Principal Component Analysis of State Level Food System Indicators

A THESIS

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Benjamin Paul Scharadin

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Robert King

February 2012

Acknowledgements

I would like to use this page to acknowledge and thank all the people who contributed to the completion of this thesis.

First and foremost I would like to thank all the professors who served on my thesis committee.

> The endless guidance and support from Professor King throughout the entire process was incredibly encouraging. I would also like to thank Professor King for giving me the opportunity to work on the project, even without prior experience in Principal Component Analysis. The completion or even the inception of this thesis would not have occurred without him.
>
> I would like to thank Professor Anderson for her comments and critiques that strengthened the final version of the thesis, as well as for her flexibility in participating in the thesis defense and in completing the necessary paperwork.
>
> Finally I would like to thank Professor Davis, for her comments and critiques that further strengthened the final thesis and her well timed warnings of necessary documents.

Secondly I would like to thank all my friends and family who are a crucial support network for accomplishing any major goal.

> My mom has been there for me in every step of my life. Her constant belief that I can accomplish bigger and better things every step of the way provides a drive that is necessary. Her support in every way imaginable is equally important to my accomplishments, as the work I contribute.
>
> My girlfriend Dianna was a savior during the thesis process. Simple actions like kindly interrupting me for a timely break, making actual dinners when I was too busy to cook, and giving me a kick in the pants to start working again ensured sanity and enjoyment every day.
>
> My sister gave me the initial inspiration to pursue graduate school. Without saying any words, but by just simply attending herself, she instilled a healthy level of competitive inspiration that allowed me to achieve more than I would have without.

Dedication

I dedicate this thesis to my mom. Her life has been spent caring for and caring about others. That lifelong care and support for me makes a part of this thesis hers as well.

Abstract

The food system's interconnectivity with almost every aspect of society makes accurately characterizing it very important. This same interconnectivity also makes the problem of accurately characterizing the food system very complex. Indicator sets that attempt to capture the holistic nature of the food system and are repeated across location and time to allow for comparisons and stability testing are inevitably very large. Using a large data set of state-level food system indicators collected for 1997, 2002, and 2007, this thesis explores the possibility of using Principal Component Analysis to develop summary measures for groups of indicators. The results show that it is possible to characterize the information presented by groups of individual indicators by component scores, although the process is very difficult. Through Principal Component Analysis and Partial Common Principal Component Analysis techniques, selected groups of indicators for each state over the three years are reduced in dimensionality and shown to be stable over time. This then allows for states to be compared nationally, regionally, and temporally on specific aspects of their food systems.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Environmental issues, such as climate change, water quality, and land use change; health issues, such as increased obesity and increased diabetes rates; and economic issues, such as the affordability of necessities and fair wages, are major concerns in American public, political, and scientific debate. All are affected by the structure and performance of the food system. The common definition of a food system is the "foundations for food production, the social aspects of consumption, and relevant government and other policies, as well as the actual growing, processing, and distributing of substances that results in foods that people consume."(Gillespie 2000, p. 2) However, this loose definition does not convey how pervasive and important the food system's impacts are. The food system is heavily embedded in the economic, health, and environmental aspects of our lives.

The extent to which the food system affects personal, state, and country level economic activities is not always apparent, however it plays a large role in each. Expenditure on food was the third highest category of expenditure in 2010, accounting for 12.5% of all average annual household expenditures (BLS 2010). Expenditure on food was only behind expenditure on other necessities, housing and transportation. On a larger scale, the food system employs a major percent of the work force and accounts for a large portion of the United States' GDP. The food system is also important to the United States' international trade, with agricultural exports being one of the only sectors to have a consistent and significant trade surplus (Jerardo 2004). Agricultural exports become even more important, on the international stage, when the number of people fed is considered. These few examples show the importance and size of the food system from both macro and micro economic perspectives.

Food consumption patterns have a strong link with health patterns, so much so that the food system is sometimes referred to as the “food and nutrition system” (Sobal 1988). Overall health in America has declined due to increasingly unhealthy consumption patterns. In the United States fruit and vegetable consumption have remained steady over time, however fast food and high sugar food consumption have increased (Casagrande 2007). The shift from nutrient rich foods to high calorie, low nutrient foods is correlated with negative health trends. In 2011 the American Diabetes Association estimated that 8.3% of the U.S. population had diabetes, with an additional 25% of the population classified as “pre-diabetes”. The average obesity rate for the United States in the 1980’s was 15% (United States CDC 2011). That rate has more than doubled, with a current average obesity rate of 33.8%. These health conditions impose a physical cost on the individual, as well as health care costs. In 2008, health costs associated with being obese were estimated to be $147 billion (Finklestein 2009). While lack of exercise also plays a role in causing these problems, the trends highlight the strong link between the consumption choices in the food system and health.

The food system also has significant effects on the environment. These can remain local or can accumulate and have impacts large distances from the cause. The common herbicide Atrazine is known to have detrimental effects to plants, wildlife, as well as humans if above the allowable concentration thresholds. A report published in 2010 showed that 66 of 153 drinking water systems sampled in the Mississippi River Basin exceeded the allowable level for Atrazine in the treated drinking water (Wu and Mae 2010). Downstream accumulation of nitrogen from fertilizers is also a major environmental issue. The accumulation of nitrogen in the Gulf of Mexico feeds massive algae blooms, which significantly lower the oxygen level in the surrounding water. The resulting “Dead Zone” ranges from 5,000 to 8,000 square miles and degrades one of the world’s most fertile fisheries (Booth 2006). When considering global warming, cattle production is the source of nearly 18% of greenhouse gas “CO2 equivalent” emissions (UNFAO 2006). The food system not only affects these environmental issues, but is also affected by them. Climate change is predicted to alter growing periods and regions and the degradation of water quality has the potential to decrease the viability of fisheries.

Due to the importance of the food system in the United States many sets of indicators have been developed to track its condition. An indicator is described as “a way to measure, indicate or point to with more or less exactness,” or “something used to show the condition of a system.”(Feenstra 2005, p 16.1). In 2005, the Vivid Picture Project published a report with 63 proposed indicators for tracking a sustainable food system from a list of over 125 (Feenstra 2005). In 2009, a list of 19 suggested indicators was proposed to conform to a healthy, fair, green, and affordable construct of the food system (Anderson 2009). More recently, in 2010, the United States Department of Agriculture’s Economic

Research Service launched the Food Environment Atlas, which is a compilation of 168 indicators in 13 different categories. There was also an international effort by the Organization for Economic Co-operation and Development in defining and collecting indicators for multiple countries (OECD 2001).

Although, extensive work has been done collecting indicators for numerous food system indicator projects, the lack of consistency across projects makes it difficult to compare indicators between regions or over time. National level indicators are unable to express variation across regions, while small scale indicator projects are often developed for only a single location. Both of these frameworks do not allow for comparison of the food system across regions. In addition, both large and small scale indicator projects have rarely allowed for repeated collection over time. With only a cross-sectional set of indicators, temporal changes in the food system cannot be analyzed. State level indicators gathered over multiple years would help address both of these problems. They would allow comparisons across states, as well as over time.

A multidisciplinary project team funded by the Healthy Food Healthy Lives Institute at the University of Minnesota has collected data for a set of state level food system indicators over multiple years. The research group sought to fill the gap in past indicator projects that either focused on a small area or the national level. Having multiple indicators from multiple geographic locations allows for the conceptual framework of the indicators to be tested. System-wide indicators, from input supply to post-consumption waste management, in five areas of interest; descriptive, health, social, environmental, and economic, have been collected. This indicator set has a large number of variables because of the complexity of the food system. If such indicator sets are to be useful in understanding changes in the food system over time and different regions, the data must be reduced while maintaining the important information captured by all the indicators.

Principal Component Analysis (PCA) is a commonly used statistical technique for data reduction and exploratory structure analysis. PCA is most commonly used in biology and psychology but has been used in many other research fields because of its wide applicability. PCA groups variables in a way that highlights their similarities and differences, while losing as little of the original information as possible. Each original variable is given a particular weight and loaded onto a component. The component represents a particular attribute, which is defined by the types of variables that load on it. The very high dimensionality of the food system indicators can be reduced using PCA. With a lower dimensionality, comparison between states and over time will be much easier. In addition, PCA exposes underlying structure in the data because no restrictions are placed on the components.

With component structures for multiple years, testing their stability over time is also important. Common Principal Component Analysis (CPCA) is a statistical technique used to test the level of similarity of component structures between different PCA groups. In biology and psychology research, different groups are commonly defined by species or gender. In economic research, groups are commonly defined by different years. CPCA uses the component weightings for each year determined through PCA and determines if they are statistically different from one another. If they are shown to not be statistically different, the component structure is stable over time. This allows for multiple years to be pooled for analysis, and a single component structure to be used for component score production. Using a single component structure for component score production allows for direct comparison of component scores over time.

The overall objective of the thesis is to explore the potential for effective data reduction in this set of state level food system indicators for all 50 states and to test the stability of the component structure over time. In order to accomplish this goal the thesis has three specific objectives:

1. Reduce the dimensionality of the state level indicators over multiple years, while exploring if there are interpretable and meaningful underlying component structures to the indicators.
2. Compare the structures over time to test for similarity. If the structures are shown to be stable, data for new time periods can be described using the current structures.
3. Compare the states across regions and over time using the structure found.

The remainder of the thesis will begin with a data chapter that discusses past indicator projects, the indicators gathered by the Healthy Food Healthy Lives research group, how the indicators are presented, and which were used in data reduction. A methods chapter explains how the data were prepared for PCA, describes the PCA process, and the partial Common Principal Components

Analysis (CPCA) technique that is used to test the stability of component structures over time. A results chapter presents and discusses the results of the PCA, partial CPCA, and the national, regional, and temporal comparisons. Finally, the concluding chapter summarizes the results, addresses challenges of the process, and discusses future applications.

Chapter 2: Conceptual Framework and Data

Review of Food System Indicators

Many different sets of food system indicators have been developed. While a number of past indicator projects are discussed in this section, there are many that are not. Small local groups have developed indicators for their particular county or community, while other publicly or privately funded groups have developed national and international indicators. These indicators also vary depending on the scope of their assessment. Some cover many aspects and activities of the food system, while others focus solely on one characteristic or just a few activities within the entire system.

Most indicator projects have been on the two ends of the scope spectrum, with the focus either being at the county level, or smaller, or the national level. There is a belief that a focus on local food system vitality is a more appropriate way to analyze food systems (Feenstra 2005). With a small physical area of focus, unique aspects of that area's food system can be uncovered, such as food deserts and pockets of high poverty. Also, these indicator sets can relate directly to local policy initiatives. Numerous studies focus on local food systems, as a way to explain trends in their immediate area (Gradwell 2002, King and Feenstra 2001). Other studies suggest that "community food system" is the appropriate scale for indicator projects. Gillespie defines a community food system as, "a part of a larger food system that is geographically located in a community."(Gillespie 2000, p. 5). The community may have boundaries very different from county boundaries. A final reduction in scope, to the individual farm level, is also proposed by Martinez because he believes public policy decisions have the greatest effect at the farm level (Martinez 2001).

On the other end of the spectrum national level indicator sets have been developed for international comparisons, usually with a particular focus on sustainability. The Organization for Economic Co-operation and Development published a four part indicator report on environmental indicators pertaining to agriculture (OECD 2001). The United Nations' Food and Agriculture Organization is also developing a sustainability indicator set (Guttenstein 2010). The indicators are divided into core and supplementary sets with stronger emphasis placed on the core indicators. Targets are then set for certain indicators. In addition to international organization efforts, individual countries have compiled indicators sets. Canada developed an indicator set that stresses the "triple bottom line" of agriculture (Anstey 2010). The "triple bottom line" is defined as earning an economic profit, while having positive effects on social and environmental welfare. For each set indicators were collected on a national level, resulting in a single indicator on a particular issue for very large nations. National level indicators are important for negotiating international agreements, where specific details of a region of a country are not of interest.

There are major drawbacks to indicator sets on both the county and national level. Small geographic scale indicator sets, with their focus on local issues and conditions, may not be consistent across projects. Without the same set of indicators being collected in multiple locations, comparisons across regions are impossible. Furthermore, national level indicator sets cannot accurately account for large variability among different regions of a country. These indicators can lead to deceiving conclusions by considering an entire country to be homogenous. For these reasons this thesis focuses on state level indicators. State level indicators can be considered to be too low resolution to capture community level issues, such as food desserts, or too high resolution to capture regional issues, such as watersheds, however they have important benefits. State level indicators describe a meaningful well defined area, are useful for policy-makers at both the state and federal level, and are easy to keep consistent. With the same set of indicators being collected across all 50 states, a direct comparison can be made between them. In addition, a state level indicator set begins to fill the gap between the two scopes used for other indicator sets.

Other food system indicator projects analyze a singular issue affected by the food system. Many of the studies focus on environmental effects and develop different sets of agri-environmental indicators. The previously mentioned Organization for Economic Cooperation and Development is an example (OECD 2001). The four categories of indicators are for: (1) agriculture in the broader economic, social, and environmental context; (2) farm management practices; (3) intensity of farm inputs and natural resource use; and (4) environmental impacts of agriculture. Focusing on one aspect of the food system allows for analysis of that single aspect to be very detailed.

There is a major drawback to this approach to indicator collection, however. The holistic nature of the food system is not conveyed when one piece is partitioned. Analyzing a single part of a system is not appropriate because each part of the system has an effect on and is affected by the other parts. By only analyzing economic performance, for example, a food system may seem very healthy. When economic performance is considered in conjunction with other aspects, such as environmental performance, the food system may seem less healthy.

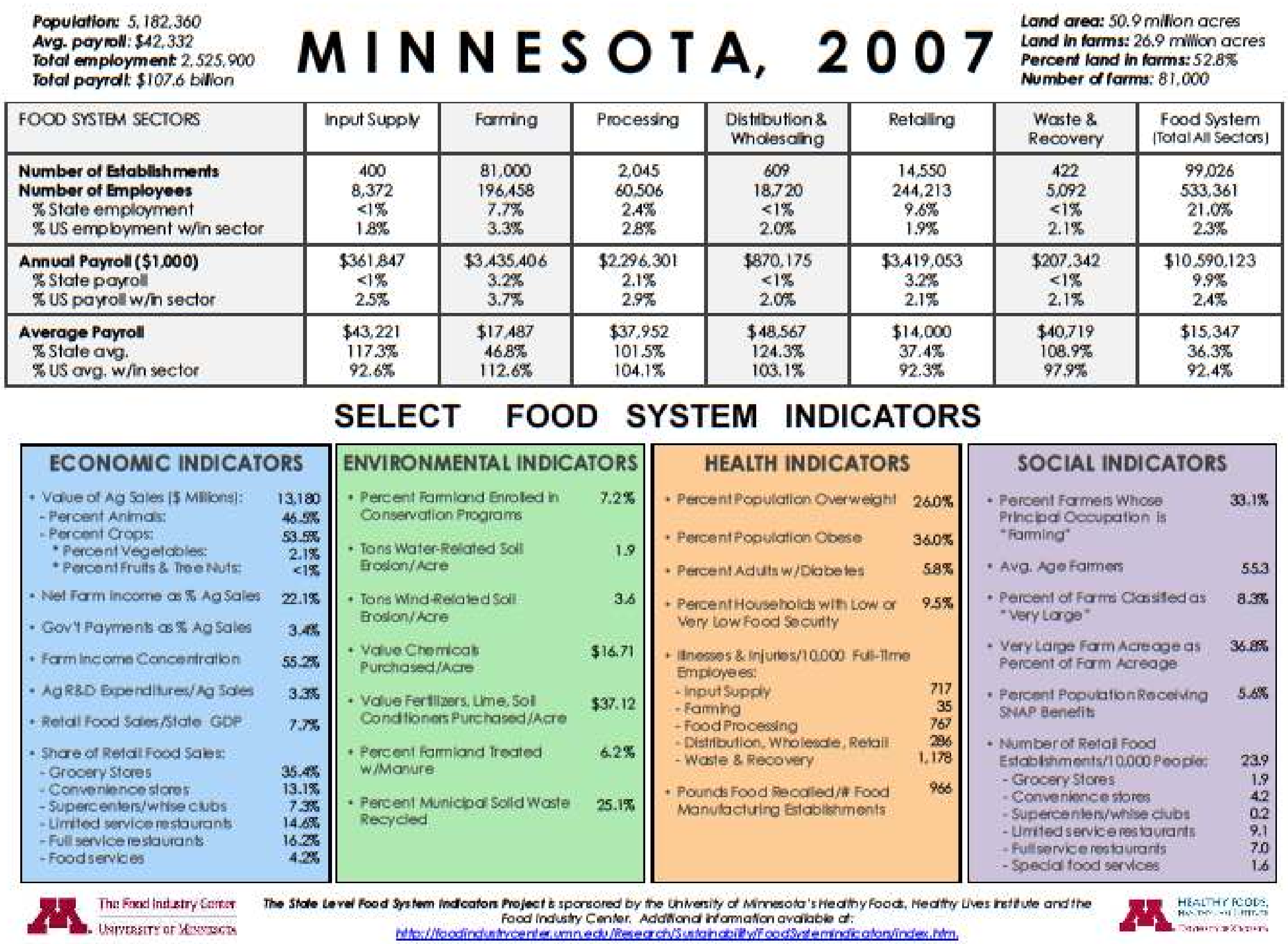
There are two recent note-worthy examples of holistic indicator projects, the Vivid Picture Project for the state of California's food system (Feenstra 2005) and the Charting Growth to Good Food project by the WK Kellogg Foundation (Anderson 2009). The Vivid Picture project tracked food system trends in order to progress towards a sustainable food system. The indicators were selected in accordance with the specific goals of the Roots of Change group and included measures for all food system activity areas, including input supply, primary production, processing and distribution, retailing, consumption, and waste. The indicators also spanned into environmental, social, health, and affordability issues. The Charting Growth to Good Food study developed a similar framework, but was more narrowly defined. The goal of the project was to define "healthy, green, fair, and affordable" as food attributes and collect credible national indicators for the United States food system. These indicators would then be used to assess the availability for "good" food in the United States. The overall indicators selected by the Healthy Food Healthy Lives research group fit within a framework that combines insights from these two projects. Indicators were collected at the state level for each of the six activity areas of the food system in categories of health, environmental, social, and economic.

