets corresponding to Table 4.59 through 4.68); and then the model in equation (4.15) is estimated without the vectors **NE**, **NW**, **CW**, **C**, and **urban** in equation (4.3). In simple words, individual SUR models will be estimated for the urban and rural sectors within each region. In stratified sampling the elements of the same stratum often tend to be more similar than randomly selected elements from the whole population; therefore, individual models will also be estimated. Individual models are not comparable to those obtained from equations (4.4) through (4.13). If comparisons between the individual models and sub-models obtained from equation (4.3) are desired, in addition to the dummy variables in equation (4.3) interaction of dummy variable with the real price variables and real expenditure variable need to be included in equation (4.3). That is, twenty more (3707×1) vectors ($7 - 2 = 5$ dummy variables times 4 regular price and expenditure variables) will have to be included in equation (4.3).

CHAPTER V
RESULTS

In this chapter results and findings from the estimation of SUR models are presented. Section 5.1 focuses on the results of the general model for the Mexican meat consumption presented in Section 4.3.1. In particular, it discusses the results of the parameter estimates corresponding to the price variables, the parameter estimates corresponding to the regional dummy variables, and the parameter estimate for the urban dummy variable. Additionally, this section also explains how to interpret the parameter estimates. Section 5.2 reports our findings when the individual SUR models presented in Section 4.3.2 are estimated for the urban and rural sector within each region. In stratified sampling the elements of the same stratum often tend to be more similar than randomly selected elements from the whole population; therefore, several individual models that analyze tastes and preferences of consumers at different urbanization levels within five Mexican regions are considered.

5.1 One General SUR Model

The results of the model in equation (4.14) or equation (4.15) are shown in Table 5.1. The sum of weights reported in Table 5.1 is the number of households that our results represent nationally. That is, the 3,707 households, who reported consumption of at lest one meat cut of beef, pork and chicken (see Table 4.56), represent 5.3 million households nationally. A total of 22.1 million households (see Table 4.54) consumed at least one meat cut during the week of the interview nationally. The number of bootstrap resamples is the number of resamples that were taken in order to estimate the standard errors and 95% confidence intervals of the estimated parameters.[1] An R-square (also called coefficient of determination) is reported for each equation. An R-square equal to 0.3641 means that 36.41% of the total variation in the dependent

[1] The size of each resample was set to the size of the input data set (i.e., the number of observations).

variable (q_i) is explained by the model (Pindyck and Rubinfeld, 1997, p. 89). The R-square of the pork equation was lower than the R-square of the beef and chicken equation. In Section 5.2, the R-square of the individual models for the urban and rural sectors within each region are reported. As explained in Section 4.3.2 the R-square of the individual models is not comparable to the R-square of the general model.

The model in equation (4.14) or equation (4.15) has three equations and thirty parameters. In Table 5.1, fifteen parameters were insignificant at the 0.05 level and twelve parameters were insignificant at the 0.1 level. Of the twelve insignificant parameters at the 0.1 level, 6 parameters correspond to price variables, 3 correspond to regional dummy variables, and 3 correspond to urbanization-level variables. However, all "own price" parameter estimates (i.e., the corresponding parameter estimates of the price of beef in the beef equation, the price of pork in the pork equation, and the price of chicken in the chicken equation) came with the correct negative sign and were statistically different from zero at the 0.05 significance level. This means that an increase in the own price of beef, pork or chicken will decrease the consumption of beef, pork or chicken, respectively. For example, increasing the price of beef by 1 real peso (i.e., 1 Mexican peso expressed in the nominal value of 2002 Mexican pesos),[2] holding all other factors affecting beef consumption constant, will decrease the average weekly per capita consumption of beef by 0.006180 kg (or 0.013624 lbs).[3] Since the average household has approximately 4 adult equivalents (which is approximately 4.14 people on average in Mexico), the average household of 4 adult equivalents (or 4.14 people) will decrease the average weekly consumption of beef by 0.02472 kg (or 0.0545 lbs). However, an increase in the price of beef by 5 real pesos (holding all other factors

[2]The exchange rate for the second half of June 2002 is US \$1 = Mexican \$9.78906. This is the exchange rate reported by Banco de Mexico and used to calculate the amount of Mexican pesos that must be used to pay debts in U.S. dollars within Mexico ("Para Pagos"). The exchange rate for the second half of June 2002 is reported to be consistent with the base period for the CPI of Banco de Mexico.

[3]1 kg $\approx$ 2.2046 lbs.

affecting beef consumption constant) will decrease the average weekly per capita consumption of beef by 0.0309 kg (or 0.068122 lbs), and the average household of 4 adult equivalents (or 4.14 people) will decrease their average weekly consumption of beef by 0.1236 kg (or 0.2724 lbs). Similarly, increasing the price of pork by 1 real peso (holding other factors affecting the consumption of pork constant) will decrease the average weekly per capita consumption of pork by 0.00475 kg. Finally, increasing the price of chicken by 1 real peso (holding other factors affecting the consumption of chicken constant) will decrease the average weekly per capita consumption of chicken by 0.009840 kg.

All the own price parameter estimates are the only price parameter estimates that are statistically different from zero at both the 0.05 and 0.1 significant levels. All other price parameter estimates in Table 5.1 were insignificant at the 0.1 level. That is, there is not enough statistical evidence to conclude that changes in the price of pork or chicken have an effect on the consumption of beef. In general, we would expect that changes in the prices of pork or chicken (holding other factors affecting the consumption of beef constant) will change the average consumption of beef. For example, if the parameter corresponding to the variable $p_{chicken}$ were statistically significant in the beef equation (see Table 5.1), then increasing the price of chicken by 1 real peso (holding other factors affecting the consumption of beef constant) will increase the average weekly consumption of beef by 0.000736 kg. On the other hand, income designated to meat expenditures is statistically different from zero at the 0.05 significant level in all three equations. Hence, increasing the household income designated to meat expenditures by 1 real peso, increases the average weekly per capita consumption of beef by 0.005389 kg, the average weekly per capita consumption of pork by 0.005158 kg, and the average weekly per capita consumption of chicken by 0.007507 kg while holding the price of beef, pork and chicken constant.

The parameter estimates corresponding to the regional dummy variables for the Northeast and Northwest regions are statistically different from zero at the 0.05 or 0.1 level in all three equations (see Table 5.1). Hence, the per capita consumption of beef in the Northeast and Northwest regions are higher than the per capita consumption of beef (first equation or beef section in Table 5.1) in the Southeast region (the excluded dummy variable) regardless of the values taken by all other variables (beef price, pork price, chicken price and expenditures on meat). Similarly, the per capita consumption of pork (second equation or pork section in Table5.1) or chicken (third equation or chicken section in Table 5.1) in the Northeast and Northwest regions is lower than the per capita consumption of pork or chicken in the Southeast region. Now consider the parameter estimates corresponding to the dummy variable CW. There is not enough statistical evidence to conclude that there is a difference between the per capita consumption of beef or pork in the Central-West region and the Southeast region. On the contrary, there is enough statistical evidence at the 0.05 significance level to conclude that the per capita consumption of chicken is lower in the Central-West region than in the Southeast region. The next parameter estimates corresponds to the dummy variable C. In this case, there is not enough statistical evidence at the 0.1 significance level to conclude that the per capita consumption of pork in the Central region and the Southeast region are different. However, there is enough statistical evidence at the 0.05 and 0.1 significant levels that per capita consumption of beef and chicken are statistically different in those regions. Finally, consider the parameter estimates corresponding to the urban dummy variable *urban*. There is not enough statistical evidence at the 0.1 significance level to conclude that the urban and rural sectors have different per capita consumption of beef, pork or chicken.