The United States Department of Agriculture's Economic Research Service has compiled an extensive set of food system indicators. In November 2011, the Food Environment Atlas (FEA) listed 168 indicators at the state and county level in the areas of food store availability, restaurant availability, food insecurity, physical activity levels, health trends, food tax rates, and community characteristics. The FEA also has multiple years gathered for it indicators. In order to compare across states and counties, the FEA can create county or state level maps for a single indicator. All indicators gathered for one state can also be viewed; however this can only be done for a single state. The FEA has a large breadth of information over the entire nation and over multiple years. However, comparison between states and across years is complex because of the vast number of indicators and limited comparison tools.

Conceptual Framework and Structure

As noted in the introduction, the data reduction and exploratory analysis of the food system indicators is a portion of a larger project being conducted by the Healthy Foods Healthy Lives Institute at the University of Minnesota. With the overall goal of the project in mind, a large number of indicators were collected for the entire food system. Conceptually, the food system was segmented into seven activity areas: input supply, primary production, processing, distribution and wholesaling, retail, consumption, and waste. Within each of these categories indicators were collected for five general aspects of food system structure and performance: descriptive, economic, environmental, health, and social. Each indicator was collected, when available, for 1997, 2002, and 2007. These years were chosen because of the large availability of indicators associated with the USDA Agricultural Census and U.S. Census Bureau's Economic census, which are conducted in these years.

For ease of presentation the Healthy Food Healthy Lives research group developed state fact sheets to present the indicators. The fact sheets present demographic information, descriptive indicators about the food system, and indicators in the four categories of economic, environmental, health, and social for each state for a single year. A sample state fact sheet is in Figure 1.1. Even in a compact form like the state fact sheet, comparisons across states and over time are complex.



Population: 5,182,360
Avg. payroll: $42,332
Total employment: 2,525,900
Total payroll: $107.6 billion

MINNESOTA, 2007

Land area: 50.9 million acres
Land in farms: 26.9 million acres
Percent land in farms: 52.8%
Number of farms: 81,000

FOOD SYSTEM SECTORS	Input Supply	Farming	Processing	Distribution & Wholesaling	Retailing	Waste & Recovery	Food System (Total All Sectors)
Number of Establishments	400	81,000	2,045	609	14,550	422	99,026
Number of Employees	8,372	196,458	60,506	18,720	244,213	5,092	533,361
% State employment	<1%	7.7%	2.4%	<1%	9.6%	<1%	21.0%
% US employment w/in sector	1.8%	3.3%	2.8%	2.0%	1.9%	2.1%	2.3%
Annual Payroll ($1,000)	$361,847	$3,435,406	$2,296,301	$870,175	$3,419,053	$207,342	$10,590,123
% State payroll	<1%	3.2%	2.1%	<1%	3.2%	<1%	9.9%
% US payroll w/in sector	2.5%	3.7%	2.9%	2.0%	2.1%	2.1%	2.4%
Average Payroll	$43,221	$17,487	$37,952	$48,567	$14,000	$40,719	$15,347
% State avg.	117.3%	46.8%	101.5%	124.3%	37.4%	108.9%	36.3%
% US avg. w/in sector	92.6%	112.6%	104.1%	103.1%	92.3%	97.9%	92.4%

SELECT FOOD SYSTEM INDICATORS

ECONOMIC INDICATORS

• Value of Ag Sales ($ Millions):	13,180
- Percent Animals:	46.5%
- Percent Crops:	53.5%
* Percent Vegetables:	2.1%
* Percent Fruits & Tree Nuts:	<1%
• Net Farm Income as % Ag Sales	22.1%
• Gov't Payments as % Ag Sales	3.4%
• Farm Income Concentration	55.2%
• Ag R&D Expenditures/Ag Sales	3.3%
• Retail Food Sales/State GDP	7.7%
• Share of Retail Food Sales:	
- Grocery Stores	35.4%
- Convenience stores	13.1%
- Supercenters/whse clubs	7.3%
- Limited service restaurants	14.6%
- Full service restaurants	16.2%
- Food services	4.2%

ENVIRONMENTAL INDICATORS

• Percent Farmland Enrolled in Conservation Programs	7.2%
• Tons Water-Related Soil Erosion/Acre	1.9
• Tons Wind-Related Soil Erosion/Acre	3.6
• Value Chemicals Purchased/Acre	$16.71
• Value Fertilizers, Lime, Soil Conditioners Purchased/Acre	$37.12
• Percent Farmland Treated w/Manure	6.2%
• Percent Municipal Solid Waste Recycled	25.1%

HEALTH INDICATORS

• Percent Population Overweight	26.0%
• Percent Population Obese	36.0%
• Percent Adults w/Diabetes	5.8%
• Percent Households with Low or Very Low Food Security	9.5%
• Illnesses & Injuries/10,000 Full-Time Employees:	
- Input Supply	717
- Farming	35
- Food Processing	767
- Distribution, Wholesale, Retail	286
- Waste & Recovery	1,178
• Pounds Food Recalled/# Food Manufacturing Establishments	966

SOCIAL INDICATORS

• Percent Farmers Whose Principal Occupation is "Farming"	33.1%
• Avg. Age Farmers	55.3
• Percent of Farms Classified as "Very Large"	8.3%
• Very Large Farm Acreage as Percent of Farm Acreage	36.8%
• Percent Population Receiving SNAP Benefits	5.6%
• Number of Retail Food Establishments/10,000 People:	23.9
- Grocery Stores	1.9
- Convenience stores	4.2
- Supercenters/whse clubs	0.2
- Limited service restaurants	9.1
- Full service restaurants	7.0
- Special food services	1.6

The Food Industry Center
University of Minnesota

The State Level Food System Indicators Project is sponsored by the University of Minnesota's Healthy Foods, Healthy Lives Institute and the Food Industry Center. Additional information available at:
http://foodindustrycenter.umn.edu/Research/Sustainability/FoodSystemindicators/index.htm.

Figure 1: State Fact Sheet developed by the Healthy Food Healthy Lives research group.

In order for the PCA and CPCA processes to produce meaningful results careful data management is necessary. Firstly, the data must be normalized in order to avoid being dominated by population or land size trends. For example there will be more grocery stores in a state with a higher population. However, in comparisons across states, we are not interested in population differences, so the number of grocery stores must be normalized by the population. The new variable is then grocery stores per a unit of population, in this case 10,000 people. Secondly, PCA and CPCA are very sensitive to different scales and outliers. This requires each variable to be standardized with mean zero and variance one. Outliers, with values greater than 2.5 or less than -2.5 are then removed. A detailed discussion for the need to standardize the data and remove outliers is given in Chapter 3.

The PCA process limits the number of variables that can to be included in a single analysis. The maximum number of included variables depends on the number of observations. Therefore, a single PCA cannot be used for the entire indicator set collected by the HFHL group and requires that the large indicator set be grouped into subsets for analysis. A detailed discussion of the appropriate number of variables per number of observations is given in Chapter 3. The focus of the PCA and CPCA is on three normalized, standardized, outlier free groups of approximately 10 indicators each for 1997, 2002, and 2007.

The first group contains variables that explain the economic structure of the food system for a state. This group includes the percent of state employment in each food system activity, percent of state population living in metropolitan areas, percent of state land devoted to agriculture, and grocery stores per 10,000 people. These indicators were collected for 2002 and 2007. A full description of variable definitions and abbreviations is given in Table 2.1. A Table of the standardized data for each state for each variable for all years is given in Appendix A.

The second group contains variables that explain the structure of agricultural production in a state. The group includes the value of chemicals applied, the value of fertilizer applied, percent of cropland that is irrigated, percent of farms considered very large, government payments to agriculture, net farm income, percent of agricultural land in conservation, percent of agricultural sales attributed to crops, and percent of agricultural land used for growing crops. These indicators were collected for 1997, 2002, and 2007. A full description of variable definitions and abbreviations is given in Table 2.2. A Table of the standardized data for each state for each variable for 1997, 2002, 2007 and is given in Appendix B.

The third group contains variables that explain consumption and health patterns in a state. It includes obesity rates, diabetes rates, food security, and percent of expenditure at different types of food establishments. These indicators were collected for 1997, 2002, and 2007. A full description of variable definitions and abbreviations is given in Table 2.3. A Table of the standardized data for each state for each variable for 1997, 2002, and 2007 is given in Appendix C.

Table 2.1: Variable Definitions and Names Economic Structure of Food System

Variable Definition	Variable Abbreviation
Percent of total state employment in input supply	Input Supply
Percent of total state employment in primary production	Primary Prod
Percent of total state employment in processing	Processing
Percent of total state employment in distribution	Distribution
Percent of total state employment in retail	Retail
Percent of total state employment in waste	Waste
Percent of total state land in agriculture	Ag Land
Percent of total state population in metropolitan areas	Met. Population
Number of grocery stores in a state per 10,000 people	Grocery stores

Table 2.2: Variable Definitions Agricultural Production Intensity Group

Variable Definition	Variable Abbreviation
Value of chemicals used per acre of agricultural land	Chemicals
Value of fertilizer used per acre of agricultural land	Fertilizer
Percent of total agricultural land in conservation	Conserve
Percent of total agricultural land that is irrigated	Irrigation
Percent of total agricultural land used for crops	Crop Land
The ratio of net farm income to agricultural sales	Income
The ratio of government payments to agriculture to agricultural sales	Payments
Percent of total agricultural sales that are crop sales	CropSales
Percent of farms that have $500,000 or greater sales	VL Farm

Table 2.3: Variable Definitions and Names for Consumption and Health Group

Variable Definition	Variable Abbreviation
Percent of adult population who are obese	Obese
Percent of adult population who have diabetes	Diabetes
Percent of population eligible for SNAP benefits	SNAP
Percent of food expenditures at grocery stores or supercenters	Grocery
Percent of food expenditures at convenience stores	Convenient
Percent of food expenditures at full service restaurants	Full Service
Percent of food expenditures at limited service restaurants	Limited Service
Percent of food expenditures at food service establishments	Food Service

Data Availability

Data availability is a major issue for indicators at the state level over multiple years. While there were numerous indicators available at the state level, many of interest, particularly in the social and environmental issue categories, are not available. The importance of the environmental and social aspects of the food system has not been recognized for as long as the affordability and health aspects of the food system. Consequently less complete data sets are available at the state level for these indicators. In addition, as the understanding of food system advances, the aspects more recently considered important will not have complete data sets for past years in order to conduct time stability analysis. Consistent and complete data collection over time is very important, because it makes it possible to research the current structure of the food system and how it has changed over time.

Chapter 3: Principal Component Analysis

Principal components analysis (PCA) is commonly used as a data reduction method, as well as for exploratory analysis. The technique is used across many fields of research, with a number of previous food system studies taking advantage of its benefits[1]. For data reduction, the technique allows for a fewer number of Principal Components (PC) than original variables to capture a large amount of the variance. For exploratory analysis, the technique allows for the data to elicit structure without a priori notions of underlying structure. PCA will be used to address the first objective of the thesis; to reduce the dimensionality of the state level indicators over multiple years, while exploring if there are interpretable and meaningful underlying component structures to the indicators.

PCA analyzes the covariance matrix of a set of variables and thus is a variance focused technique. The statistical technique of PCA began with papers by Pearson and Hotelling. These two papers were 32 years apart and discussed two different methods for developing PCA. Pearson focused on geometric optimization problems, using lines and planes to calculate a best fit for a set of points (Pearson 1901). Hotelling introduced the traditional algebraic derivation that will be discussed in this section (Hotelling 1933). The following is an in depth

[1] There are a number PCA studies that have been conducted on topics pertaining to food each focusing on a small portion of the food system. An analysis of the sourcing and promotion of local foods by food cooperatives was recently conducted by Katchova and Woods in 2011. The paper aimed to study multiple strategies that food cooperatives used, while working with in the local food system. A second study analyzed supermarket management practices (King and Jacobson 2001). Using survey information developed by the Supermarket Panel, a PCA was conducted and the component scores were used in a regression analysis predicting store performance. PCA studies have also been conducted on food chemistry and sensory characteristics of food (Csomos 2002, Borgognone 2001).

discussion of the definition and derivation of PC's. It is derived from multiple sources on PCA (Jackson 1991, Joliffe 2002, and Hardle 2007).

Technical Review of Principal Component Analysis

Definition of a Principal Component. A PC is a linear combination of random variables, such that the coefficients on each variable maximize the variance of that PC. The variance of each PC is maximized orthogonal to – i.e., independent of - the variance of all other PC's. In addition, the PC's are ordered with the variance captured by each PC decreasing from highest to lowest. Therefore the first PC is found by maximizing the variance explained across the entire sample, the second PC is found by maximizing the variance explained out of the remaining variance across the entire sample, and so on. The number of PC's will equal the total number of variables in the analysis. Thus if there are 15 variables in the PCA there will be 15 PC's. However, if the data are appropriate for PCA, the first few PC's will account for a majority of the variance. The notation used for defining PC's follows that used by Joliffe (Joliffe 2002 pg. 2).

Given that,

X = a vector of p random variables
α_1 = a vector of p constants
′ = the transpose of the vector

The first PC is then the linear equation

$$\alpha_1^{'}\mathrm{x} = \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1p}x_p = \sum_{j=1}^{p} \alpha_{1j}\, x_j$$

The process is repeated with each principal component following the same form and capturing the maximum amount of remaining variance. If there are k variables in the analysis, then the final PC found will be of the form

$$\alpha_k' \mathrm{x} = \alpha_{k1}x_1 + \alpha_{k2}x_2 + \cdots + \alpha_{kp}x_p = \sum_{j=1}^{p} \alpha_{kj}\, x_j$$

Deriving the Principal Components. Analysis of the covariance matrix of variables solves the linear equations discussed above under the restraint of each having the maximum orthogonal variance. In practice the sample covariance matrix is used because the true covariance matrix for the population is unknown. Under the constraint of unit length, or $\alpha_k' \alpha_k = 1$, the maximization technique used is Lagrange Multipliers. The derivation of PC's continues to use the notation used by Joliffe (Joliffe 2002 pg. 4).

Given that,

λ = the Lagrange multiplier
I_p = the identity matrix of size pxp
$\sum$ = the covariance matrix for the random variable vector x
var($\alpha_k' \mathrm{x}$) = $\alpha_k' \sum \alpha_k$.

The derivation using a Lagrange Multiplier is as follows.

$\max_k \alpha_k' \sum \alpha_k$ subject to $\alpha_k' \alpha_k = 1$

$$\alpha_k' \sum \alpha_k - \lambda(\alpha_k' \alpha_k - 1)$$

Differentiate with respect to α_k and set solutions equal to 0

$$\sum \alpha_k - \lambda \alpha_k = 0$$

Factor out α_k

$$\left(\Sigma - \lambda I_p\right)\alpha_k = 0$$

Solving this gives λ as the eigenvalue and α_k as the eigenvector. With the identity matrix being of dimension pxp there are p eigenvectors, however only one maximizes the var($\alpha_k' \mathrm{x}$), and thus is the first PC. The eigenvector that corresponds to the largest value of λ is the first PC. The corresponding coefficients of the linear combination are referred to as the component loadings.

A rearrangement of the first order conditions, substituting the result into the initial maximization condition, makes the need to maximize λ clear.

$$\sum \alpha_k - \lambda\alpha_k = 0 \;\Rightarrow\; \sum \alpha_k = \lambda\alpha_k$$
$$\max_k \alpha_k^{'} \sum \alpha_k \;\Rightarrow\; \max_k \alpha_k^{'} \lambda\alpha_k \;\Rightarrow \max_k \lambda\alpha_k^{'} \alpha_k \;\Rightarrow\; \max \lambda$$

Application of Principal Component Analysis

The application of PCA is made quite simple by statistical packages such as SAS and Stata, however there are still numerous issues needed to be addressed for PCA to produce meaningful results. First, proper management of the data before the PCA process begins is necessary to ensure proper analysis. This includes the standardization of variables and the removal of outliers from the analysis. Second, careful consideration must be given to each of the following decisions in the PCA process: addressing the number of observations needed to be included, appropriateness of the included variables, the number of components to retain, and the rotation method. A number of sources have been consulted in order to consider the large amount of variation in PCA procedures (Jackson 1991, Joliffe 2002, Garson 2005, Hardle 2007).