The estimate of the sub-models in equations (4.4) through (4.13) can be obtained by replacing the true parameters in equations (4.4) through (4.13) by the corresponding parameter estimates of the i^{th} commodity (i = 1, 2, 3 = beef, pork, chicken) in Table 5.1. However, the underlying assumption of the model in equation (4.3) is that regional or urbanization factors shift the consumption of the i^{th} commodity in a parallel fashion. If this assumption is true, then there would be no need to have a model for the urban and rural sectors because the parameter estimates corresponding to the urban dummy variables are statistically insignificant at the 0.1 level in all three equations. For the same reason, there would be no need to distinguish between consumption of beef or pork in the Central-West and Southeast regions (estimate of parameter β_{18} or estimate of parameter β_{28} respectively), and the consumption of pork in the Central and Southeast regions (estimate of parameter β_{29}). Therefore, if the above assumption is true, instead of having thirty sub-models (ten for each commodity–see equations (4.4) through (4.13)), there would only be twelve sub-models (four sub-models for beef, three sub-model for pork, and five sub-models for chicken). However, if the above assumption is false, the estimates of the parameters β_{18}, β_{110}, β_{28}, β_{29}, β_{210} and β_{310} may be statistically insignificant simply because the parallel shifts do not reflect the real situation for each urban or rural sector within each region.

In section 5.2, an individual model is estimated for the urban and rural sector within each region. In stratified sampling elements of the same stratum often tend to be more similar than randomly selected elements from the whole population; therefore, several individual models rather than one general model are considered.

Table 5.1: SUR Parameter Estimates, All Strata.

Number of Observations	3707
Sum of Weights	5303145
Number of Bootstrap Resamples	1000
Beef-Equation R-square	0.3641
Pork-Equation R-square	0.0841
Chicken-Equation R-square	0.3208

Par.	**Variable**	**Parameter Estimate**	**Approx. Bootstr. Std. Err.**	**Approx. Bootstr. 95% LCL**	**Approx. Bootstr. 95% UCL**	**Approx. Bootstr. p-value**
			Beef			
β_{11}	Intercept	0.346868**	0.03858	0.27724	0.42846	< 0.00001
β_{12}	p_{beef}	-0.006180**	0.00068	-0.00737	-0.00470	< 0.00001
β_{13}	p_{pork}	-0.000090	0.00028	-0.00059	0.00051	0.88031
β_{14}	$p_{chicken}$	0.000736	0.00061	-0.00027	0.00213	0.12984
β_{15}	m	0.005389**	0.00146	0.00206	0.00777	0.00075
β_{16}	NE	0.114659**	0.02601	0.06388	0.16584	0.00001
β_{17}	NW	0.103198**	0.02593	0.05430	0.15209	0.00009
β_{18}	CW	0.027630	0.02309	-0.02295	0.06756	0.33404
β_{19}	C	0.028585*	0.01459	-0.00174	0.05545	0.06567
β_{110}	$urban$	0.010061	0.01483	-0.01786	0.04025	0.45014
			Pork			
β_{21}	Intercept	0.389742**	0.12727	0.13817	0.63708	0.00232
β_{22}	p_{beef}	0.000121	0.00056	-0.00095	0.00125	0.78708
β_{23}	p_{pork}	-0.004750**	0.00188	-0.00828	-0.00091	0.01457
β_{24}	$p_{chicken}$	-0.000750	0.00081	-0.00238	0.00080	0.32874
β_{25}	m	0.005158**	0.00064	0.00378	0.00630	< 0.00001
β_{26}	NE	-0.133980*	0.07282	-0.27600	0.00944	0.06720
β_{27}	NW	-0.117400**	0.05978	-0.23586	-0.00152	0.04710
β_{28}	CW	-0.108500	0.07319	-0.25365	0.03326	0.13218
β_{29}	C	-0.089910	0.06291	-0.21515	0.03147	0.14435
β_{210}	$urban$	-0.057790	0.05116	-0.15761	0.04293	0.26233
			Chicken			
β_{31}	Intercept	0.480953**	0.05131	0.38113	0.58227	< 0.00001
β_{32}	p_{beef}	0.000157	0.00065	-0.00103	0.00154	0.69602
β_{33}	p_{pork}	-0.000240	0.00038	-0.00097	0.00510	0.54208
β_{34}	$p_{chicken}$	-0.009840**	0.00145	-0.01234	-0.00668	< 0.00001
β_{35}	m	0.007507**	0.00123	0.00472	0.00956	< 0.00001
β_{36}	NE	-0.073920**	0.02643	-0.12630	-0.02270	0.00482
β_{37}	NW	-0.056650*	0.03209	-0.12100	0.00477	0.07011
β_{38}	CW	-0.128570**	0.02552	-0.18322	-0.08317	< 0.00001
β_{39}	C	-0.070190**	0.02095	-0.11242	-0.03028	0.00066
β_{310}	$urban$	0.025974	0.01974	-0.01074	0.06665	0.15683

Note: Bootstrap significance levels of 0.05 and 0.1 are indicated by ** and * respectively.

5.2 Individual SUR Models

Individual SUR models were estimated for the urban and rural sectors within each region. That is, the SUR model in equation (4.15) was estimated without the vectors **NE**, **NW**, **CW**, **C**, and **urban** in equation (4.3) for each corresponding dataset of Table 4.59 through Table 4.68. Tables 5.2 through 5.11 report the parameter estimates as well as their standard errors and 95% confidence intervals. In addtion, the number of observations in the sample, the sum of weights, the number of bootstrap resamples, and a R-square for each equation was reported. The sum of weights is the total number of households that the number of households in the sample (the number of observations) represent in the corresponding sector and region. The number of bootsrap resamples considered was 1,000.

The SUR models estimated in Tables 5.2 through 5.11 have three equations and fifteen parameters. Out of the total fifteen parameters estimated, there are usually seven or six insignificant parameter estimates in each table at the 0.05 or 0.1 level respectively. The parameter estimates corresponding to own prices are all with the correct sign. Similarly, they are all statistically different from zero at both 0.05 and 0.10 significance levels except for the own price of chicken in Table 5.9 and the own price of pork in Tables 5.3, 5.10 and 5.11. In the beef equation, the price of pork is always insignificant at the 0.1 level, and the price of chicken is in three occasions significant at the 0.05 or 0.1 level (one occasion positive and significant in Table 5.8, two occasions negative and significant in Tables 5.2 and 5.8). In the pork equation, the price of beef is only in one occasion significant at the 0.1 level (negative and significant in Table 5.5), and the price of chicken is in four occasions significant at the 0.05 or 0.1 level (four occasions negative and significant in Tables 5.2, 5.3, 5.4, and 5.8). In the chicken equation, the price of beef is in two occasions significant at the 0.05 level (two occasions negative and significant in Tables 5.5 and 5.9), the price of pork is in two occasions significant at the 0.05 or 0.1 level (one occasion positive and significant in Table 5.5 and one occasion negative and significant in Table 5.10). Therefore, in the pork and chicken equations, when the price of beef is significant, it is negative. Only in the pork equation, when the price of chicken is significant, it is negative.

Furthermore, income designated to meat expenditures is always significant and positive in Tables 5.2 through Table 5.11. When significant, price parameter estimates can be interpreted in the same fashion that they were interpreted in Section 5.1. For example, if the parameter estimate of the price of the j^{th} commodity is positive in the i^{th} equation; then, an increase (decrease) of one real peso in the price of the j^{th} commodity will increase (decrease) the average weekly per capita consumption of the i^{th} commodity by the value of the parameter estimate, holding other factors affecting the per capita consumption of the i^{th} commodity constant. Similarly, if the parameter estimate of the price of the j^{th} commodity is negative in the i^{th} equation; then, an increase (decrease) of one real peso in the price of the j^{th} commodity will decrease (increase) the average weekly per capita consumption of the i^{th} commodity by the absolute value of the parameter estimate, holding other factors affecting the per capita consumption of the i^{th} commodity constant.

Finally, parameter estimates from individual models are not comparable to those obtained from the general model. If comparisons between the individual models and sub-models obtained from the general model (equation (4.3)) are desired, in addition to the dummy variables in equation (4.3) interaction of dummy variable with the real price variables and real expenditure variable need to be included in equation (4.3). That is, twenty four more (3707×1) vectors (6 dummy variables times 4 regular price and expenditure variables) will have to be included in the general model (equation (4.3)).