This section provides a step-by-step discussion of these issues using a sample variable set consisting of the first seven indicators for 2002 from the Economic section of the State Fact Sheet. The sample variable group is used for illustrative purposes. The explanation for the standardization of the data, the removal of outliers, and the decisions made during the PCA process are made clearer when accompanied by a concrete example. The sample variable group was only chosen with the intention to strengthen the description of PCA application and not for the purpose of interpretation.

The sample variable set contains the percent of agricultural sales attributed to crops, the percent of agricultural sales attributed to fruits and nuts, the percent of agricultural sales attributed to vegetables, net farm income, government payments to agriculture, farming diversity, and expenditure on agriculture research. The variable definitions and abbreviations for the sample variable set are given in Table 3.1. The unstandardized data for the sample variable set are given in Appendix D.

Table 3.1: Sample Variable Set Definitions and Abbreviations

Variable Definition	Variable Abbreviation
Percent of agricultural sales from crop sales	Crop
Percent of agricultural sales from vegetable sales	Vegetable
Percent of agricultural sales from fruit and nut sales	Fruit and Nut
The ratio of net farm income to agricultural sales	Income
The ratio of agricultural government payments to agricultural sales	Payments
Percent of agricultural sales by the largest 3 commodities	Top3
The ratio of agricultural research expenditure to agricultural sales	Research

Number of Observations. Although there are some common theories on the number of observations needed to perform PCA accurately, methodologists differ widely in opinion on this matter. The most stringent recommendations are the rules of 200 and 300 (Gorsuch 1983, Norusis 2005). These two recommendations state that the sample should be at least 200 or 300 observations, respectively. However, significantly lower suggestions also have been made. The rule of 100 recommends that the sample size be at least 5 times larger than the number of variables in the analysis or a minimum of 100 observations (Hatcher 1994). The most popular and most often followed recommendations are based on a subject to variable ratio. These recommendations state that the ratio of observations to variables should be at least between 5 and 10 (Bryant and Yarnold 1995). The need for the analysis to be conducted through three separate groups is caused by the subject to variable ratio. There are 50 subjects, each state being a subject, in the analysis of state level indicators, and therefore PCA can only be conducted on a maximum of 10 variables at a time under the least restrictive rule.

Despite the many and varying recommendations there are some general agreements among the methodologists. The lower range of sample size recommendations can used when dealing with highly multicollinear variables (Garson 2005). Highly multicollinear variables are closely related to each other, producing an underlying structure that is more clearly defined. This allows for the sample size to be smaller, while still determining clean components. Methodologists also agree that more observations are needed for a clear PCA result when the amount of variance captured by the retained components is lower (Garson 2005). This means that each of the retained components captures less of the variance of each variable. This makes a larger sample size necessary for clear component structure.

Correlation Matrix Consideration. The correlation matrix can be used as a preliminary measure of the appropriateness of each variable's inclusion in a group. The correlation matrix shows the Pearson-correlation coefficient between each variable. Since this is a measure of how related the variables are, a variable that is appropriate for the PCA group should have some high positive correlation coefficients with other variables in the group. If a variable has only low correlation coefficients, there is a large possibility it will not load strongly on any one component. Therefore, the variable should not be included in the group because it will not allow for adequate data reduction and clean component loadings.

Table 3.2 displays the Pearson Correlation matrix for the sample variable set. By looking at the table, the variable "Top3" would be considered inappropriate for this group. It has only one low positive correlation coefficient of 0.1046. In addition, the 5 remaining negative correlation coefficients are low as well. The variable for farm diversity, "Top3", will most likely not load with other variables in the PCA process because it has no positive correlation coefficients larger than 0.20 and is negatively correlated with nearly all the other variables. As a result, it should not be included in the procedure.

Table 3.2: Pearson Correlation Matrix for Example Variable Set

	Crop	Fruit and Nut	Vegetable	Income	Payments	Top3	Research
Crop	1	0.1329	0.0451	-0.1732	0.1294	-0.1310	0.2301
Fruit and Nut	0.1329	1	0.6184	0.0138	-0.4883	-0.1268	0.4743
Vegetable	0.0451	0.6184	1	0.2849	-0.5840	-0.2077	0.2133
Income	-0.1732	0.0138	0.2849	1	-0.2606	-0.1670	-0.0585
Payments	0.1294	-0.4883	-0.5840	-0.2606	1	0.1046	-0.3018
Top3	-0.1310	-0.1268	-0.2077	-0.1670	0.1046	1	-0.0805
Research	0.2301	0.4743	0.2133	-0.0585	-0.3018	-0.0805	1

Covariance Matrix and Standardization of the data. Variables measured on different scales in the data analyzed by PCA can cause inaccurate component structures (Jackson 1991, Hardle 2007). This issue must be carefully considered and rectified before the techniques are performed. If the variables are allowed to remain in differing scales, the assumption of homogeneous variances for PCA will be violated, since heterogeneous scaling of the variables is a major cause of heterogeneous variances. Greater weight is given to the variables with larger variances, and with differing scales larger variance may be due solely to a larger scale (Jackson 1991). Standardizing the scale of each variable by subtracting the mean from each observation and then dividing by the standard deviation results in equal weight being given to each variable. This guarantees that the homogeneous variances assumption holds. All data used in the PCA techniques in this thesis are standardized.

While standardization has no effect on the correlation matrix, it has a significant effect on the covariance matrix. PCA is a variance based technique, so the issues discussed above are very serious. The sample variable group has variables with 2 different scales, making standardization necessary. Table 3.3 presents the covariance matrix using unstandardized and standardized variables. The covariance matrix using unstandardized data has very low values. Looking at these values, it seems that none of the variables change together. However, this is caused by the differing scales and means of each variable. Once the data are standardized, giving each a mean of 0 and standard deviation of 1, the values in the covariance matrix are much higher. In addition to higher covariance coefficients, standardization makes decisions, such as factor interpretation and scree plot analysis easier. These differences will be discussed in the rotation method and scree plot sections respectively.

Table 3.3: Unstandardized and Standardized Covariance Matrices for Sample Variable Group

NAME	Crop	Fruit and Nut	Vegetable	Income	Payments	Top3	Research
			Unstandardized Data				
Crop	0.027	0.000	-0.001	-0.002	0.001	-0.003	0.000
Fruit and Nut	0.000	0.007	0.003	0.002	-0.001	-0.004	0.000
Vegetable	-0.001	0.003	0.005	0.002	-0.001	-0.003	0.000
Income	-0.002	0.002	0.002	0.009	-0.001	-0.003	0.001
Payments	0.001	-0.001	-0.001	-0.001	0.001	0.000	0.000
Top3	-0.003	-0.004	-0.003	-0.003	0.000	0.018	-0.001
Research	0.000	0.000	0.000	0.001	0.000	-0.001	0.002
Mean	0.348	0.044	0.065	0.187	0.032	0.623	0.036
Std. Dev.	0.165	0.082	0.072	0.094	0.021	0.132	0.043

Table 3.3 continued: Covariance Matrices for Sample Variable Group

			Standardized Data				
Crop	1.020	-0.004	-0.066	-0.156	0.239	-0.134	0.025
Fruit and Nut	-0.004	1.020	0.532	0.219	-0.465	-0.363	0.050
Vegetable	-0.066	0.532	1.020	0.330	-0.485	-0.288	0.094
Income	-0.156	0.219	0.330	1.020	-0.241	-0.255	0.181
Payments	0.239	-0.465	-0.485	-0.241	1.020	0.145	-0.159
Top3	-0.134	-0.363	-0.288	-0.255	0.145	1.020	-0.235
Research	0.025	0.050	0.094	0.181	-0.159	-0.235	1.020
Mean	0	0	0	0	0	0	0
Std. Dev.	1.01	1.01	1.01	1.01	1.01	1.01	1.01

Outliers in PCA. Inaccurate component structure can be caused by outliers being included in the PCA (Hardle 2007). The large variance associated with outliers can detract from the underlying variance of the majority of the data. If the outliers remain in the data, the variance of each variable may be significantly different, from the variance of each variable without the outliers. In practice this can cause variables to load together on a component because each has large outliers. It may also cause a variable with a truly low variance to load with variables of high variance.

There are techniques available that weight the data to allow for the outliers to be included in the analysis. However, these techniques weight each data point, which is inappropriate for data that have only a few outliers (Jackson 1991). Using standardized data, outliers are defined in this study as having standardized values greater than 2.5 or less than -2.5. These values are removed from the data before PCA is performed. Removing these outliers has clear benefits to component interpretation and scree plot analysis. These differences will be discussed in the rotation method and scree plot sections respectively.

Determining the Number of Components. There are numerous criteria for determining the number of components to retain in the PCA. Many references detail the multiple criteria available to determine the number of components, but only three criteria will be discussed here: the Guttman-Kaiser criterion, the scree plot, and the variance explained criterion (Joliffe 2002, Garson 2005, Hardle 2007). These criteria are commonly used in current PCA research because of their ease of use and accurate results when considered in conjunction with each other.

The Guttman-Kaiser criterion was first used by Guttman in 1954 and later by Kaiser in 1960 (Garson 2005). Only components with eigenvalues of 1 or larger are considered significant. If a component has an eigenvalue of 1 or larger, then it explains at least as much variance as adding a single original variable by itself. There is much debate on the appropriateness of this method. Despite being a very commonly used cut-off criterion, some studies show that it consistently overestimates the true number of relevant components. Others have determined that there are more accurate criteria when communalities, a measure of the variance accounted for by the retained components for each variable, are low. However the Guttman-Kaiser criterion performs well when there are high communalities (Joliffe 2002). It is suggested that when using this criterion a statement of communalities is made. A detailed discussion on communality is presented in a following section.

Another popular component selecting criterion is the proportion of variance explained (Jackson 1991, Garson 2005). This method simply retains the number of components that account for the pre-determined percent of the original variance. This method is not considered to be reliable when used alone. With no rules on the amount of total variance that should be kept, the percentage chosen is arbitrary. Researchers not concerned with the number of components may use 90% as their cut-off, while others more concerned with data reduction may choose a percentage as low as 50%. Usually accounting for less total variance allows for fewer components to be retained, achieving greater data reduction. Despite the many caveats against this criterion, the most appropriate case for its use is under exploratory PCA (Joliffe 2002). With little knowledge of the underlying structure of the sample this criterion can be used to establish initial PCA models.

Table 3.4 shows the eigenvalue and proportion of variance explained by each component in a PCA using the standardized sample variable group. Using the Guttman-Kaiser criterion, two components would be retained. This is also appropriate when the proportion of variance explained is considered. Retaining two components would account for 65% of the original variance to be explained. This is well above the 50% mark sometimes used for data reduction.

Table 3.4: Eigenvalues and Variance Explained for Sample PCA group

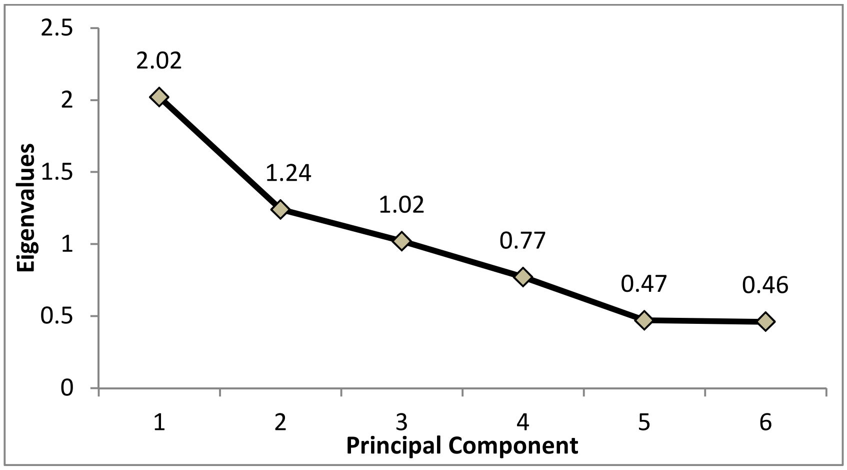
Component Number	Eigenvalue	Proportion of Variance Explained
1	2.43	0.41
2	1.45	0.65
3	0.69	0.76
4	0.67	0.88
5	0.44	0.95
6	0.31	1.00

A final criterion, the scree plot, is a graphical technique used to determine the number of components. The name scree plot comes from the term for rubble at the bottom of a cliff. As in nature, the scree, or rubble, will be found at the bottom of the cliff made by the graph. The graph plots the eigenvalue of each component. A clear scree plot will have a distinct drop off, or cliff. At the bottom of the cliff is a turning point, or “elbow” where the graph begins to level off. Components up to and including this point are retained, while components after this point are deleted (Garson 2005). The subjectivity of this technique is often criticized, stating the researcher can choose an “elbow” that coincides with his or her agenda. In addition, some scree plots do not have clear “elbows” and may be smooth curves. Despite these critiques, scree plots are very commonly used for component retention.

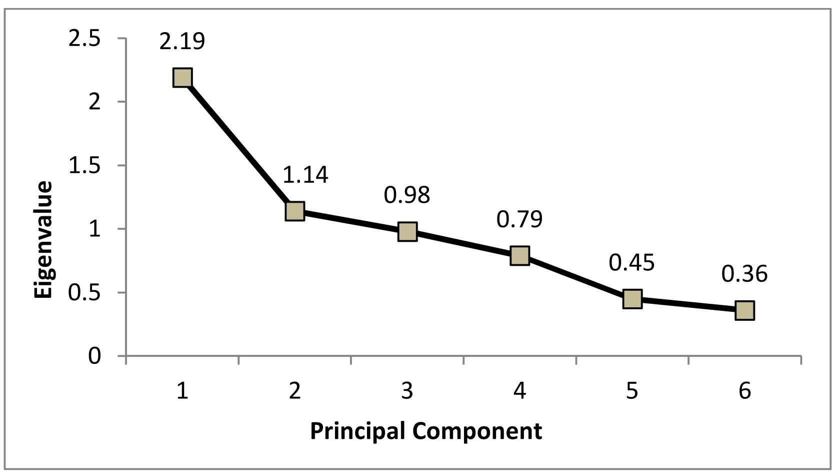
Using the data management practices discussed earlier, standardizing the data and removing the outliers, scree plot analysis is made much clearer. Scree plots for the unstandardized, standardized, and standardized without outliers data for the sample variable group are shown in the three panels of Figure 3.1. The scree plot in Panel a has no clear “elbow”. Deciding on the number of components only using the scree plot is not possible for the unstandardized data. However, as the data are properly handled an obvious “elbow” becomes present in Panels b and c. The “elbow” occurs at the second component. This means that 2 components should be retained, coinciding with the other criteria.

Communality Consideration. Communalities are another measure of how appropriate a single variable is in a particular PCA group. However, unlike the other measures, a communality can only be considered after a PCA is run because it depends on the number of retained components (Garson 2005). It is calculated by the summing the squared loadings for one variable across the retained components. The measure will be less than 1 if the number of retained components is less than the number of variables, which is often the case, given the objective of data reduction. The simple summation across components is possible because the components are orthogonal. Therefore, the variation of the variable in one component will not overlap the variation of the variable in another component.

A low communality suggests that the variable may not be a good fit for the group, whereas a high communality means that the variable may be a good fit for the group (Garson 2005). This is logical because a low communality shows that little of the variation of a variable is captured by the retained components, while a high communality shows the opposite. There are no standards for high and low communality values, although a general rule of thumb is that greater than 0.65 is high and less than 0.3 is low (Garson 2005).