Table 5.2: SUR Parameter Estimates, Urban Sector in Northeast Region.

Number of Observations	216
Sum of Weights	241965
Number of Bootstrap Resamples	1000
Beef-Equation R-square	0.5315
Pork-Equation R-square	0.4305
Chicken-Equation R-square	0.3903

Par.	**Variable**	**Parameter Estimate**	**Approx. Bootstr. Std. Err.**	**Approx. Bootstr. 95% LCL**	**Approx. Bootstr. 95% UCL**	**Approx. Bootstr. p-value**
			Beef			
β_{11}	Intercept	0.512882**	0.11038	0.28550	0.71816	< 0.00001
β_{12}	p_{beef}	-0.008370**	0.00124	-0.01082	-0.00595	< 0.00001
β_{13}	p_{pork}	-0.000310	0.00229	-0.00437	0.00461	0.95896
β_{14}	$p_{chicken}$	-0.005170**	0.00200	-0.00924	-0.00139	0.00792
β_{15}	m	0.009649**	0.00125	0.00720	0.01208	< 0.00001
			Pork			
β_{21}	Intercept	0.245647**	0.05330	0.13848	0.34741	< 0.00001
β_{22}	p_{beef}	0.000095	0.00093	-0.00166	0.00197	0.86806
β_{23}	p_{pork}	-0.004650**	0.00083	-0.00618	-0.00293	< 0.00001
β_{24}	$p_{chicken}$	-0.002190**	0.00111	-0.00446	-0.00011	0.03967
β_{25}	m	0.005080**	0.00094	0.00323	0.00690	< 0.00001
			Chicken			
β_{31}	Intercept	0.448229**	0.09204	0.25850	0.61928	< 0.00001
β_{32}	p_{beef}	-0.000210	0.00139	-0.00294	0.00252	0.87795
β_{33}	p_{pork}	-0.000300	0.00173	-0.00339	0.00339	0.99997
β_{34}	$p_{chicken}$	-0.010660**	0.00196	-0.01459	-0.00889	< 0.00001
β_{35}	m	0.008506**	0.00108	0.00640	0.01062	< 0.00001

Note: Bootstrap significance levels of 0.05 and 0.1 are indicated by ** and * respectively.

Table 5.3: SUR Parameter Estimates, Rural Sector in Northeast Region.

Number of Observations	65
Sum of Weights	58040
Number of Bootstrap Resamples	1000
Beef-Equation R-square	0.6279
Pork-Equation R-square	0.5093
Chicken-Equation R-square	0.1802

Par.	**Variable**	**Parameter Estimate**	**Approx. Bootstr. Std. Err.**	**Approx. Bootstr. 95% LCL**	**Approx. Bootstr. 95% UCL**	**Approx. Bootstr. p-value**
			Beef			
β_{11}	Intercept	0.595225**	0.22702	0.19903	1.08895	0.00456
β_{12}	p_{beef}	-0.015500**	0.00505	-0.02690	-0.00710	0.00077
β_{13}	p_{pork}	0.002124	0.00313	-0.00409	0.00818	0.51345
β_{14}	$p_{chicken}$	0.001613	0.00520	-0.00744	0.01293	0.59759
β_{15}	m	0.011809**	0.00373	0.00454	0.01918	0.00150
			Pork			
β_{21}	Intercept	0.080998	0.12934	-0.21300	0.29402	0.75414
β_{22}	p_{beef}	0.000739	0.00198	-0.00286	0.00491	0.60544
β_{23}	p_{pork}	-0.001740	0.00171	-0.00465	0.00207	0.45221
β_{24}	$p_{chicken}$	-0.005700*	0.00363	-0.01368	0.00056	0.07112
β_{25}	m	0.008047**	0.00452	0.00040	0.01812	0.04054
			Chicken			
β_{31}	Intercept	0.409135**	0.16600	0.11247	0.76318	0.00835
β_{32}	p_{beef}	0.001362	0.00261	-0.00416	0.00607	0.71373
β_{33}	p_{pork}	0.000478	0.00218	-0.00409	0.00444	0.93420
β_{34}	$p_{chicken}$	-0.009780*	0.00346	-0.01546	-0.00192	0.01193
β_{35}	m	0.002919	0.00300	-0.00393	0.00784	0.51499

Note: Bootstrap significance levels of 0.05 and 0.1 are indicated by ** and * respectively.

Table 5.4: SUR Parameter Estimates, Urban Sector in Northwest Region.

Number of Observations	291
Sum of Weights	465235
Number of Bootstrap Resamples	1000
Beef-Equation R-square	0.5469
Pork-Equation R-square	0.3233
Chicken-Equation R-square	0.2641

Par.	**Variable**	**Parameter Estimate**	**Approx. Bootstr. Std. Err.**	**Approx. Bootstr. 95% LCL**	**Approx. Bootstr. 95% UCL**	**Approx. Bootstr. p-value**
			Beef			
β_{11}	Intercept	0.292169**	0.08414	0.13098	0.46079	0.00044
β_{12}	p_{beef}	-0.007570**	0.00189	-0.01146	-0.00405	0.00004
β_{13}	p_{pork}	-0.000570	0.00051	-0.00154	0.00046	0.28679
β_{14}	$p_{chicken}$	0.001200	0.00109	-0.00089	0.00339	0.25193
β_{15}	m	0.011655**	0.00141	0.00900	0.01451	< 0.00001
			Pork			
β_{21}	Intercept	0.190671**	0.03600	0.11440	0.25552	< 0.00001
β_{22}	p_{beef}	-0.000970	0.00079	-0.00249	0.00062	0.23815
β_{23}	p_{pork}	-0.002000**	0.00054	-0.00289	-0.00078	0.00067
β_{24}	$p_{chicken}$	-0.001680**	0.00077	-0.00321	-0.00018	0.02850
β_{25}	m	0.004692**	0.00079	0.00307	0.00617	< 0.00001
			Chicken			
β_{31}	Intercept	0.507386**	0.10979	0.27019	0.70057	< 0.00001
β_{32}	p_{beef}	0.001055	0.00255	-0.00348	0.00649	0.55455
β_{33}	p_{pork}	-0.000006	0.00076	-0.00141	0.00156	0.91877
β_{34}	$p_{chicken}$	-0.014430**	0.00407	-0.02243	-0.00649	0.00038
β_{35}	m	0.007658**	0.00134	0.00507	0.01034	< 0.00001

Note: Bootstrap significance levels of 0.05 and 0.1 are indicated by ** and * respectively.

Table 5.5: SUR Parameter Estimates, Rural Sector in Northwest Region.

Number of Observations	50
Sum of Weights	42828
Number of Bootstrap Resamples	1000
Beef-Equation R-square	0.5962
Pork-Equation R-square	0.4337
Chicken-Equation R-square	0.5057

Par.	**Variable**	**Parameter Estimate**	**Approx. Bootstr. Std. Err.**	**Approx. Bootstr. 95% LCL**	**Approx. Bootstr. 95% UCL**	**Approx. Bootstr. p-value**
			Beef			
β_{11}	Intercept	0.740964**	0.34910	0.23082	1.59929	0.00876
β_{12}	p_{beef}	-0.017030**	0.00822	-0.03585	-0.00363	0.01633
β_{13}	p_{pork}	0.003104	0.00334	-0.00306	0.01003	0.29674
β_{14}	$p_{chicken}$	0.001602	0.00381	-0.00535	0.00959	0.57780
β_{15}	m	0.005925	0.00455	-0.00545	0.01239	0.44581
			Pork			
β_{21}	Intercept	0.481360**	0.19212	0.18076	0.93387	0.00372
β_{22}	p_{beef}	-0.004260*	0.00257	-0.00996	0.00013	0.05612
β_{23}	p_{pork}	-0.004100**	0.00172	-0.00740	-0.00066	0.01908
β_{24}	$p_{chicken}$	-0.000990	0.00162	-0.00437	0.00198	0.46243
β_{25}	m	0.002762	0.00276	-0.00402	0.00678	0.61637
			Chicken			
β_{31}	Intercept	0.615571**	0.22320	0.20155	1.07648	0.00420
β_{32}	p_{beef}	-0.007100**	0.00356	-0.01457	-0.00063	0.03256
β_{33}	p_{pork}	0.004984*	0.00286	-0.00038	0.01082	0.06779
β_{34}	$p_{chicken}$	-0.009340**	0.00337	-0.01528	-0.00206	0.01010
β_{35}	m	0.004909	0.00260	-0.00135	0.00884	0.15951

Note: Bootstrap significance levels of 0.05 and 0.1 are indicated by ** and * respectively.

Table 5.6: SUR Parameter Estimates, Urban Sector in Central-West Region.

Number of Observations	887
Sum of Weights	834796
Number of Bootstrap Resamples	1000
Beef-Equation R-square	0.3265
Pork-Equation R-square	0.7230
Chicken-Equation R-square	0.4600

Par.	**Variable**	**Parameter Estimate**	**Approx. Bootstr. Std. Err.**	**Approx. Bootstr. 95% LCL**	**Approx. Bootstr. 95% UCL**	**Approx. Bootstr. p-value**
			Beef			
β_{11}	Intercept	0.416511**	0.092523	0.24890	0.61159	< 0.00001
β_{12}	p_{beef}	-0.004720**	0.001492	-0.00733	-0.00148	0.00316
β_{13}	p_{pork}	0.000746	0.000975	-0.00077	0.00305	0.24346
β_{14}	$p_{chicken}$	-0.000350	0.000784	-0.00170	0.00137	0.83511
β_{15}	m	0.002697	0.002286	-0.00314	0.00582	0.55702
			Pork			
β_{21}	Intercept	0.225220**	0.035018	0.15767	0.29494	< 0.00001
β_{22}	p_{beef}	-0.000480	0.000462	-0.00134	0.00048	0.35301
β_{23}	p_{pork}	-0.003240**	0.000431	-0.00398	-0.00229	< 0.00001
β_{24}	$p_{chicken}$	-0.000410	0.000618	-0.00160	0.00082	0.52992
β_{25}	m	0.004338**	0.000434	0.00327	0.00497	< 0.00001
			Chicken			
β_{31}	Intercept	0.381302**	0.060955	0.27297	0.51191	< 0.00001
β_{32}	p_{beef}	0.000270	0.001057	-0.00166	0.00249	0.69386
β_{33}	p_{pork}	0.000773	0.000881	-0.00077	0.00296	0.27612
β_{34}	$p_{chicken}$	-0.008480**	0.001427	-0.01114	-0.00555	< 0.00001
β_{35}	m	0.004991**	0.001484	0.00129	0.00710	0.00471

Note: Bootstrap significance levels of 0.05 and 0.1 are indicated by ** and * respectively.

Table 5.7: SUR Parameter Estimates, Rural Sector in Central-West Region.

Number of Observations	207
Sum of Weights	291773
Number of Bootstrap Resamples	1000
Beef-Equation R-square	0.7501
Pork-Equation R-square	0.3197
Chicken-Equation R-square	0.3562

Par.	**Variable**	**Parameter Estimate**	**Approx. Bootstr. Std. Err.**	**Approx. Bootstr. 95% LCL**	**Approx. Bootstr. 95% UCL**	**Approx. Bootstr. p-value**
			Beef			
β_{11}	Intercept	0.202539	0.11252	-0.05452	0.38656	0.14010
β_{12}	p_{beef}	-0.004740**	0.00182	-0.00803	-0.00091	0.01396
β_{13}	p_{pork}	-0.001520	0.00102	-0.00365	0.00033	0.10250
β_{14}	$p_{chicken}$	0.000317	0.00265	-0.00479	0.00558	0.88015
β_{15}	m	0.010974**	0.00258	0.00694	0.01707	< 0.00001
			Pork			
β_{21}	Intercept	0.235395**	0.08941	0.06927	0.41976	0.00624
β_{22}	p_{beef}	0.000317	0.00124	-0.00197	0.00290	0.70979
β_{23}	p_{pork}	-0.002560**	0.00061	-0.00358	-0.00120	0.00009
β_{24}	$p_{chicken}$	-0.002030	0.00146	-0.00499	0.00074	0.14533
β_{25}	m	0.003496	0.00169	-0.00056	0.00608	0.10329
			Chicken			
β_{31}	Intercept	0.465114**	0.14459	0.22630	0.79309	0.00042
β_{32}	p_{beef}	0.000017	0.00239	-0.00499	0.00436	0.89514
β_{33}	p_{pork}	-0.000280	0.00132	-0.00299	0.00220	0.76639
β_{34}	$p_{chicken}$	-0.011510**	0.00321	-0.01799	-0.00542	0.00026
β_{35}	m	0.006863**	0000191	0.00250	0.00999	0.00109

Note: Bootstrap significance levels of 0.05 and 0.1 are indicated by ** and * respectively.

Source: Computed by author.

Table 5.8: SUR Parameter Estimates, Urban Sector in Central Region.

Number of Observations	918
Sum of Weights	1986442
Number of Bootstrap Resamples	1000
Beef-Equation R-square	0.4607
Pork-Equation R-square	0.3615
Chicken-Equation R-square	0.2195

Par.	**Variable**	**Parameter Estimate**	**Approx. Bootstr. Std. Err.**	**Approx. Bootstr. 95% LCL**	**Approx. Bootstr. 95% UCL**	**Approx. Bootstr. p-value**
			Beef			
β_{11}	Intercept	0.343246**	0.043271	0.25653	0.42615	< 0.00001
β_{12}	p_{beef}	-0.006400**	0.000702	-0.00783	-0.00508	< 0.00001
β_{13}	p_{pork}	-0.000090	0.000398	-0.00087	0.00069	0.81362
β_{14}	$p_{chicken}$	0.001221*	0.000804	-0.00005	0.00310	0.05769
β_{15}	m	0.006343**	0.000691	0.00490	0.00761	< 0.00001
			Pork			
β_{21}	Intercept	0.273154**	0.035296	0.20320	0.34156	< 0.00001
β_{22}	p_{beef}	-0.000640	0.000543	-0.00167	0.00046	0.26182
β_{23}	p_{pork}	-0.003830**	0.000365	-0.00456	-0.00313	< 0.00001
β_{24}	$p_{chicken}$	-0.001710**	0.000496	-0.00269	-0.00075	0.00053
β_{25}	m	0.005097**	0.000720	0.00370	0.00652	< 0.00001
			Chicken			
β_{31}	Intercept	0.461323**	0.091832	0.26132	0.62129	< 0.00001
β_{32}	p_{beef}	-0.000920	0.001207	-0.00343	0.00130	0.37813
β_{33}	p_{pork}	-0.001410	0.000956	-0.00332	0.00043	0.13027
β_{34}	$p_{chicken}$	-0.009430**	0.002945	-0.01406	-0.00252	0.00487
β_{35}	m	0.009350*	0.001532	-0.00628	0.01228	< 0.00001

Note: Bootstrap significance levels of 0.05 and 0.1 are indicated by ** and * respectively.

Table 5.9: SUR Parameter Estimates, Rural Sector in Central Region.