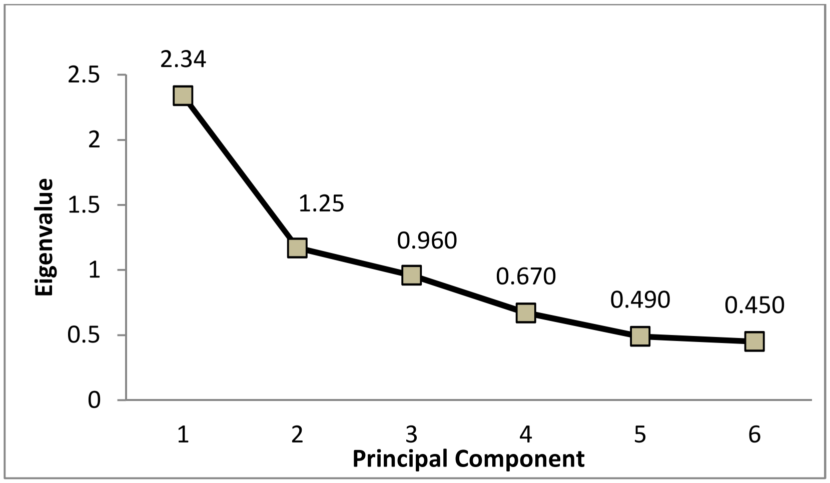


Panel a: Scree Plot of unstandardized Example Variable Group



Panel b: Scree Plot of standardized Example Variable Group

Figure 3.1: Scree plots for the unstandardized, standardized, and standardized without outliers data for the sample variable group



Panel c: Scree Plot of Standardized Example Variable Group without outliers

Figure 3.1 cont.: Scree plots for the unstandardized, standardized, and standardized without outliers data for the sample variable group

The communality measure must not be used independent of interpretability. Interpretability is the ability for the variables that load on a component to make sense or have a real world interpretation. If a factor has a high communality but is loaded on a component that does not make sense, the variable may not be a good fit because it does not add to the interpretation. The opposite is also true. If a factor has a low communality, but is loaded on a component that does make sense, the variable may be left in the PCA because it contributes to the interpretation. The communalities for each variable in the sample PCA group are given in Table 3.5.When looking at the communalities for the rotated factor pattern, the communality for the variable research might be considered low, although it is in the acceptable range. A value of 0.51 means that 51% of the total variance of the variable is captured by the two retained PC's. The variable would remain in the PCA because it adds interpretability to the first PC.

Rotation Method and Simple Structure. Rotation of components is designed to obtain simple structure of the resulting components. When the structure is simple, each variable will only identify strongly with one component (Garson 2005). This means that variables should have a loading close to 1 on one component and loadings close to 0 on the other components. In practice simple structure is hard to obtain, but rotation methods allow for the components to be close. With a simpler structure, the components are more easily interpreted because the loadings are clearly on one component. There are two groups of rotation methods with multiple types in each group (Jackson 1991, Hardle 2007). The first group is orthogonal rotation methods, where the components remain uncorrelated with each other. The second group is oblique rotations methods, which allow for the components to be correlated with each other.

There are a number of orthogonal rotations, but the most popular by far is the varimax rotation (Garson 2005). This method maximizes the squared loadings for each component for each variable. The result is a high loading on one component and low loadings on other components. The popularity of this rotation method has led to many examinations of it. Each confirms it as a proper rotation method given the desire for orthogonal components (Jackson 1991, Joliffe 2002). Other orthogonal methods such as quartimax and equamax can be further studied in other texts (Jackson 1991, Garson 2005).

There are also a number of oblique rotations, although their use is less common. The most popular oblique rotation is promax rotation (Jackson 1991, Garson 2005). This is a two stage rotation process that first performs an orthogonal rotation, followed by an oblique rotation. Many criticisms have been made of oblique rotations because the main purpose of rotation, obtaining simple structure for ease of interpretation, is not always achieved (Jackson 1991). Allowing for the components to be correlated allows for cross-loading, high loadings on more than one component, which decreases ease of interpretation. Other oblique rotation methods, such as oblimin, maxplane, and covarimin can be further studied in many texts (Jackson 1991, Garson 2005).

Table 3.5 below demonstrates the benefits of component rotation. Continuing to use the sample variable group, the original and rotated factor patterns are given for the standardized data without outliers. The varimax rotation method is used for all PCA in this study. Component loadings with absolute values less than 0.4 have been left out of the table for ease of comparison. In the original factor pattern multiple variables are cross-loaded. This complicates interpretation. There are no cross-loadings in the rotated factor pattern. In addition, all communality values are larger in the rotated factor pattern. Although the effect of rotation in this example is minor because of the low number of variables and few retained components, a PCA with more variables and more retained components benefits greatly from component rotation.

Table 3.5: Original and Rotated Factor Patterns and Communalities

Original Factor Pattern and Communalities			
Variable	Factor 1	Factor 2	Communality
Vegetable	0.82		0.72
Fruit and Nut	0.87		0.74
Research	0.64		0.48
Payments	-0.72	0.42	0.68
Crop	0.47	0.73	0.60
Income		-0.78	0.61
Rotated Factor Pattern and Communalities			
Vegetable	0.85		0.74
Fruit and Nut	0.85		0.76
Research	0.68		0.51
Payments	-0.78		0.71
Crop		0.77	0.64
Income		-0.77	0.63

Obtaining simple structure can also be achieved through the data management practices discussed earlier. Table 3.6 shows the rotated factor pattern for the unstandardized, standardized, and standardized without outliers data for the sample variable group. Values less than 0.4 were again left out for ease of comparison. The factor pattern for the unstandardized data has multiple cross-loadings, as well as a variable that does not load on either component, and is not easily interpreted. Using the standardized data eliminates the cross-loading on government payments to agriculture and causes expenditure on research to load on the first component. In addition, most of communalities have increased.

Comparing the standardized component structure to the component structure without outliers shows obvious benefits to removing outliers. The component structure without outliers has clear and large loadings on the 2 components. There are no cross-loadings, allowing for easier interpretation. Net farm income loaded positively on the first component and negatively on the second component when outliers were included, but loads negatively without outliers. In addition, research expenditures has a strong loading on the first component when outliers are removed. This complete change in component structure shows the large effect outliers and scales can have on the results of the PCA process.

Table 3.6: Rotated Factor Pattern for Example Group

Variable	Factor 1	Factor 2	Communality
Unstandardized Factor Pattern			
Vegetable	0.82		0.67
Fruit and Nut	0.80		0.65
Research			0.07
Payments	-0.70	0.40	0.60
Crop		0.95	0.90
Income	0.46	-0.41	0.35
Standardized Factor Pattern			
Vegetable	0.83		0.71
Fruit and Nut	0.79		0.69
Research	0.42		0.28
Payments	-0.72		0.64
Crop		0.85	0.86
Income	0.42	-0.56	0.55
Standardized without Outliers Factor Pattern			
Vegetable	0.85		0.74
Fruit and Nut	0.85		0.76
Research	0.68		0.51
Payments	-0.78		0.71
Crop		0.77	0.64
Income		-0.77	0.63

Chapter 4: Partial Common Principal Component Analysis

Common Principal Component Analysis is a statistical technique used to confirm or deny a level of similarity between the component structures of two groups (Flury 1988, Schott 1988, Schott 1999). The CPCA process analyzes similarities among the covariance matrices of the different groups. There is a hierarchy of similarity and a maximum likelihood estimate can be found for each hierarchy level (Flury 1988). CPCA begins at the third level of the hierarchy, followed by a restricted case of Partial CPCA.

Partial CPCA will be used to complete the second objective of the thesis: comparing the component structures over multiple years to test for similarity. Data on the same species separated by gender is a common group distinction for CPCA in the literature, but the application of CPCA is quite widespread in fields other than biology. Applications include uses in multiple economic studies[2] (Katsuura 2001, Hoang 2009). While all levels of Flury's hierarchy are described below, attention is given to partial CPCA because it is the level tested for similarity in component structures across years for the food system indicators.

[2] Partial CPCA is used extensively in fields, such as biology and psychology, but has also found uses in social science fields. With a large amount of economic and financial data being collected for multiple time periods, partial CPCA has become popular in comparing the component structure of data sets of the same variables for multiple years. One study used partial CPCA to compare similarities among phases of the business cycle between the 1960's and 1990's(Katsuura 1990). A second study used partial CPCA to investigate similarities of regional welfare and economic disparities in Vietnam over three years, 1998, 2002, 2004 (Hoang 2009).

Hierarchy of Similarity Covariance Matrix Analysis

The hierarchy of similarity ranges from equality to no statistical similarity.

1. The highest level of similarity between covariance matrices is equality. This implies that the covariance matrix of k groups are all equal. Mathematically

$$\Sigma_1 = \Sigma_2 \dots = \Sigma_k$$

2. The second highest level of similarity is proportionality between covariance matrices. This implies that the covariance matrices of k groups are just proportional to a single covariance matrix by some set of constants. Mathematically

$$\Sigma_i = \rho_i \Sigma_1 \quad \text{for all} \quad i = 2,3,\dots,k$$

where ρ_i is a constant proportion.

3. The third highest level of similarity is the CPCA model. This model implies that the covariance matrices of k groups produce the same characteristic roots or components. Mathematically

$$\Sigma_i = \alpha \Lambda_i \alpha'$$

where Λ_i is a diagonal matrix of eigenvalues

4. The fourth highest level of similarity is the partial CPCA model. This model implies that the covariance matrices of k groups are similar up to q components and the rest may be specific to each matrix, where q is less than p and p is the total number of components. Mathematically

$$\Sigma_i = \alpha^i \Lambda_i \alpha^{i\prime} \quad \text{where} \quad \alpha^i = (\alpha_1, \dots \alpha_q, \alpha^i_{q+1} \dots, \alpha^i_p)$$

5. The final and lowest level of similarity is arbitrary covariance matrices. Each covariance matrix is independent of the others and must be analyzed on its own.

Technical Review of Partial Common Principal Component Analysis

Partial CPCA is the fourth level of the similarity hierarchy and allows for a certain number, q, of the first components of covariance matrices to be common and for the remainder, p-q, to be specific to each covariance matrix (Flury 1988). This type of analysis is appropriate when trying to compare component structures over multiple groups, while only retaining the first q components. The last p-q components are discarded for similar reasons discussed earlier in the review of PCA. If CPCA is conducted rather than partial CPCA the hypothesis of similar component structure may be rejected due to components that are not to be retained anyway.

If a particular model of partial CPCA fails to be rejected, pooling the data for a PCA is appropriate as long as only the first q components are being retained. For example, if a partial CPCA model comparing the first three PC's is not rejected, the group's variables can be pooled and one PCA run (Schott 1999, Flury 1988). The first three component structures of the pooled PCA will be similar to each individual PCA, but they will be of a higher quality because the number of observations will have been significantly increased. In addition, the component structure can be considered stable over time, when the groups are multiple years.

The null hypothesis for partial CPCA is that the first q components are similar. The null hypothesis is defined in the opposite manner to normal hypothesis testing. Usually the null hypothesis is defined in a way that causes rejection to be favorable, i.e. in a standard significance test the null hypothesis would be $H_0: \beta = 0$. Therefore by rejecting the null hypothesis a statistically significant coefficient has been found. In partial CPCA failing to reject the null hypothesis is the favorable outcome and shows similarity between the covariance matrices. The null hypothesis is given by

$$H_{CPC}(q): \quad \Sigma_i = \alpha^{(i)} \Lambda_i \alpha^{(i)'} \quad \text{for} \quad i = 1, \ldots k$$

where

$\Sigma_i =$ pxp covariance matrix

$\Lambda_i =$ diagonal matrix of eigenvalues

$\alpha^{(i)} = (\alpha_c, \alpha_s^{(i)})$

where α_c is the common part and $\alpha_s^{(i)}$ is the specific part (Flury 1988 pg. 67).

The null hypothesis can be evaluated with a maximum likelihood statistic following a χ^2 distribution. The maximum likelihood statistic is

$$\chi^2_{CPC}(q) = \sum_{i=1}^{k} n_i \log \frac{det\Sigma_i}{detS_i}$$

Where $S_i =$ the unrestricted covariance matrix. The statistic has $\frac{(k-1)q(2p-q-1)}{2}$ degrees of freedom (Flury 1988, p. 72). If the p-value is larger than 0.10, we fail to reject the partial CPCA model (Phillips 2000).

Application of Partial Common Principal Component Analysis

When testing partial CPCA, there are a number of approaches for considering the best model. First a predicted model level can be tested against the unrelated structure case (Flury 1988). This approach is commonly used when a strong a priori structure is being tested. However, for exploratory analysis, it is common to use the “jump up” approach (Flury 1988, Schott 1999). In this approach tests are begun at the lowest level of similarity and move on to higher levels of similarity if the lower level is not rejected. The lowest level of similarity for a partial CPCA would be having the first PC be similar. If this hypothesis is not rejected a test for the first two PC’s being similar would be conducted. The process continues until a model of a certain level is rejected.

Often multiple levels of CPCA will be tested and will fail to be rejected. In this case the reduced chi squared value is used as the best model selection criterion (Phillips 2000). The reduced chi squared value is the chi squared value of the maximum likelihood test divided by the degrees of freedom for the level being tested. Each level of the CPCA has different degrees of freedom; therefore, direct comparison of their chi squared values for a measure of best fit cannot be done. The reduced chi squared value standardizes the statistic to degrees of freedom allowing for direct comparison. Therefore, although a chi squared test may fail to reject multiple models, the model with the minimized reduced chi squared statistic is the most appropriate model (Flury 1998).

Continuing to use the sample data set from the PCA chapter, results for the individual PCA’s for 1997, 2002, 2007 using standardized and outlier free data are given in Table 4.1. A simple comparison of component structure between the three years can be done by looking at the individual PCA results. The component structures and magnitudes of the loadings for all 3 years are similar. The only difference is in the 1997 PCA, when government payments to agriculture loads on both components. However, statistically sound conclusions cannot be drawn by looking at the individual component structures.

Table 4.1: Sample Variable PCA Results

Variable	Factor 1	Factor 2	Communality
1997			
Vegetable	0.88		0.78
Fruit and Nut	0.75		0.57
Research	0.60		0.50
Payments	-0.55	0.63	0.70
Crop		0.90	0.83
Income		-0.46	0.40
2002			
Vegetable	0.85		0.74
Fruit and Nut	0.85		0.76
Research	0.68		0.51
Payments	-0.78		0.71
Crop		0.77	0.64
Income		-0.77	0.63
2007			
Vegetable	0.87		0.76
Fruit and Nut	0.86		0.76
Research	0.77		0.66
Payments	-0.72		0.71
Crop		0.79	0.72
Income		-0.58	0.44

The results for the CPCA are given in Table 4.2 using standardized and outlier free data. The equality, proportionality, and full CPCA models are all rejected because their p-values are smaller than 0.10. However, the partial CPCA models with both 1 and 2 similar components fail to be rejected. Using the reduced chi squared value, the partial CPCA model with 2 similar components is considered to be the best fitting model because it has the lowest value.

The partial CPCA results show two significant results. The first is that the component structure between 1997, 2002, and 2007 is consistent up to two components. This suggests that the component structure is stable over time. Data from multiple years can be pooled with stable component structures, allowing for analysis with more observations leading to more accurate results. In addition, a stable component structure suggests that it will be appropriate to apply the stable structure to out of sample data sets to obtain component scores. Fewer resources will be needed to analyze this group of variables in the future.

Table 4.2: Partial CPCA results for sample variable group for 1997, 2002, and 2007

Higher Model	Lower Model	P-Value	Reduced x^2
Equality	Proportionality	<0.0001	17.731
Proportionality	CPCA	0.0008	2.222
CPCA	Partial CPCA (2)	0.0157	2.068
Partial CPCA (2)	Partial CPCA (1)	0.9612	0.367
Partial CPCA (1)	Unrelated	0.6913	0.820

Table 4.3 shows the component structure for the pooled sample variable group. The structure and magnitude of the ladings in the two retained components are similar to each individual year, which follows the similarity shown by the partial CPCA.

Table 4.3: Pooled Sample Variable PCA Results

Variable	Factor 1	Factor 2	Communality
Vegetable	0.86		0.73
Fruit and Nut	0.82		0.58
Research	0.69		0.55
Payments	-0.64		0.72
Crop		0.84	0.81
Income		-0.60	0.47

Chapter 5: Results

The data groups defined in Chapter 2; the Economic Structure of Food System Group for 2002 and 2007, the Agricultural Production Intensity Group for 1997, 2002, and 2007, and the Health and Consumption Group for 1997, 2002, and 2007, are analyzed using the methods discussed in the PCA and partial CPCA chapters. As noted previously, the groups are necessary because of the subjects to variable ratio requirement. Each group has no more than 10 variables in order to not exceed the 5 to 1 ratio. The three individual groups that are analyzed were chosen out of many possible groups. The large number of food system indicators available means there are many possible combinations of 10 variables. However, each group was chosen with prior consideration given to importance, relatedness, and meaningfulness of the indicators included in the group. The variable definitions and abbreviations are given for each group in a table at the beginning of each group's section.

The variables for each year in each group were standardized and outliers were removed. A PCA was performed for data from each year of each group to test the component structure of the years independently. The SAS statistical program was used to perform the PCA. A combination of scree plot analysis and the Guttman-Kaiser criterion was used to determine the number of retained components. The varimax rotation method was used for all PCA's in order to minimize the chance of cross-loadings.

Partial CPCA was then used to do a formal comparison across years for each indicator group. The data set with standardized variables and outliers removed was also used for the partial CPCA. SAS was used to create a covariance matrix for each set of indicators. The covariance matrices were compared using a CPCA program written by Patrick Phillips (Phillips 2000). The program analyzed the covariance matrix to perform hierarchical comparisons of similarity. PCA and partial CPCA results for each group are discussed below.

The years were then pooled and a PCA was conducted to the level found appropriate by partial CPCA. The component structure for the pooled PCA was then used to calculate component scores for each year.