Number of Observations	198
Sum of Weights	4415894
Number of Bootstrap Resamples	1000
Beef-Equation R-square	0.5087
Pork-Equation R-square	0.1908
Chicken-Equation R-square	0.3560

Par.	**Variable**	**Parameter Estimate**	**Approx. Bootstr. Std. Err.**	**Approx. Bootstr. 95% LCL**	**Approx. Bootstr. 95% UCL**	**Approx. Bootstr. p-value**
			Beef			
β_{11}	Intercept	0.295745**	0.07597	0.14328	0.44106	0.00012
β_{12}	p_{beef}	-0.003700**	0.00137	-0.00626	-0.00089	0.00916
β_{13}	p_{pork}	-0.000660	0.00097	-0.00268	0.00113	0.42396
β_{14}	$p_{chicken}$	-0.003510**	0.00140	-0.00631	-0.00081	0.01124
β_{15}	m	0.007911**	0.00103	0.00603	0.01007	< 0.00001
			Pork			
β_{21}	Intercept	0.223519**	0.09089	0.04675	0.40305	0.01335
β_{22}	p_{beef}	0.000245	0.00123	-0.00226	0.00255	0.90524
β_{23}	p_{pork}	-0.003440**	0.00074	-0.00485	-0.00194	< 0.00001
β_{24}	$p_{chicken}$	-0.000900	0.00147	-0.00361	0.00217	0.62465
β_{25}	m	0.004756**	0.00101	0.00267	0.00662	< 0.00001
			Chicken			
β_{31}	Intercept	0.503726**	0.19494	0.12359	0.88775	0.00949
β_{32}	p_{beef}	-0.007220**	0.00288	-0.01288	-0.00160	0.01190
β_{33}	p_{pork}	0.001672	0.00148	-0.00107	0.00475	0.21458
β_{34}	$p_{chicken}$	-0.006060	0.00385	-0.01333	0.00177	0.13347
β_{35}	m	0.009075**	0.00208	0.00452	0.01268	0.00003

Note: Bootstrap significance levels of 0.05 and 0.1 are indicated by ** and * respectively.

Table 5.10: SUR Parameter Estimates, Urban Sector in Southeast Region.

Number of Observations	617
Sum of Weights	581798
Number of Bootstrap Resamples	1000
Beef-Equation R-square	0.2720
Pork-Equation R-square	0.1656
Chicken-Equation R-square	0.4628

Par.	**Variable**	**Parameter Estimate**	**Approx. Bootstr. Std. Err.**	**Approx. Bootstr. 95% LCL**	**Approx. Bootstr. 95% UCL**	**Approx. Bootstr. p-value**
			Beef			
β_{11}	Intercept	0.404887**	0.07822	0.25235	0.55897	< 0.00001
β_{12}	p_{beef}	-0.008170**	0.00219	-0.01226	-0.00369	0.00026
β_{13}	p_{pork}	0.000025	0.00051	-0.00114	0.00085	0.77230
β_{14}	$p_{chicken}$	-0.000500	0.00071	-0.00190	0.00088	0.47064
β_{15}	m	0.007282**	0.00127	0.00476	0.00976	< 0.00001
			Pork			
β_{21}	Intercept	0.201835	0.10365	-0.04589	0.36042	0.12922
β_{22}	p_{beef}	0.000015	0.00122	-0.00248	0.00229	0.94012
β_{23}	p_{pork}	-0.002890	0.00271	-0.00709	0.00353	0.51072
β_{24}	$p_{chicken}$	-0.000570	0.00092	-0.00241	0.00120	0.51205
β_{25}	m	0.005065**	0.00060	0.00391	0.00625	< 0.00001
			Chicken			
β_{31}	Intercept	0.335855**	0.09966	0.13096	0.52160	0.00106
β_{32}	p_{beef}	0.007250	0.00139	-0.00205	0.00342	0.62404
β_{33}	p_{pork}	-0.001120*	0.00069	-0.00267	0.00002	0.05406
β_{34}	$p_{chicken}$	-0.009470**	0.00265	-0.01394	-0.00353	0.00100
β_{35}	m	0.011001**	0.00151	0.00809	0.01400	< 0.00001

Note: Bootstrap significance levels of 0.05 and 0.1 are indicated by ** and * respectively.

Table 5.11: SUR Parameter Estimates, Rural Sector in Southeast Region.

Number of Observations	258
Sum of Weights	384374
Number of Bootstrap Resamples	1000
Beef-Equation R-square	0.5910
Pork-Equation R-square	0.1170
Chicken-Equation R-square	0.6817

Par.	**Variable**	**Parameter Estimate**	**Approx. Bootstr. Std. Err.**	**Approx. Bootstr. 95% LCL**	**Approx. Bootstr. 95% UCL**	**Approx. Bootstr. p-value**
			Beef			
β_{11}	Intercept	0.263960**	0.07682	0.11503	0.41615	0.00054
β_{12}	p_{beef}	-0.006110**	0.00128	-0.00859	-0.00356	$<$ 0.00001
β_{13}	p_{pork}	0.000726	0.00092	-0.00111	0.00248	0.45500
β_{14}	$p_{chicken}$	0.000167	0.00157	-0.00277	0.00399	0.84323
β_{15}	m	0.006337**	0.00084	0.00453	0.00781	$<$ 0.00001
			Pork			
β_{21}	Intercept	1.393350	1.08052	-0.66362	3.57192	0.17837
β_{22}	p_{beef}	0.005656	0.00786	-0.00960	0.02121	0.46039
β_{23}	p_{pork}	-0.049240	0.03940	-0.12808	0.02636	0.19672
β_{24}	$p_{chicken}$	0.016849	0.01744	-0.01664	0.05172	0.31448
β_{25}	m	0.011363*	0.00578	-0.00013	0.02253	0.05269
			Chicken			
β_{31}	Intercept	0.277109**	0.12304	0.01834	0.50067	0.03494
β_{32}	p_{beef}	0.001718	0.00177	-0.00176	0.00518	0.33525
β_{33}	p_{pork}	0.002516	0.00178	-0.00080	0.00619	0.13118
β_{34}	$p_{chicken}$	-0.016720**	0.00286	-0.02242	-0.01122	$<$ 0.00001
β_{35}	m	0.018520**	0.00182	0.00966	0.01678	$<$ 0.00001

Note: Bootstrap significance levels of 0.05 and 0.1 are indicated by ** and * respectively.

CHAPTER VI
CONCLUSION

The general objective of this research was to provide an understanding of the SUR procedure and to explain some of its current trends. To understand SUR this study dedicated Chapter II to explain the estimation procedure, some properties of the SUR estimator, the efficiency gained by the SUR estimator, and how to test for aggregation bias using SUR. With respect to the efficiency gained by the SUR estimator, Zellner (1962) found that the regression coefficient estimators are at least asymptotically more efficient than the least squares equation-by-equation estimators. Specifically, a quite large gain in efficiency can be obtained when independent variables in different equations are not highly correlated and when error terms in different equations are highly correlated. The test for aggregation bias consists of a test for the equality of all regression equation coefficients. Particularly, Zellner's (1962) test can be used to determine if aggregated data (macro-data) has an aggregation bias problem or if disaggregated data (micro-data) can be aggregated without suffering from aggregation bias.

Additional literature of the SUR procedure discussed in this study include SUR with unequal number of observations, the different alternative estimators of the variance-covariance matrix of the error term ($\mathbf{\Sigma}$), the conditions under which one estimator of $\mathbf{\Sigma}$ will perform better than another, and whether it is relevant to use better estimates of $\mathbf{\Sigma}$. SUR with unequal number of observations (i.e., the case where one or more equations have missing observations) focuses on how to handle a set of regression equations when the data is time-series, cross-sectional or panel data. With respect to whether it is relevant to use better estimates of $\mathbf{\Sigma}$, it has been found that better estimates of $\mathbf{\Sigma}$ or $\mathbf{\Sigma}^{-1}$ need not imply better estimates of regression coefficients. Furthermore, a feasible GLS estimator of the regression coefficients that ignores the extra observations in estimating $\mathbf{\Sigma}$ (but not necessarily in estimating $\mathbf{\Sigma}^{-1}$ or $\boldsymbol{\beta}$) compares favorably to a feasible GLS estimator of the regression coefficients that seem to use all extra observations.

Specific objectives of this research were to provide an empirical application of a SUR model, and to explain the relevant findings from this empirical application. The Mexican household meat consumption was selected as the empirical application in this research. To familiarize with the world and Mexican meat markets before estimating SUR models, a discussion of the role meat plays in the agricultural sector was presented in Section 1.1, a review of the meat world market was provided in Section 1.2, and an analysis of Mexican meat production and consumption in Section 1.4. In particular, from Section 1.2, we saw that the United States, the EU-25, Brazil, China, Mexico, and Canada were leading producing and consuming countries of beef, pork, and chicken. Additionally, when we considered the combined production and consumption of beef, pork, and chicken, on average for the period 1997-2006, Mexico was a net meat consumer with excess consumption of 0.859 million MT while Brazil, United States, EU-25, Canada, and China were net meat producers with excess production of 2.808, 2.054, 1.664, 0.956, and 0.218 million MT respectively. Therefore, Mexico is a very important market for all net meat producers.

To analyze the Mexican meat consumption, this study used a nation-wide Mexican survey on household income and weekly expenditures *(Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH))* that is published by a Mexican governmental institution (Instituto Nacional de Estadísitca, Geogría e Informática or INEGI). The data used from ENIGH corresponded to the year 2006 and it was collected between August and November 2006. ENIGH's sampling methods are probabilistic, multi-staged, stratified, and conglomerated. The sampling method is *probabilistic* because the sampling units have a probability of being selected, which is known and different from zero. Additionally, the sampling method is *multi-staged* because the sampling units are selected in multiple stages. It is *stratified* because the target population is divided into groups with similar characteristics, which form the strata. Finally, it is *conglomerated* because the sampling units (households) are made up from the observation units (household members).

Variables of interest were selected from ENIGH database. To organize the data, this study created and modified new variables. Particularly, the variable adult equivalents was created in order to calculate per capita meat consumption and per capita real expenditure. Not adjusting household meat consumption and expenditures by adult equivalents presents a problem when estimating quantity consumed (quantity demand) as a function of prices and total meat expenditure. For example, suppose one household demands q amount of beef and suppose a bigger household who pays a higher price demands more beef. If we compare these two households without adjusting by adult equivalents, price increases but does quantity decrease? On the other hand, adjusting by adult equivalents (i.e., computing per capita beef consumption) in our example, price will always increase but this time quantity will decrease. In addition, nominal variables were transformed to real variables. Then, meat cuts were aggregated into meat categories (Table 4.6) to reduce the excessive number of missing observations resulting from the nature of the survey. To avoid doing price imputations, the number of missing observations was reduced even further by excluding non-relevant meat categories (processed meat and seafood in Table 4.57) from the analysis and considering the total intersection of the non-missing prices in the remaining categories (beef, pork and chicken in Table 4.57). Hence, the results from this study can only be generalized to those households who consumed at least one meat cut of beef, pork and chicken during week of the interview. Since we are dealing with a stratified sample, we know the beef, pork and chicken dataset, which consists of 3,707 households, represents 5.3 million households nationally of the total of 22.1 million households (Table 4.54) who consumed at least one meat cut during the week of the interview.

Since ENIGH is a stratified sample, any descriptive statistic or regression model estimated in this study incorporated the stratification variables (weight and strata). "A data analyst who ignores stratification variables and dependence among observations is not fitting a good model to the data but is simply being lazy" (Lohr, 1999, p. 229). Particularly, this study weighted all the observations by the sampling weight variable as it is done in weighted least squares when estimating the SUR models. In fact, the weighted estimator is consistent (Wooldridge, 2001, p. 464). Specifically, this study used the SAS software to estimate the system of equations using Seemingly Unrelated Regressions as well as the sampling weight of each observation. When "[we] use weights w_i in the weighted least squares estimation, [we] will obtain the same point estimates...; however, in complex surveys, the standard errors and hypothesis tests the software provides will be incorrect and should be ignored" (Lohr, 1999, p. 355). Hence, to calculate better estimates of the standard errors and hypothesis tests for the parameter estimates, this study applied the bootstrap by using SAS software. The Bootstrap is a resampling method that can be used to estimate standard errors where other estimation methods are inappropriate. This approach was prefered over the alternative formulae provided by Wooldridge (2001, p. 464) to calculate asymptotic variances of the parameter estimates because SUR deals with a system of equations instead of one equation.

SUR models were estimated for one general model and for individual models. The general model assumes that regional or urbanization factors shift the consumption of the i^{th} commodity in a parallel fashion. That is, regional and urbanization-level differences in consumption of the i^{th} commodity can be appropriately modeled by parallel shifting sub-models. If this assumption is false and there are differences in consumption of the i^{th} commodity among Mexican regions and urbanization level, the individual SUR models provide more precise parameter estimates for each case. Additionally, in stratified sampling elements of the same stratum often tend to be more similar than randomly selected elements from the whole population; therefore, individuals models were considered.

The results of the SUR general model were reported in Table 5.1. All "own price" parameter estimates (i.e., the corresponding parameter estimates of the price of beef in the beef equation, the price of pork in the pork equation, and the price of chicken in the chicken equation) came with the correct negative sign and statistically different from zero at the 0.05 significance level. Our results indicate that increasing the price of beef by 1 real peso (i.e., 1 Mexican peso expressed in the nominal value of 2002 Mexican pesos), holding all other factors affecting beef consumption constant, will decrease the weekly per capita consumption of beef by 0.006180 kg (or 0.013624 lbs). Similarly, increasing the price of pork by 1 real peso (holding other factors affecting the consumption of pork constant) will decrease the weekly per capita consumption of pork by 0.00475 kg. Finally, increasing the price of chicken by 1 real peso (holding other factors affecting the consumption of chicken constant) will decrease the weekly per capita consumption of chicken by 0.009840 kg. All other price parameter estimates in Table 5.1 resulted insignificant at the 0.1 level. That is, there is not enough statistical evidence to conclude that changes in the price of pork or chicken have an effect on the consumption of beef. In addition, our results indicate that income designated to meat expenditures is statistically different from zero at the 0.05 significant level in all three equations. Hence, increasing the household income designated to meat expenditures by 1 real peso, increases the average weekly per capita consumption of beef by 0.005389 kg, the average weekly per capita consumption of pork by 0.005158 kg, and the average weekly per capita consumption of chicken by 0.007507 kg while holding the price of beef, pork and chicken constant. The estimate of the regional dummy variables and urbanization level dummy variable indicate that there is not enough statistical evidence at the 0.1 significance level to conclude that there are differences between consumption of beef or pork in the Central-West and Southeast regions, the consumption of pork in the Central and Southeast regions, and the urban and rural sectors.

The results of the SUR individual models are reported in Tables 5.2 through 5.11. The parameter estimates corresponding to own prices are all with the correct sign. Similarly, they are all statistically different from zero except for the own price of chicken in Table 5.9 and the own price of pork in Tables 5.10 and 5.11. In the beef equation, the price of pork is always insignificant at the 0.1 level, and the price of chicken is in three occasions significant at the 0.05 or 0.1 level (one occasion positive and significant in Table 5.8, two occasions negative and significant in Tables 5.2 and

5.8). In the pork equation, the price of beef is only in one occasion significant at the 0.1 level (negative and significant in Table 5.5), and the price of chicken is in four occasions significant at the 0.05 or 0.1 level (four occasions negative and significant in Tables 5.2, 5.3, 5.4, and 5.8). In the chicken equation, the price of beef is in two occasions significant at the 0.05 level (two occasions negative and significant in Tables 5.5 and 5.9), the price of pork is in two occasions significant at the 0.05 or 0.1 level (one occasion positive and significant in Table 5.5 and one occasion negative and significant in Table 5.10). Therefore, in the pork and chicken equations, when the price of beef is significant, it is negative. Only in the pork equation, when the price of chicken is significant, it is negative. When significant, price parameter estimates can be interpreted in the same fashion that they were interpreted for the general SUR model.

BIBLIOGRAPHY

A. C. Aitken. On least squares and linear combinations of observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48, November 1934-1935.

Javier Usabiaga Arroyo. Desarrollo rural en México. Technical report, Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA), 2002. URL `www.sagarpa.gob.mx/desarrollorural/publicaciones/files/ponencias/desrur_mx.pdf`. Accessed on May 28, 2008.

B. H. Baltagi. Pooling: An experimental study of alternative testing and estimation procedures in a two-way error component model. *Journal of Econometrics*, 17: 21–49, 1981.

Badi H. Baltagi, Susan Garvin, and Stephen Kerman. Further Monte Carlo evidence on seemingly unrelated regressions with unequal number of observations. *Annales d'Économie et de Statistique*, 14:103–115, 1989.