The component scores for each component and year are compared for four states within each U.S. Census region, with at least one state in each section of the region. The states selected for the Northeast Region are New Jersey, New York, Pennsylvania, and Vermont. The states selected for the Midwest Region are Indiana, Iowa, Minnesota, and Nebraska. The states selected for the South Region are Arkansas, Mississippi, Texas, and Virginia. The states selected for the West Region are Arizona, California, Colorado, and Oregon. A map of the U.S. Census regions is provided in Appendix E.

In each section the results are presented in the same order. First, the PCA results for the group are presented and discussed. Secondly the partial CPCA results are presented and discussed. The PCA results for each group with years pooled together are then presented, along with the component scores for each state. Finally, three separate types of component score comparisons are given using 2 types of charts.

1. A scatter chart with a 45 degree line is used to compare component scores for one component for each state across time. For the Economic Structure group 2002 and 2007 are compared. For the Agricultural Production Intensity and Health and Consumption Group 1997 and 2007 are compared.
2. A radar chart makes it possible to compare levels of all retained components for multiple states for one year. It is used to do a regional comparison, comparing all retained components of the four selected states in each region for 2007.
3. A radar chart also makes it possible to compare levels of all retained components for one state over time. It is used to do a state comparison over time, comparing all retained components for Minnesota over each year.

Economic Structure of Food System

Table 5.1 reproduces the variable definitions and abbreviations for the Economic Structure of the Food system group given in Chapter 2.

Table 5.1: Variable Definitions for Economic Structure of Food System

Variable Definition	Variable Abbreviation
Percent of total state employment in input supply	Input Supply
Percent of total state employment in primary production	Primary Prod
Percent of total state employment in processing	Processing
Percent of total state employment in distribution	Distribution
Percent of total state employment in retail	Retail
Percent of total state employment in waste management	Waste
Percent of total state land in agriculture	Ag Land
Percent of total state population in metropolitan areas	Met. Population
Number of grocery stores in a state per 10,000 people	Grocery stores

PCA Results. The PCA has 50 observations for each year, one for each state. The communalities for each variable, accumulated percent variance explained, and rotated component loadings with an absolute value greater than 0.40 are presented for 2002 and 2007 in Table 5.2.The three retained components adequately explain a majority of the variance of the original variables. The accumulated variance explained by the three components is 69% for 2002 and 70% for 2007. In addition, the communalities in both years are quite high. Both of these results show that a majority of the original information conveyed by the 9 variables is conveyed by the three components.

The component structures for the Economic Structure group are strong and interpretable. In both 2002 and 2007 the first component can be interpreted as upstream production activities, the second component as retail activities, and the third component as waste activities.

Component 1: Upstream Production Activities

The variables for employment in input supply, employment in processing, employment in primary production, and the percent of land in agriculture load heavily on the first component. States heavily involved in one upstream aspect of the food system also tend to be involved in the others. This result is logical. Although, chemicals and fertilizers are not necessarily more likely to be produced in an area that is heavily involved in primary production, employees are needed to sell and distribute the inputs to the farms. More retail and distribution employment for chemicals and fertilizer is needed in areas of high primary production activity. Food processing follows the same logic. Food processing is not necessarily more likely to take place near the primary production area, but more employment in the sector is needed to transport the large amount of food from the primary production area to the processing centers.

Component 2: Retail

Retail activities of food system products follow a separate and major factor. The variables for employment in distribution, employment in retail, and the number of grocery stores per 10,000 people load heavily on the second component. The grouping makes sense because there would be increased distribution employment and retail employment in states where there are many grocery stores. More employees would be needed to work in the grocery stores and more drivers would be needed to delivery to the grocery stores. These are likely to be populous, highly urbanized states.

Component 3: Waste

The variables for employment in waste activities and percent of population in metropolitan areas load heavily on the third component. Metropolitan areas have a large concentration of people and have a large amount of waste to process. Therefore, larger employment in waste treatment activities is needed. In addition states with fewer large waste facilities because of a less dense population may opt to export their waste to an established waste treatment area in another state, because it is cheaper than building their own site.

Table 5.2: PCA results for Economic Structure Group

Variable Name	Communalities	Factor 1: Upstream Production	Factor 2: Retail	Factor 3: Waste
		2002		
Input Supply	0.81	0.90		
Ag Land	0.73	0.82		
Processing	0.65	0.79		
Primary Prod	0.76	0.77		
Grocery Stores	0.64		0.76	
Distribution	0.57		0.78	
Retail	0.71		0.77	
Waste	0.83			0.87
Met. Population	0.74			0.80
Accumulated Percent Variance Explained		0.36	0.57	0.69
		2007		
Input Supply	0.75	0.89		
Ag Land	0.73	0.81		
Processing	0.62	0.78		
Primary Prod	0.81	0.89		
Grocery Stores	0.66		0.80	
Distribution	0.57		0.74	
Retail	0.50		0.56	
Waste	0.77			0.87
Met. Population	0.69			0.78
Accumulated Percent Variance Explained		0.37	0.53	0.70

Partial CPCA results. Table 5.3 presents the partial CPCA. Partial CPCA is restricted to three similar components because both of the individual PCA results retain the first three components. With each model's p-value larger than 0.10, all models fail to be rejected. The component structure of the group is very stable over time. Considering what the group represents, this result is not surprising. Large farms, grocery store chains, and waste management facilities do not change much in location or size over time. This consistency is reflected in stability of the constructs over time.

When multiple models fail to be rejected based on the p-value, the reduced chi squared statistic for the higher model must be considered. As discussed earlier, this statistic is minimized. The partial CPCA with three similar components has the smallest reduced chi squared statistic value, meaning that it is the most appropriate model. It is appropriate for the Economic Structure Group to be pooled together for PCA, as long as only the first 3 components are retained and analyzed.

Table 5.3: Similarity Hierarchy Results for Economic Structure Group

Higher Model	Lower Model	P-Value	Reduced x^2
Equality	Proportionality	0.2217	0.713
Proportionality	CPCA	0.2801	1.072
CPCA	Partial CPCA (3)	0.5456	0.492
Partial CPCA (3)	Partial CPCA (2)	0.9799	0.155
Partial CPCA (2)	Partial CPCA (1)	0.6978	0.670
Partial CPCA (1)	Unrelated	0.7509	0.761

Pooled PCA Results. The pooled PCA is performed with a data set consisting of the 2002 and 2007 data sets. This gave the pooled PCA 100 observations, one for each state in each year.

The pooled PCA component structure is very similar to the structure for each individual year. This is not surprising because the results of the partial CPCA shows the similarity of the two groups. The component structure for the pooled data set can be used to construct component scores for individual in-sample and out-of-sample years, because the component structure is stable over time. Table 5.4 shows the communalities, accumulated percent of total variance explained, and the component structure for the Economic Structure pooled group.

Table 5.4: PCA Results for Pooled Economic Structure Group

Variable Name	Communalities	Factor 1: Upstream Production	Factor 2: Retail	Factor 3: Waste
Input Supply	0.81	0.87		
Ag Land	0.73	0.85		
Processing	0.65	0.78		
Primary Prod	0.75	0.86		
Grocery Stores	0.63		0.81	
Distribution	0.59		0.73	
Retail	0.58		0.64	
Waste	0.76			0.85
Met. Population	0.72			0.79
Total Variance Explained		0.37	0.55	0.69

These component loadings can then be used to create components scores, which can be compared across states and time. The standardized data from each state is multiplied by the weight given to each variable in the pooled component structure. Each product is then summed across the component to get the component score. Using standardized data causes usual component scores tend to be between values of -3 and 3. Table 5.5 shows the component scores for each component for each state for 2002 and 2007.

Table 5.5: Component Scores for 2002 and 2007 for the Economic Structure Group

	2002			2007		
State	Factor1: Upstream Activities	Factor2: Retail	Factor3: Waste	Factor1: Upstream Activities	Factor2: Retail	Factor3: Waste
AL	0.03136	-0.22622	-0.46781	-0.06637	-0.41384	-0.36541
AK	0.08221	0.76767	0.12242	-0.17634	0.96169	-0.51236
AZ	-1.1616	-0.5667	0.00552	-1.06665	-0.8768	-0.62792
AR	1.03148	-0.2117	-0.08919	0.86096	-0.34877	0.00555
CA	-0.26947	0.22148	0.62243	-0.30839	0.30812	0.63208
CO	-0.49733	-0.34093	0.19301	-0.47483	-0.6437	0.11165
CT	-0.44669	-0.46746	1.11356	-0.52941	-0.23245	0.91692
DE	-0.65606	-1.76601	-0.5235	-0.64133	-1.59206	-0.78811
FL	-0.71507	-0.12948	0.10706	-0.7959	-0.22284	0.00847
GA	-0.50735	-0.32109	-0.32832	-0.46578	-0.38342	-0.30947
HI	-0.94031	2.07887	-0.61858	-0.8813	1.59953	-0.22071
ID	1.24467	1.18819	-0.03913	0.83375	0.54331	-0.45582
IL	0.31681	-0.56346	0.4663	0.19544	-0.3636	0.21783
IN	0.20399	-1.16047	0.00709	0.31945	-1.26135	0.04996
IA	2.87748	-0.71614	0.42754	3.15605	-1.05355	0.4313
KS	1.84563	-0.44904	1.12831	1.99979	-0.59257	2.54124
KY	0.49019	0.13452	-0.16497	0.34506	-0.22572	-0.2803
LA	-0.40714	0.50634	0.18076	-0.27729	0.06413	0.69622
ME	-0.78134	1.81255	0.6319	-0.4752	1.2351	1.70524
MD	-0.74095	-0.00325	0.5116	-0.84854	-0.04461	0.35573
MA	-1.10392	-0.26053	0.09754	-1.12529	-0.12455	0.05586
MI	-0.58562	-0.08722	0.41531	-0.62767	-0.06695	0.01158
MN	0.4056	-0.50131	-0.1227	0.32214	-0.58498	-0.14953
MS	0.50146	0.35768	-0.20423	0.33591	-0.23361	-0.367
MO	0.40627	-0.89883	-0.59239	0.32	-1.06069	-0.49963
MT	0.68091	1.23312	-0.8841	0.54321	0.1827	-1.32131
NE	2.98636	0.0572	0.42641	2.67699	-0.10966	0.11345
NV	-1.06328	-1.03905	0.14231	-0.9671	-1.15553	0.21962

Table 5.5 continued: Component Scores for 2002 and 2007 for the Economic Structure Group

	2002			2007		
State	Factor1: Upstream Activities	Factor2: Retail	Factor3: Waste	Factor1: Upstream Activities	Factor2: Retail	Factor3: Waste
NH	-1.15349	-0.16377	-0.29927	-1.2164	-0.17139	-0.82848
NJ	-1.02561	0.70395	1.68693	-1.06876	1.21713	1.74225
NM	-0.03091	-0.61932	-0.58031	-0.03657	-0.5496	-0.69952
NY	-1.17793	1.39568	-0.41399	-1.2231	1.69734	-0.41856
NC	-0.33677	-0.19713	-0.32571	-0.39516	-0.3696	-0.63916
ND	2.41881	2.05681	-1.01047	2.57117	1.97239	-1.11848
OH	-0.24273	-0.28528	-0.19101	-0.26279	-0.37969	-0.06711
OK	0.55007	-0.43593	-0.68585	0.62776	-0.79249	-0.6001
OR	0.244	0.63013	0.78538	0.25277	0.49443	0.96059
PA	-0.56134	-0.1541	0.2859	-0.55213	-0.27587	0.43713
RI	-1.23549	-0.18694	-0.82029	-1.11539	-0.10811	-0.53506
SC	-0.61472	-0.36793	-0.0164	-0.6263	-0.34766	0.04035
SD	1.91265	0.51326	-0.62578	1.97425	-0.10003	-0.49056
TN	-0.02026	0.03338	-0.29514	-0.26928	-0.37179	-1.04379
TX	0.23162	-1.01228	0.05065	0.19096	-1.21338	-0.36027
UT	-0.50423	-0.7545	-0.27151	-0.50789	-1.20501	-0.6152
VT	-0.73608	4.00066	0.56617	-0.72279	2.60426	-1.03288
VA	-0.57469	-0.63798	-0.67497	-0.58892	-0.5638	-0.4318
WA	0.15493	1.07697	1.10222	-0.02867	0.61085	0.42207
WV	0.07334	-0.02799	5.31967	0.03114	-0.49	4.39447
WI	0.31761	0.02654	-0.14194	0.33724	-0.65009	-0.41203
WY	0.58069	-0.81291	-1.15664	0.51999	-1.03655	-0.46812

Scores for Component 1: Upstream Production Activities

The component scores for upstream production activities range from 2.98 to -1.24 in 2002 and from 3.15 to -1.22 in 2007. Although the specific order changes, the same five states have the highest component scores in each year. The states are Iowa, Nebraska, South Dakota, North Dakota, and Kansas. Employment in the upstream activities of input supply, primary production, and food processing is consistently a large aspect of these states' overall employment. The four states with lowest components scores also remain the same between 2002 and 2007. The states are New York, New Hampshire, Massachusetts, and Rhode Island. Employment in the upstream activities of the food system is a smaller portion of state employment than in other states.

Figure 5.1 compares the value of the upstream activities component scores for 2002 and 2007. The 45 degree line represents the points where a state's component score is the same for 2002 and 2007. If a state is above the line, its component score increases from 2002 to 2007. If a state is below the line, its component score decreases from 2002 to 2007. A majority of the states are very near the 45 degree line, meaning many state's employment in upstream activities did not change much over the 5 year period. Nebraska and Indiana's component scores decrease the most between the two years, while Iowa's component score increases the most.

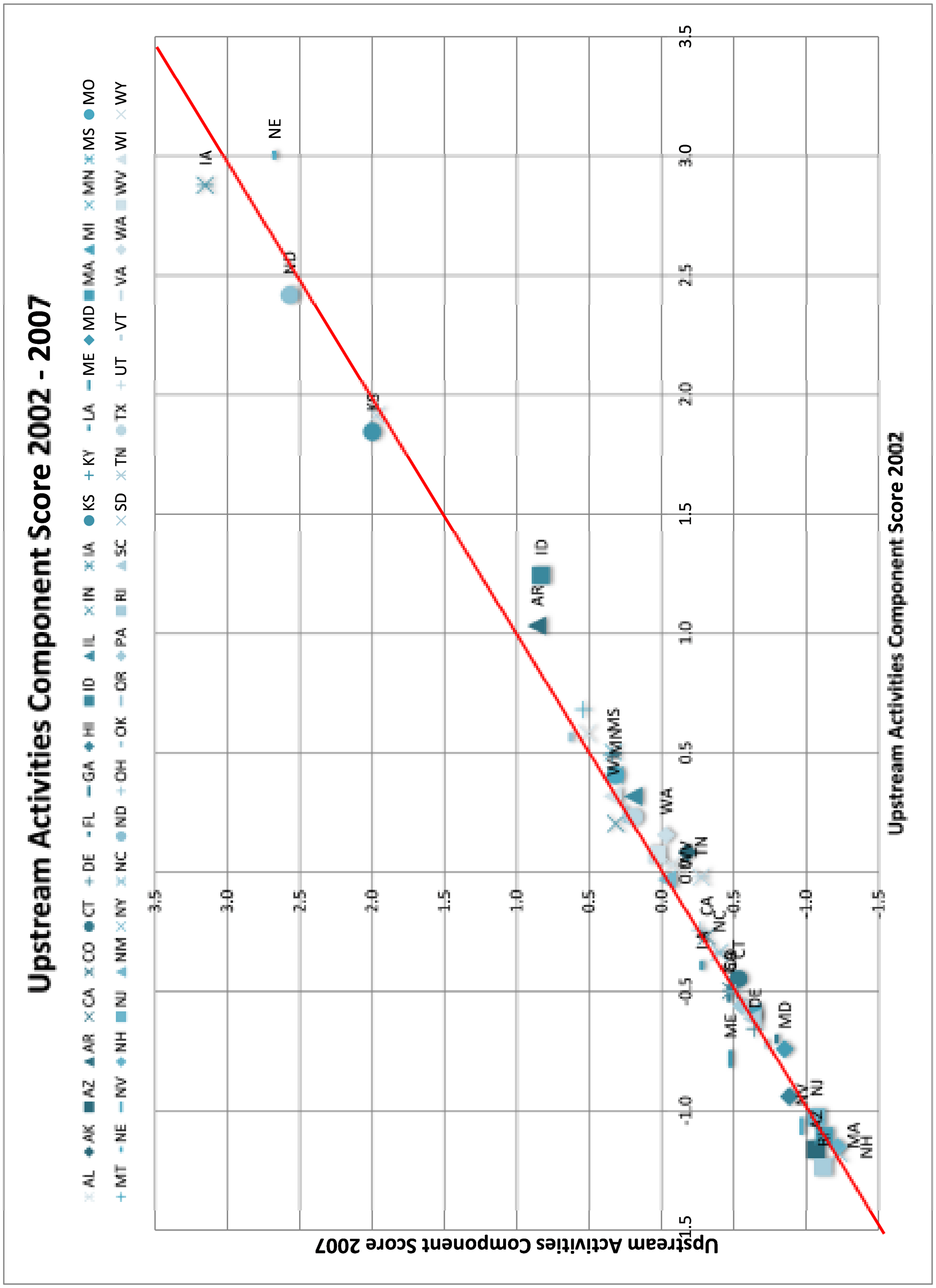


Figure 5.1: Upstream Component Score Comparison for 2002 and 2007

Scores for Component 2: Retail

The component scores for the retail components range from 2.54 to -1.32 in 2007 and from 4.00 to -1.76 in 2002. The order of the five states with the largest scores changes, but the group does not. The five states with the highest scores for both 2002 and 2007 are Vermont, Hawaii, North Dakota, Maine, and New York. The retailing of food products is consistently important in these states' economies. In both 2002 and 2007, Delaware, Indiana, Texas, and Nevada are among the five states with the lowest scores. However, the fifth state in that group changes between years. In 2002 the fifth state is Utah, while in 2007 it is Missouri.