Laura Elena de Moral Barrera, Brenda Patricia Ramírez Gómez, and Alma Rosa Muñoz Jumilla. Crecimiento regional de la producción de la carne de cerdo en México, 1980-2005. Technical report, 2008. URL `http://www.analisiseconomico.com.mx/pdf/5215.pdf`. Accessed on May 28, 2008.

P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9:1196–1217, 1981.

J. Blaylock. The impact of equivalence scales on the analysis of income and food spending distributions. *Western Journal of Agricultural Economics*, 16:11–20, 1991.

J. Blokland. *Continuous Consumer Equivalence Scales*. The Hague: Martinas Nijhoff, 1976. ISBN 9-024-71847-3.

Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. New York: Springer-Verlag, Inc., 1987. ISBN 0-387-96406-1.

J. A. Brown and A. S. Deaton. Surveys in applied economics: Models of consumer behaviour. *Economic Journal*, 82:1145–1236, 1972.

Keith C. Brown and Rao K. Kadiyala. The estimation of missing observations in related time series data: Further results. *Communications in Statistics - Simulation and Computation*, 14(4):973–981, 1985.

D. Brownstone and C. Kazimi. Applying the Bootstrap, manuscript, Department of Economics, University of California, Irvine, 1998.

R. C. Buse and L. E. Salathe. *American Journal of Agricultural Economics*, 60: 460–468, 1978.

A. Colin Cameron and Pravin K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press, 2005. ISBN 0-521-84805-9.

Ronald Christensen. *Plane Answer to Complex Questions: The Theory of Linear Models*. New York: Springer-Verlag, Inc., third edition, 2002. ISBN 0-387-95361-2.

Harald Cramér. *Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946.

Angus Deaton. Quality, quantity and spatial price variation. *American Economic Review*, 78:418–430, 1988.

Angus Deaton and John Muellbauer. *Economic and Consumer Behaviour*. Cambridge: Cambridge University Press, 1980. ISBN 0-521-22850-6.

Angus Deaton and John Muellbauer. On measuring child costs: With applications to poor countries. *Journal of Political Economy*, 94(4):720–744, August 1986.

P. M. Dixon. The bootstrap and the jacknife: Describing the precision of ecological indices. In S. M. Scheiner and J. Gurevitch, editors, *Design and Analysis of Ecological Experiments*, pages 290–318. New York: Chapman & Hall, 1993. ISBN 0-195-13187-8.

D. Dong, J. S. Shonkwiler, and Oral Capps. Estimation of demand functions using cross-sectional household data: The problem revisited. *American Journal of Agricultural Economics*, 40:466–473, August 1998.

E. M. Duvall. *Marriage and Family Development*. Philadelphia: J. B. Lippincott Company, 1977. ISBN 0-397-47362-1.

John H. Dyck and Kenneth E. Nelson. Structure of the global markets for meat. Washington, D.C.: U.S. Department of Agriculture, Economic Research Service, Market and Trade Economics Division, Agriculture Information Bulletin No. 785, 2003.

Bradley Efron. Bootstrap methods: Another look at the jacknife. *Annals of Statistics*, 7:1–26, 1979.

Bradley Efron. The jacknife, the bootstrap, and other resampling plans. Philadelphia: SIAM, 1982.

Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. London: Chapman & Hall, 1993. ISBN 0-412-04231-2.

Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH). Aguascalientes, México: Instituto Nacional de Estadística, Geografía e Informática (INEGI). URL `www.inegi.gob.mx`.

Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)—Síntesis Methodológica. Aguascalientes, México: Instituto Nacional de Estadística, Geografía e Informática (INEGI), 2006. URL `www.inegi.gob.mx`.

Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH)—Síntesis Methodológica. Aguascalientes, México: Instituto Nacional de Estadística, Geografía e Informática (INEGI), 2006. URL `www.inegi.gob.mx`.

E. F. Fama, L. Fisher, M. Jensen, and R. Roll. The adjustment of stock prices to new information. *International Economic Review*, 10:1–21, 1969.

R. E. Fay. Alternative paradigms for the analysis of imputed survey data. *Journal of the American Statistical Association*, 91:490–498, 1996.

Denzil G. Fiebig and Jae H. Kim. Estimation and inference in SUR models when the number of equations is large. *Econometric Reviews*, 19(1):105–130, 2000.

D. A. Freedman. On bootstrapping two-stage least-squares estimates in stationary linear models. *Annals of Statistics*, 3:827–842, 1984.

Brian W. Gould and Hector J. Villarreal. Adult equivalence scales and food expenditures: An application to mexican beef and pork purchases. *Applied Economics*, 34(9):1075–1088, 2002.

William E. Griffiths, Carter R. Hill, and George G. Judge. *Learning and Practicing Econometrics*. New York: John Wiley & Sons, Inc., 1992. ISBN 0-471-51364-4.

S. Gross. 1980. pages 181–184, 1980.

P. Hall. *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag, Inc., 1992. ISBN 0-387-97720-1.

M. H. Hansen and W. N. Hurwitz. The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41:517–529, 1946.

David A. Harville. *Matrix Algebra from a Statistician's Perspective*. New York: Springer-Verlag, Inc., 1997. ISBN 0-387-94978-X.

J. A. Hausman and D. A. Wise. Stratification on an endogenous variable and estimation: The Gary income maintenance experiment. In C. F. Manski and D. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 365–391. Cambridge, Massachusetts: MIT Press, 1981. ISBN 0-262-13159-5.

J. S. U. Hjorth. *Computer Intensive Statistical Methods*. London: Chapman & Hall, 1994. ISBN 0-412-49160-5.

R. R. Hocking and W. B. Smith. Estimation of parameters in the multivariate normal distribution with missing observations. *Journal of the American Statistical Association*, 63:159–173, March 1968.

J. L. Horowitz. The bootstrap. In J. J. Heckman and E. Leamer, editors, *Handbook of Econometrics*. Volume 5, 3159-3228, Amsterdam, North-Holland, 2001. ISBN 0-444-53200-5.

Hae-Shin Hwang. Estimation of linear SUR model with unequal numbers of observations. Mimeo, 1987. Texas A&M University.

Hae-Shin Hwang. Estimation of a linear SUR model with unequal numbers of observations. *The Review of Economics and Statistics*, 72(3):510–515, August 1990.

G. W. Imbens and T. Lancaster. Efficient estimation and stratified sampling. *Journal of Econometrics*, 74:289–318, 1996.

International Monetary Fund–World Economic Outlook Database. URL `http://www.imf.org/external/pubs/ft/weo/2008/01/weodata/weoselgr.aspx`. Accessed on May 12, 2008.

T. Kloek. Note on convenient matrix notations in multivariate statistical analysis and in the theory of linear aggregation. pages 1–8, August 1960. Report 6017 of the Econometric Institute of the Netherlands School of Economics.

J. Kmenta. *Elements of Econometrics*. New York: Macmillan, 1971. ISBN 0-023-65060-5.

J. Kmenta and R. F. Gilbert. Small sample properties of alternative estimators of seemingly unrelated regressions. *Journal of the American Statistical Association*, 63:1180–1200, 1968.

E. P. Lazear and R. T. Michael. Family size and the distribution of real per capita income. *American Economic Review*, 70:91–107, 1980.

D. C. Levinson, C. N. Darrow, E. B. Klein, M. H. Levinson, and B. McKee. *The Seasons of a Man's Life*. New York: Ballantine Books, 1978. ISBN 0-394-40694-X.

R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, Inc., 1987. ISBN 0-471-80254-9.

Sharon L. Lohr. *Sampling: Design and Analysis*. New York: Duxbury Press, 1999. ISBN 0-534-35361-4.

J. G. Mackinnon. Bootstrap inference in econometrics. *Canadian Journal of Economics*, 35:615–645, 2002.

G. S. Maddala and T. D. Mount. A comparative study of alternative estimators for variance components models used in econometric applications. *Journal of the American Statistical Association*, 68(342):324–328, June 1973.

J. Lastra Marín, María A. Peralta Arias, Luis Villamar, Cesar Segura, Marco A. Barrera, Herminio Guzmán, and Ramón Domímguez. La producción de carnes en México y sus perspectivas 1990-2000. Technical report, Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA), 2000. URL `http://www.sagarpa.gob.mx/Dgg/estudio/carne.pdf`. Accessed on March 16, 2007.

Microsoft Encarta Online Encyclopedia. European Union. Copyright 1997-2008 Microsoft Corporation. All Rights Reserved, 2008. URL `http://encarta.msn.com`. Accessed on May 16, 2008.

Alexander McFarlane Mood. *Introduction to the Theory of Statistics*. New York: McGraw-Hill Book Company, Inc., first edition, 1950.

J. A. Nelson. Quality variation and quantity qggregation in consumer demand for food. *American Journal of Agricultural Economics*, 73:1204–1212, 1991.

Northeast-Midwest Institute. URL `http://www.nemw.org/gdp1.htm`. Accessed on May 12, 2008.

R. W. Parks. Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. *Journal of the American Statistical Association*, 62:500–509, 1967.

C. F. Perali. *Consumption, Demographics and Welfare Measurement: Metric and Policy Implications to Columbia*. PhD thesis, Department of Agricultural and Applied Economics, University of Wisconsin-Madison, 1993.

Robert S. Pindyck and Daniel L. Rubinfeld. *Econometric Models and Economic Forecasts*. Massachusetts: Irwin McGraw-Hill, Inc., fourth edition, 1997. ISBN 0-07-913292-8.

J. N. K. Rao. On variance estimation with imputed survey data. *Journal of the American Statistical Association*, 91:499–506, 1996.

J. N. K. Rao and C. F. J. Wu. Resampling inference with complex survey data. *Journal of American Statistical Association*, 83:231–241, 1988.

S. N. Roy. *Some Aspects of Multivariate Analysis*. New York: John Wiley & Sons, Inc., 1957.

Ricardo Sabates, Brian W. Gould, and Hector Villarreal. Household composition and food expenditures: A cross-country comparison. *Food Policy*, 26:571–586, July 2001.

SAS Institute Inc. Sample 24982: Jacknife and bootstrap analyses. URL http://support.sas.com/kb/24/982.html. Accessed on July 1, 2008.

Peter Schmidt. Estimation of seemingly unrelated regressions with unequal numbers of observations. *Journal of Econometrics*, 5:365–377, 1977.

J. Shao and D. Tu. *The Jacknife and the Bootstrap*. Springer-Verlag, Inc., 1995. ISBN 0-387-94515-6.

V. K. Sharma. Estimation of seemingly unrelated regressions with unequal numbers of observation. *Sankhyā*, Series B, 55:135–138, 1993.

K. Singh. On the asymptotic accuracy of efron's bootstrap. *Annals of Statistics*, 9: 1187–1195, 1981.

Sistema de Información Agropecuaria de Consulta (SIACON). Servicio de Información Agroalimentaria y Pesquera (SIAP), Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA). URL http://www.sagarpa.gob.mx/Dgg/indexave.htm. Accessed on March 17, 2007.

Sistema de Información Agropecuaria de Consulta (SIACON). Servicio de Información Agroalimentaria y Pesquera (SIAP), Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA). Comportamiento del gasto de los hogares en alimentos en el sector rural 2000-2004, September 2006. URL www.siap.sagarpa.gob.mx/Integracion/EstadisticaDerivada/IndicadoresEconomicos/IndMacroeconomicos/Gastos.pdf. Accessed on May 28, 2008.

R. Sitne. An introduction to bootstrap methods: Examples and ideas. 18:243–291, 1990.

R. R. Sitter. Comparing three bootstrap methods for survey data. *Canadian Journal of Statistics*, 20:135–154, 1992.

J. N. Srivastava and M. K. Zaatar. A Monte Carlo comparison of four estimators of dispersion matrix of a bivariate normal population, using incomplete data. *Journal of the American Statistical Association*, 68:180–183, March 1973.

Virendra K. Srivastava and David E. A. Giles. *Seemingly Unrelated Regression Equations Models*. New York: Marcel Dekker, Inc., 1987. ISBN 0-8247-7610-0.

W. E. Taylor. Small sample considerations in estimation from panel data. *Journal of Econometrics*, 13:203–223, 1980.

John R. Tedford, Oral Capps, and Joseph Havlicek. Adult equivalent scales once more: A developmental approach. *American Journal of Agricultural Economics*, 68(2):322–333, May 1986.

Lester G. Telser. Iterative estimation of a set of linear regression equations. *Journal of the American Statistical Association*, 59(307):845–862, September 1964.

The World Bank and International Monetary Fund. Mexico: Financial assessment program update—technical note—financing of the private sector, November, 2006. URL `http://www.imf.org/external/pubs/ft/scr/2007/cr07170.pdf`. Accessed on May 12, 2008.

H. Theil. Qualities, prices and budget inquiries. *Review of Economic Studies*, 19(3): 129–147, 1952–1953.

H. Theil. *Linear Aggregation of Economic Relations*. Netherlands: North-Holland Publishing Company, 1954.

H. Theil. *Principles of Econometrics*. New York: John Wiley & Sons, Inc., 1971. ISBN 0-471-85845-5.

United States Department of Agriculture. Economic Research Service (ERS), Production, Supply and Distribution (PSD) online database. URL `http://www.fas.usda.gov/psdonline/psdQuery.aspx`. Accessed on March 17, 2007.

J. W. Vander Zanden. *Human Development*. New York: Alfred A. Knopf Co., 1978. ISBN 0-394-32103-0.

T. J. Wales and A. D. Woodland. Sample selectivity and the estimation of labour supply functions. *International Economic Review*, 21:437–468, 1980.

H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48:817–838, 1980.

Wikipedia. Zonas económicas de México—Wikipedia, The Free Encyclopedia, 2008. URL `http://es.wikipedia.org/w/index.php?title=Zonas_Econ%C3%B3micas_d_M*%C3%A9xico&oldid=15356206`. Accessed on May 28, 2008.

S. S. Wilks. Moments and distributions of estimates of population parameters from fragmentary samples. *Annals of Mathematical Statistics*, 3:167–195, 1932.

S. S. Wilks. *Mathematical Statistics*. New Jersey: Princeton University Press, 1943.

Jeffrey M. Wooldridge. Asymptotic properties of weighted M-estimators for variable probability samples. *Econometrica*, 67(6):1385–1406, November 1999.

Jeffrey M. Wooldridge. Asymptotic properties of weighted M-estimators for standard stratified samples. *Econometric Theory*, 17(2):451–470, April 2001.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: MIT Press, 2002. ISBN 0-262-23219-7.

Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach*. New York: Thomson South-Western, third edition, 2006. ISBN 0-324-28978-2.

Arnold Zellner. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57 (298):348–368, June 1962.

Arnold Zellner. Estimators for seemingly unrelated regression equations: Some exact finite sample results. *Journal of the American Statistical Association*, 58(304): 977–992, December 1963.

Arnold Zellner. On the aggregation problem: A new approach to an old problem. In K. et al. Fox, editor, *Economic models, Estimation, and Risk Programing: Essays in Honor of Gerhard Tintner*, pages 362–374. Berlin: Springer-Verlag, Inc., 1969.

Arnold Zellner. *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley & Sons, Inc., 1971. ISBN 0-471-98165-6.

Arnold Zellner and David S. Hwang. Further properties of efficient estimators for seemingly unrelated regression. *International Economic Review*, 3(3):300–313, September 1962.

Arnold Zellner and C. Montmarquette. A study of some aspects of temporal aggregation problems in econometric analyses. *Review of Economics and Statistics*, pages 335–342, 1971.

Francisco Morales Zepeda. *El Impacto de la Biotecnología en la Formación de Redes Institucionales en el Sector Hortofrutícola de Sinaloa, México*. PhD thesis, 2007. URL `www.tdr.cesca.es/TESIS_UB/AVAILABLE/TDX-1003107-110809//14.FMS_ANEXOS.pdf`. Accessed on May 28, 2008.