Figure 5.2 compares the retail component scores of each state for 2002 and 2007. Unlike the upstream component scores, many of the states have moved off of the 45 degree line. This shows greater variability in food retail activities compared to upstream production activities. A majority of the states are below the 45 degree line, showing that the percentage of employment in food retail decreased for many states. Of these states, Vermont's component score decreased the most between the two years. Of the few states that have higher component scores in 2007, New Jersey is the farthest above the line. However, it does not stand apart from the other states above the line as Vermont does from the states below the line.

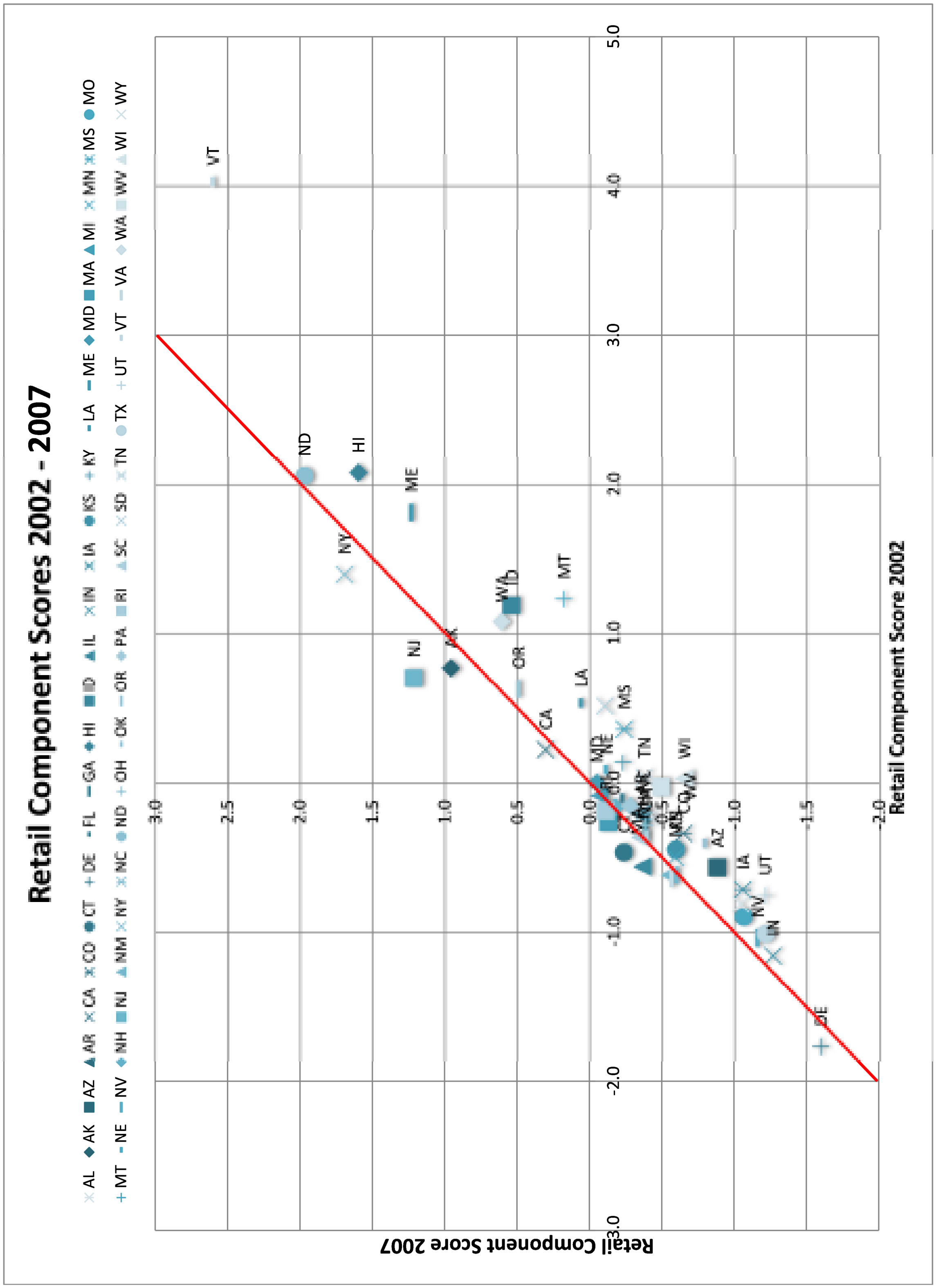


Figure 5.2: Retail Component Comparison for 2002 and 2007

Scores for Component 3: Waste

The component scores for the waste component range from 4.39 to -1.32 in 2007 and from 5.29 to -1.16 in 2002. Only the three states with the highest component scores remain the same between years for this component. The states are West Virginia, Kansas, and New Jersey. Waste disposal associated with food is a larger portion of these states economies than other states. Montana and North Dakota have the lowest component scores for both 2002 and 2007.

Figure 5.3 compares the component scores for the waste component for 2002 and 2007. Similar to the retail component a majority of the states are below the 45 degree line. Most states have a smaller portion of the employment in the food waste sector in 2007 than in 2002. West Virginia stands out as having the largest portion of employment in the food waste sector despite having the second largest drop in component score between the two years. Kansas and Maine both have large increases in component scores from 2002 to 2007.

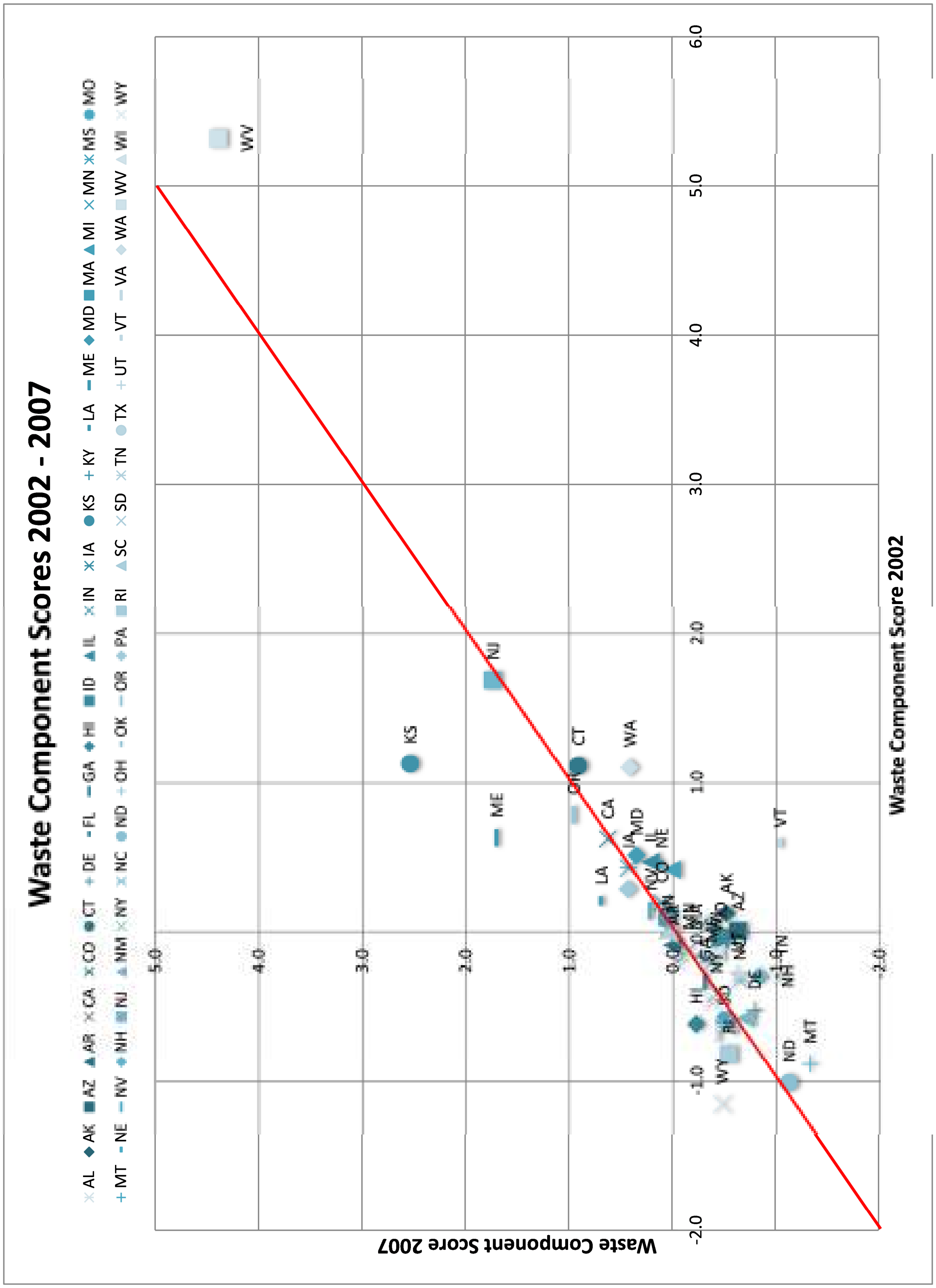


Figure 5.3: Waste Component Comparison for 2002 and 2007

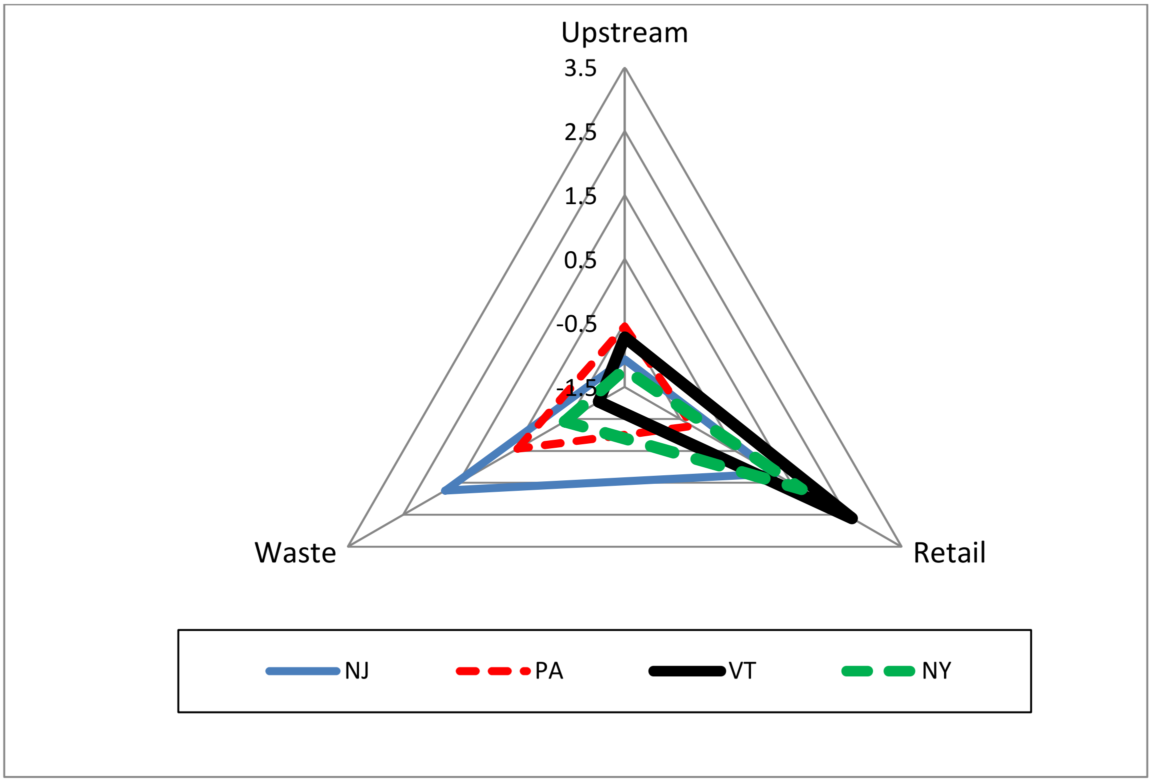
Regional Comparisons.

Northeast Region. Figure 5.4 Panel a shows the radar chart for the Northeast Region. The triangles created by the component scores for each state are skewed largely in the retail direction except in the case of Pennsylvania. The retailing of food is the most important component in the Northeast. This makes sense because of the high concentration of people. In addition, New Jersey and Pennsylvania have a long leg in the waste direction. The post-consumer waste component of the food system is important in these states, which is also due to the high concentration of people.

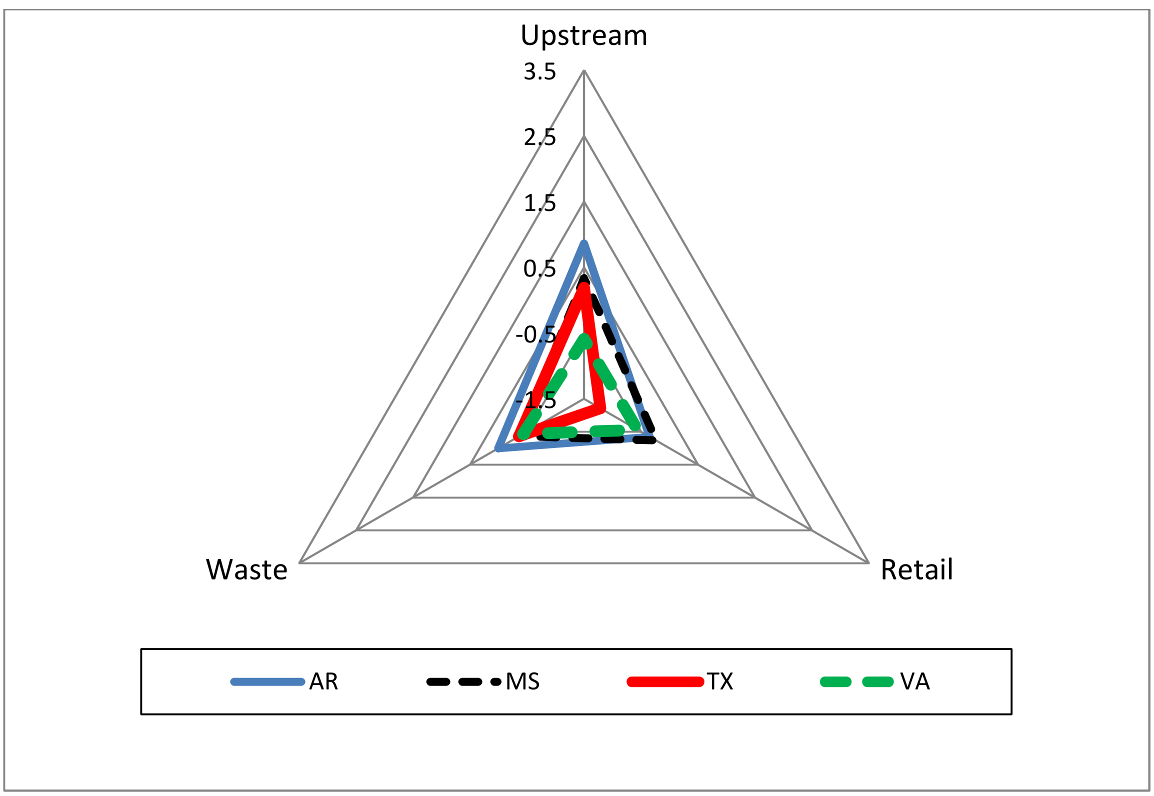
South Region. Figure 5.4 Panel b shows the radar chart for the South Region. Overall the South Region has smaller triangles than the other regions. The food system accounts for a relatively smaller portion of overall employment compared to other regions. The triangles are skewed toward the upstream activities component, showing that employment in the food system is mostly in primary production, processing, and input supply industries. Comparing Virginia to Arkansas, the triangles are very different. Arkansas has a much higher percentage of its employment in the food system, and is more skewed towards the upstream activities. Virginia has a very balanced, but smaller, employment distribution. This shows a large amount of variation can be present within a single region.

Midwest Region. Figure 5.4 Panel c shows the radar chart for the Midwest Region. The Midwest's employment in the food system is heavily centered around upstream activities. All of the triangles are skewed in the upstream activities component direction, with Iowa and Nebraska having extremely large values. Minnesota and Indiana have a more balanced distribution of employment, but the percentage of employment in the food system is less than in Nebraska and Iowa. Overall the Midwest is heavily involved in the food system, with a large percent of its workforce associated with it.

West Region. Figure 5.4 Panel d shows the radar chart for the West Region. Each of the states in the West Region has balanced employment distribution in the food system, but the states have very different overall levels of employment. The size of the triangle ranges from Oregon's large triangle, showing it is involved in each aspect of the food system, to Arizona's very small triangle, having no score larger than -1.5. A very minor percent of Arizona's employment is associated with the food system compared to other states. The contrasting sizes also show the large variation present in the region. This variation is due to the numerous and diverse nature of the states in the region.

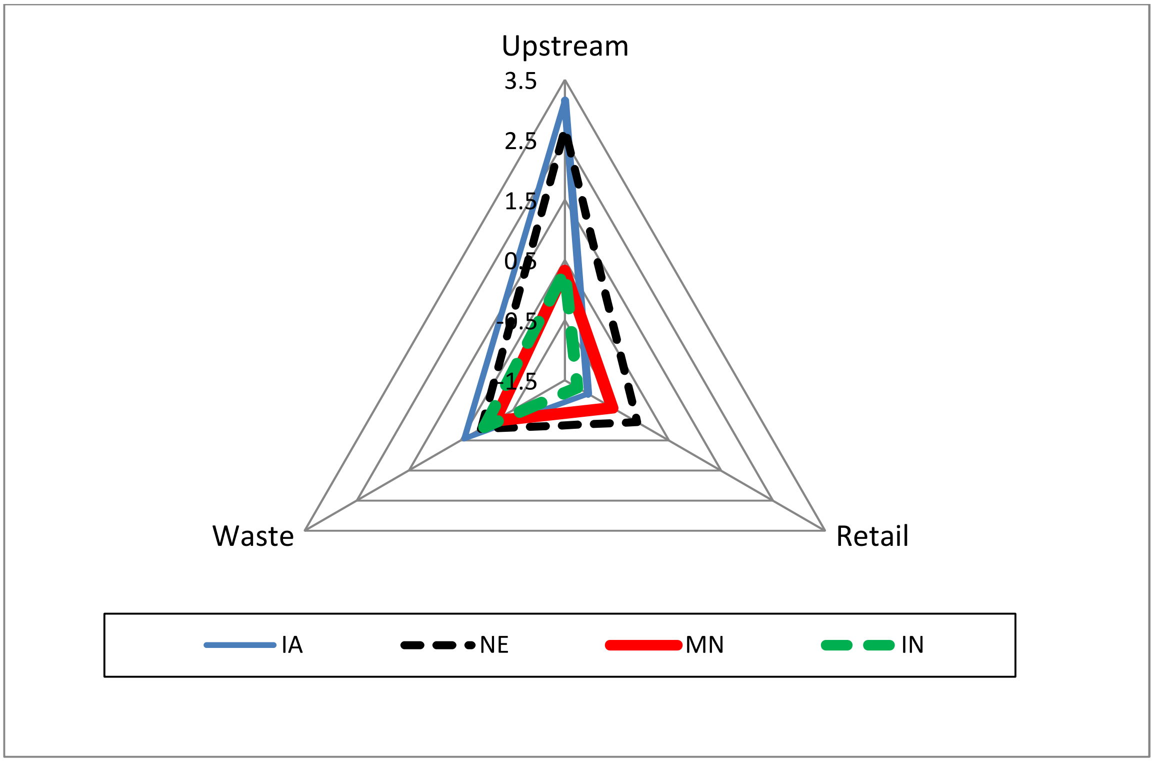


Panel a: Northeast Component Score Comparison

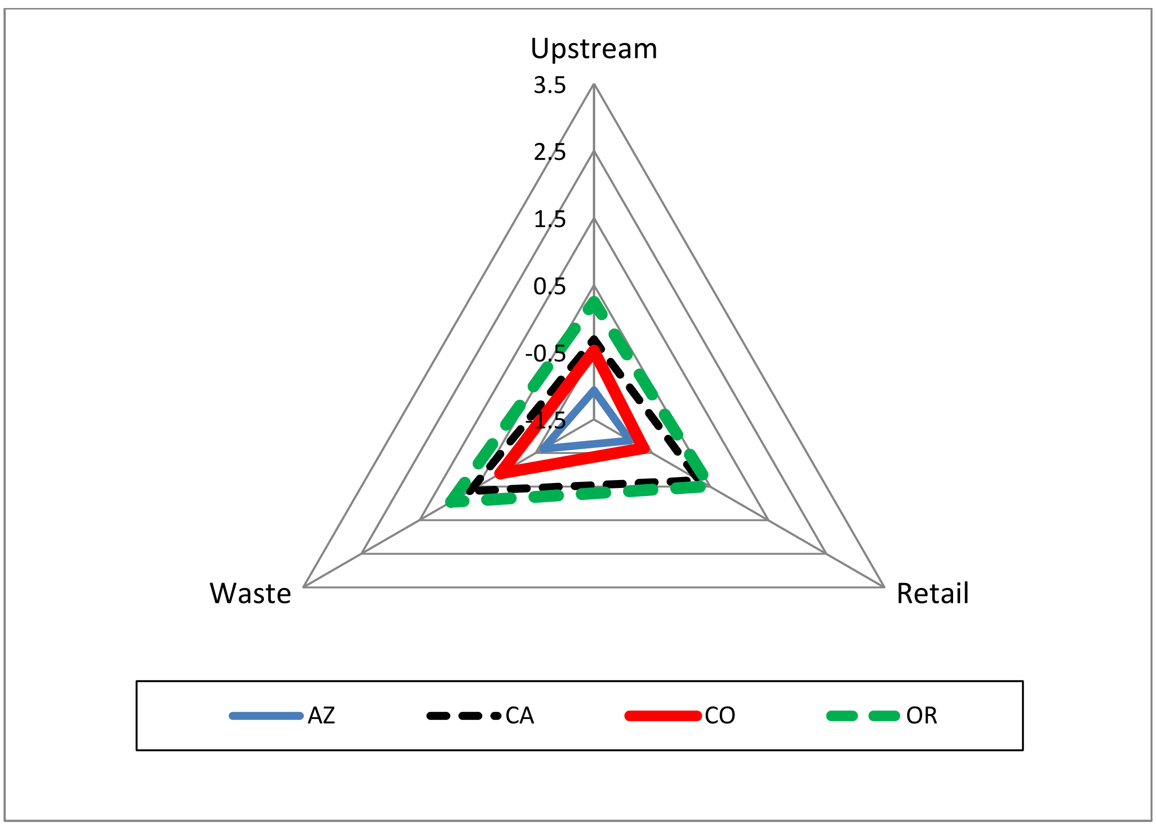


Panel b: South Component Score Comparison

Figure 5.4: Radar Charts of Four States in Each Census Region



Panel c: Midwest Component Score Comparison



Panel d: West Component Score Comparison

Figure 5.4 cont.: Radar Charts of Four States in each Census Region

State Comparison

Figure 5.5 shows the radar chart for Minnesota for 2002 and 2007. The shapes of the triangles are very similar to each other. The economic structure of the Minnesota food system did not change much over the 5 years. The triangles are skewed in the upstream activities component direction. This component has the only positive scores. The retail component has the smallest component scores. Overall the 2002 triangle is slightly larger than the 2007 triangle in every component, meaning that the food system accounted for a slightly larger percent of the state's employment in 2002.

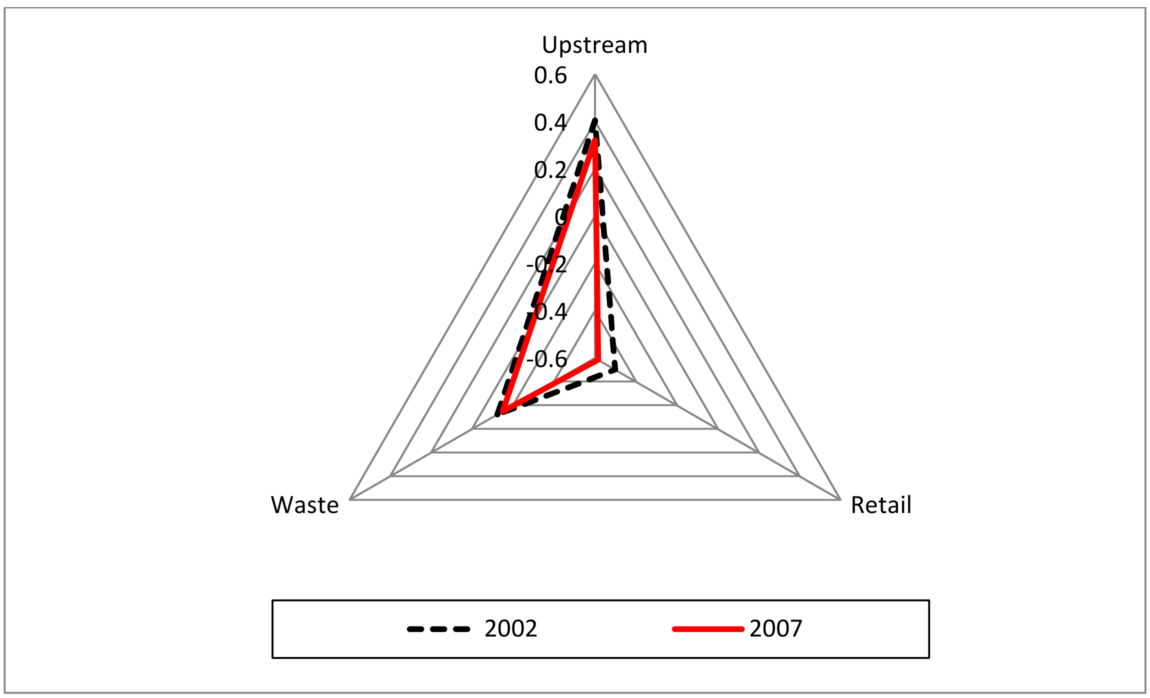


Figure 5.5: Radar chart for Minnesota for 2002 and 2007

Agricultural Production Intensity Group

Table 5.6 reproduces the variable definitions and abbreviations for the Agricultural Production Intensity group given in Chapter 2.

Table 5.6: Variable Definitions and Names for Agricultural Production Intensity Group

Variable Definition	Variable Abbreviation
Value of chemicals used per acre of agricultural land	Chemicals
Value of fertilizer used per acre of agricultural land	Fertilizer
Percent of total agricultural land in conservation	Conserve
Percent of total agricultural land that is irrigated	Irrigation
Percent of total agricultural land used for crops	Crop Land
The ratio of net farm income to agricultural sales	Income
The ratio of government payments to agriculture to agricultural sales	Payments
Percent of total agricultural sales that are crop sales	CropSales
Percent of farms that have $500,000 or greater sales	VL Farm

PCA Results. After accounting for missing values, standardizing the data, and removing outliers there were 44 observations in the 1997 group, 46 observations in the 2002 group, and 41 observations in the 2007 group.

The communalities for each variable, accumulated percent variance explained, and component loadings are presented for 1997, 2002, and 2007 in Table 5.7. The communalities for all groups are very high. Only the communality for NFI in 2007 could be considered low, but it remained in the PCA for consistency across years. The 3 retained components convey a large portion of the variance of each variable.

The three retained components also explain a large portion of the overall variance of the group. The accumulated variance explained by the three components is 72% for 1997, 75% for 2002, and 71% for 2007. A majority of the original information conveyed by the 9 variables is conveyed by the three components for all three years. The component structures for the Crop Production Intensity group are strong and interpretable. In all three years the first component is interpreted as Crop Input Supply Intensity, the second component as Government Payments and Conservation, and the third component as Irrigated Farming.

Table 5.7: PCA results for Agricultural Production Intensity Group

Variable	Communalities	Factor 1: Crop Input Intensity	Factor 2: Government Payments and Conservation	Factor 3: Irrigated Farming
		1997		
Fertilizer	0.92	0.95		
Chemicals	0.91	0.90		
Crop Land	0.94	0.82		-0.48
CropSales	0.55	0.60		
Payments	0.71		0.82	
Conserve	0.61		0.78	
VL Farms	0.70		0.81	
Irrigation	0.70			0.81
Income	0.46			0.56
Total Variance Explained		0.33	0.57	0.72

Table 5.7 continued: PCA results for Agricultural Production Intensity Group

Variable	Communalities	Factor 1: Crop Input Intensity	Factor 2: Government Payments and Conservation	Factor 3: Irrigated Farming
2002				
Fertilizer	0.92	0.96		
Chemicals	0.94	0.93		
Crop Land	0.90	0.67		-0.54
CropSales	0.71	0.76		
Payments	0.62		0.58	
Conserve	0.73		0.86	
VL Farms	0.74		0.85	
Irrigation	0.58			0.72
Income	0.63			0.79
Total Variance Explained		0.33	0.60	0.75
2007				
Fertilizer	0.92	0.95		
Chemicals	0.92	0.85		
Crop Land	0.85	0.86		
CropSales	0.51	0.55		
Payments	0.82		0.85	
Conserve	0.63		0.78	
VL Farms	0.65		0.74	
Irrigation	0.85			0.92
Income	0.25			0.45
Total Variance Explained		0.34	0.57	0.71

Component 1: Crop Input Intensity

The variables for the value of fertilizer used per acre, the value of chemicals used per acre, percent of agricultural land that is cropland, and the percent of agricultural sales attributed to crop sales heavily load on the first component. States that are heavily involved in crop production have a higher input use per acre. This structure is expected for multiple reasons. Crop producing states will have higher inputs per acre than non-crop producing states because of the nature of crop production. Fertilizer will be spread over a crop field, but would not be spread over a grazing pasture. In addition, states that are consistently involved in crop production may be causing the soil quality to decrease, causing a higher amount of fertilizer and chemical use per acre in order to maintain the level of yields.

Component 2: Government Payments and Conservation

The variables for government payments to agriculture per value of agriculture sales, percent of agricultural land in conservation programs, and the percent of total farms that have $500,000 or greater sales heavily load on the second component. This supports two major findings about aspects of government payments. First it supports that payments for land in conservation are a significant portion of government payments. In the years analyzed payments to conservation accounted for, on average, 20% of direct government payments (Effland and Stout 2011). The component also supports that government payments tend to go to states with very large farms. This continues to be true, with the largest 12.4% of farms receiving 62.2% of government payments to agriculture in 2009 (Effland and Stout 2011). There are multiple reasons this may occur. States with a higher percent of very large farms may be more important to the national food system. Also very large farms may pursue federal funds more vigorously, or be more willing and able to enroll in conservation programs.

Component 3: Irrigated Farming

The variables for the percent of agricultural land that is irrigated and the value of net farm income to agricultural sales heavily load on the third component. The variable for the percent of agricultural land used for crop production negatively loads on the third component for two of the three years. Irrigation is a laborious and expensive process. Irrigated crops are not grown when a large percent of the agricultural land is devoted to crops. This is because it is too costly to irrigate very large areas. In addition, the net farm income variable shows that farms with developed irrigated farming systems tend to have higher profits. This supports a report of the USDA, stating that although only 16% of cropland is irrigated, it accounts for half of the value of national crop sales (Schaible 2004).

Partial CPCA results. Table 5.8 presents the results of the partial CPCA. The analysis is restricted to 3 similar components in accordance with the number of retained components in each individual PCA. All models fail to be rejected, with p-values larger than 0.10. The component structure of the group is stable over time. The group represents farming practices that have been used for a long time, as well as steady government payments to agriculture. The consistency of these activities is present in the stability of the component construct.

The reduced chi squared statistic for the higher model must be minimized, in order to choose the best fitting model. The partial CPCA with three similar components has the smallest reduced chi squared statistic, meaning that it is the most appropriate model. It is, therefore, appropriate for the Crop Production Intensity group to be pooled together for PCA, as long as only the first three components are retained.

Table 5.8: Similarity Hierarchy Results for Crop Production Intensity Group

Higher Model	Lower Model	P-Value	Reduced x^2
Equality	Proportionality	0.2735	0.686
Proportionality	CPCA	0.2911	1.161
CPCA	Partial CPCA (3)	0.8508	0.615
Partial CPCA (3)	Partial CPCA (2)	0.9451	0.325
Partial CPCA (2)	Partial CPCA (1)	0.8549	0.617
Partial CPCA (1)	Unrelated	0.8946	0.599

Pooled PCA Results. The pooled PCA is performed with a pooled data set consisting of the 1997, 2002 and 2007 data sets. The pooled PCA has a total of 131 observations, after accounting for outliers.

The communalities, accumulated percent of total variance explained, and the component structure for the Crop Production Intensity group are given in Table 5.9. The pooled PCA component structure is very similar to the structure of each individual year. This is not surprising because the results of the partial CPCA showed the similarity of the three groups. The component structure for the pooled data set is used to construct component scores.

Table 5.9: PCA results for pooled Agricultural Production Intensity Group

Variable	Communalities	Factor 1: Crop Input Intensity	Factor 2: Government Payments and Conservation	Factor 3: Irrigated Farming
Fertilizer	0.90	0.94		
Chemicals	0.91	0.90		
Crop Land	0.88	0.81		-0.40
CropSales	0.50	0.58		
Payments	0.63		0.76	
Conserve	0.71		0.84	
VL Farms	0.64		0.77	
Irrigation	0.58			0.73
Income	0.50			0.69
Total Variance Explained		0.33	0.55	0.70

The component scores for each component for each state for all three years are given in Table 5.10. The component scores are created using the pooled components scores. They will again tend to be between -3 and 3 because standardized data is used.

Table 5.10: Component Scores for Crop Production Intensity Group

	1997			2002			2007		
State	**Crop Inputs**	**Payment and Cons**	**Irrigated Farming**	**Crop Inputs**	**Payment and Cons**	**Irrigated Farming**	**Crop Inputs**	**Payment and Cons**	**Irrigated Farming**
AK	-1.897	0.357	2.926	-1.785	-0.008	1.280	-1.505	-0.490	-0.217
AL	-0.460	0.625	0.090	-0.645	0.856	0.239	-0.724	1.017	-0.233
AR	0.352	0.160	1.610	0.211	0.376	1.213	0.411	0.865	1.970
AZ	N/A	N/A	N/A	-1.274	-1.270	0.903	N/A	N/A	N/A
CA	1.227	0.282	2.877	1.639	0.313	2.757	1.603	0.494	3.154
CO	-1.170	0.052	0.043	-1.161	0.208	-0.181	-1.179	0.711	0.452
CT	1.048	-1.186	-0.344	1.453	-1.429	-0.510	0.627	-1.342	0.076
DE	1.565	-1.257	-0.825	1.863	-0.981	-0.478	2.031	-0.882	0.314
FL	1.215	-0.904	2.442	1.036	-0.924	2.390	1.191	-0.704	1.774
GA	0.452	0.624	1.038	0.075	0.765	1.119	0.340	0.777	0.782
HI	N/A	N/A	N/A	-0.215	-1.600	0.960	N/A	N/A	N/A
IA	0.667	1.532	-0.122	0.732	1.385	-0.921	1.089	2.471	-0.354
ID	0.055	0.206	1.500	-0.184	0.362	1.673	-0.082	0.477	1.845
IL	1.277	0.524	-0.354	1.446	0.700	-1.165	1.858	1.332	-0.252
IN	1.211	0.211	-0.495	1.283	0.117	-1.384	1.714	0.453	-0.540
KS	-0.480	0.663	-0.487	-0.483	0.542	-1.238	-0.308	1.162	-0.556
KY	-0.107	-0.142	0.242	-0.237	0.041	-0.742	-0.315	-0.093	-0.710
LA	0.776	-0.147	0.728	0.513	0.683	-0.053	0.238	1.401	0.940
MA	0.716	-1.270	0.538	0.853	-1.371	-0.297	0.632	-1.317	0.189
MD	0.877	-0.796	-0.923	0.830	-0.312	-1.217	0.809	-0.245	-0.456
ME	0.430	-0.683	-0.898	0.203	-0.371	-0.508	0.072	-0.245	-0.042
MI	1.186	-0.170	-1.007	1.086	0.057	-1.027	1.202	0.018	-0.418
MN	0.721	0.825	-1.197	0.628	1.153	-1.166	0.779	1.913	-0.306

Table 5.10 continued: Component Scores for Crop Production Intensity Group

State	1997 Crop Inputs	1997 Payment and Cons	1997 Irrigated Farming	2002 Crop Inputs	2002 Payment and Cons	2002 Irrigated Farming	2007 Crop Inputs	2007 Payment and Cons	2007 Irrigated Farming
MO	-0.175	0.728	-0.352	-0.241	0.793	-0.962	-0.218	1.203	0.031
MS	0.179	1.048	0.785	-0.011	1.470	0.326	-0.151	2.333	0.889
MT	-1.312	1.224	-0.495	-1.426	1.729	-0.894	-1.354	1.270	-0.110
NC	0.878	0.059	0.463	0.978	0.049	-0.510	0.865	0.132	-0.435
ND	-0.219	1.583	-1.064	-0.242	1.793	-0.754	0.064	1.953	0.049
NE	-0.562	0.179	0.417	-0.471	0.175	-0.130	-0.297	0.776	0.864
NH	-0.233	-1.106	-1.139	-0.263	-0.890	-1.067	-0.357	-1.146	-0.731
NJ	1.777	-1.369	0.435	1.959	-1.406	0.346	1.978	-1.158	1.181
NM	-1.793	-0.790	0.587	-1.830	-0.595	0.250	-1.885	-0.622	0.917
NV	N/A	N/A	N/A	-1.385	-1.354	0.334	N/A	N/A	N/A
NY	0.329	-0.830	-1.603	0.143	0.039	-1.427	0.155	-0.367	-0.616
OH	0.748	0.137	0.148	0.699	0.192	-1.106	1.082	0.345	-0.766
OK	-1.043	-0.070	-0.693	-1.198	0.149	-0.227	-1.096	0.151	-0.927
OR	-0.497	-0.500	0.880	-0.430	-0.630	0.452	-0.468	-0.279	0.801
PA	0.374	-0.662	-1.231	0.275	-0.325	-1.463	0.187	0.042	-0.543
RI	N/A	N/A	N/A	1.069	-1.405	0.495	N/A	N/A	N/A
SC	0.470	0.310	0.101	0.164	0.302	-0.693	0.086	0.629	-0.687
SD	-0.966	0.387	0.120	-0.804	0.346	-0.981	-0.718	0.614	0.123
TN	0.083	-0.137	-0.514	0.107	-0.433	-1.110	0.141	-0.118	-1.426
TX	-1.208	0.425	0.329	-1.353	0.576	0.703	-1.161	1.095	0.281
UT	-1.541	-0.714	0.601	-1.537	-0.679	0.166	-1.530	-0.774	0.496
VA	-0.079	-0.697	-0.723	-0.214	-0.473	-0.922	-0.270	-0.656	-0.927
VT	-0.538	-0.989	-1.244	-0.529	-0.288	-1.749	-0.631	-0.818	-0.488
WA	0.185	0.337	0.655	0.252	0.775	0.704	0.167	1.263	1.319
WI	0.237	0.372	-1.059	0.165	0.684	-0.957	0.354	0.704	-0.270
WV	-0.861	-1.192	-1.720	-0.904	-1.216	-2.000	-1.087	-1.294	-1.563
WY	-1.757	-0.882	0.237	-1.747	-0.501	-0.613	-1.606	-0.932	-0.847

Scores for Component 1: Crop Input Intensity

The component scores for crop input intensity range from 1.78 to -1.90 in 1997, 1.96 to -1.83 in 2002 and 2.03 to -1.88 in 2007. The same four states, New Jersey, Delaware, Illinois, and California, have the highest component scores in each year. In these states primary production is consistently focused on crop production and is input intensive. The 4 states with lowest components scores, Montana, Wyoming, Utah, and New Mexico, also remain the same over the three years. In these states primary production is focused on non-crop production. As a result these states also have a low amount of inputs to their agricultural land.

Figure 5.6 shows a scatter plot comparing the 1997 component scores and the 2007 component scores for crop input intensity. Many of the states with large component scores in 1997 have larger component scores in 2007. Therefore, states with input intensive practices are increasing the amount of inputs needed. Conversely many of the states that use moderate or low levels of inputs have a smaller component score in 2007 than in 1997, showing these states moved to less input intensive practices.

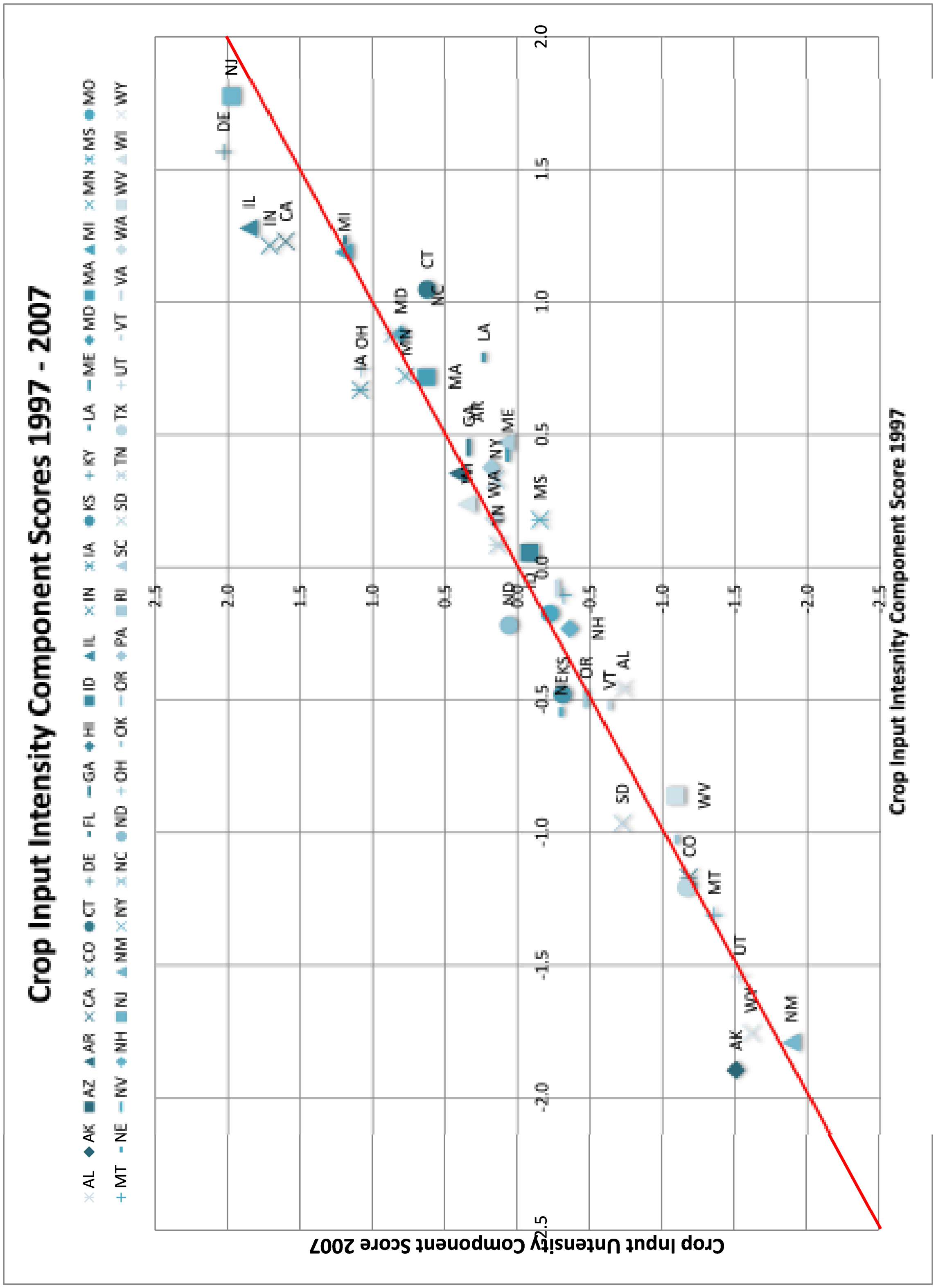


Figure 5.6: Crop Input Intensity Component Score Comparison for 1997 to 2007

Scores for Component 2: Government Payments and Conservation

The component scores for the government payments and conservation components range from 1.58 to -1.37 in 1997, from 1.79 to -1.60 in 2002, and from 2.47 to -1.34 in 2007. The 5 states with the largest scores for all three years are North Dakota, Minnesota, Iowa, Mississippi, and Montana. These states receive a large amount of government payments compared to their total agricultural sales or have a large amount of agricultural land in conservation programs. In each year, New Jersey, Massachusetts, and Connecticut are among the states with the lowest scores. This is likely due to farms being of a smaller size in the Northeast.

Figure 5.7 compares the Government Payments and Conservation component scores for 1997 and 2007. The component scores are higher in 2007 than in 1997 for almost every state. Increases in government payments to agriculture, which have increased since 1997, are the cause of this trend (Effland, Anne and Stout 2011). Only one state, Alaska, is well below the red line. Alaska may either have received less government payments to agriculture or had less agricultural land enrolled in conservation programs in 1997. Mississippi and Louisiana are well above the red line. These states have been affected more by the increased number and intensity of the natural disasters in the gulf area than other states.

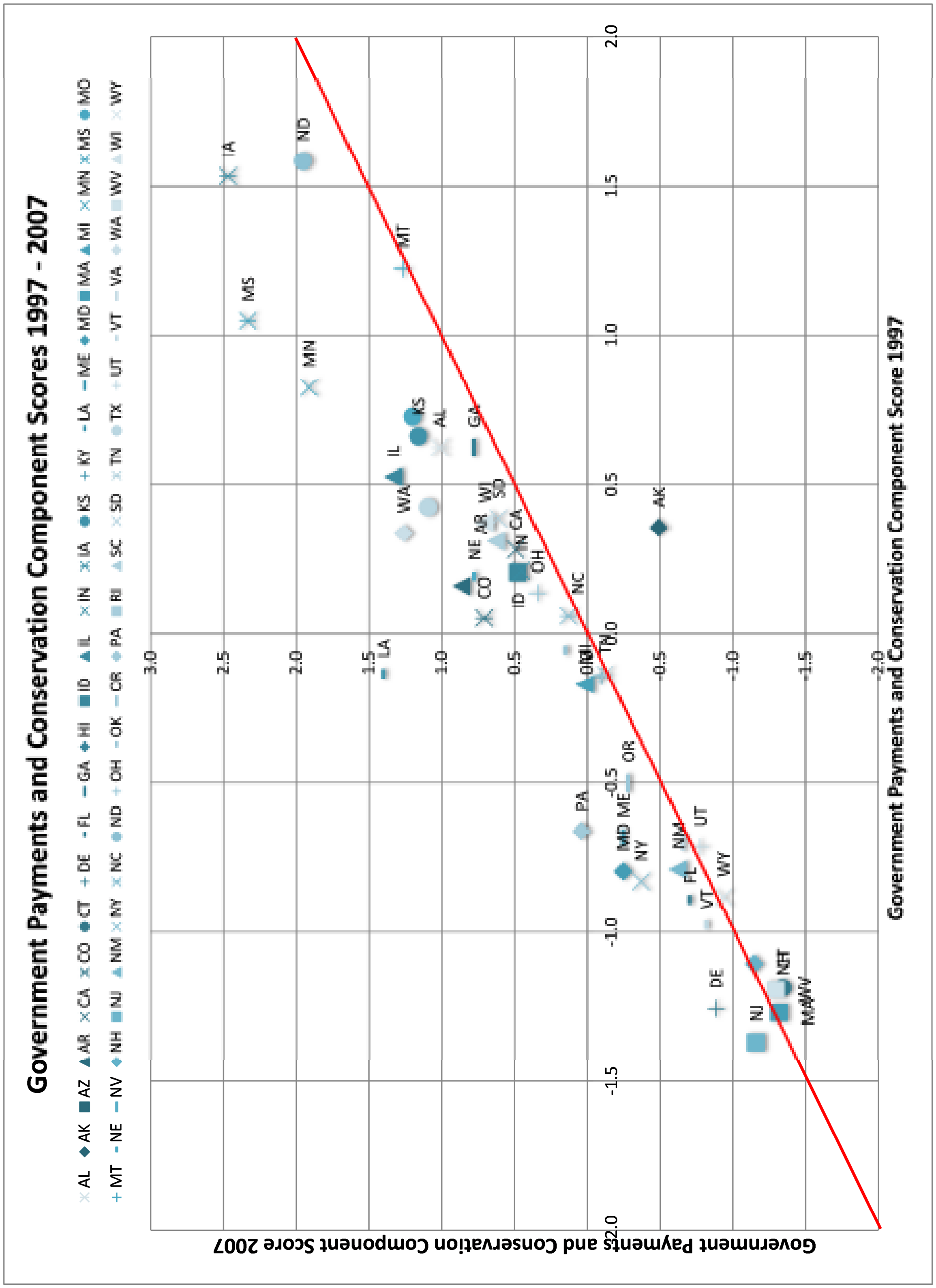


Figure 5.7: Government Payments and Conservation Component Score Comparison for 1997 to 2007

Scores for Component 3: Irrigated Farming

The component scores for the irrigated farming components range from 2.93 to -1.72 in 1997, from 2.76 to -2.99 in 2002, and from 3.15 to -1.56. The states with the highest component scores, Arkansas, California, Idaho, and Florida, remain the same over the three years. Irrigated farming is an important aspect of crop production in these states. Only West Virginia remains in the lowest 5 component scores for all three years.

Figure 5.8 compares the irrigated farming component scores for 1997 and 2007. The states on the two extremes, West Virginia and California, do not have much difference between the two years. California has by far the largest component score in both years. The arid land used for agricultural production in California requires much irrigation. There is also a large amount of variability in states that have component scores near 0.

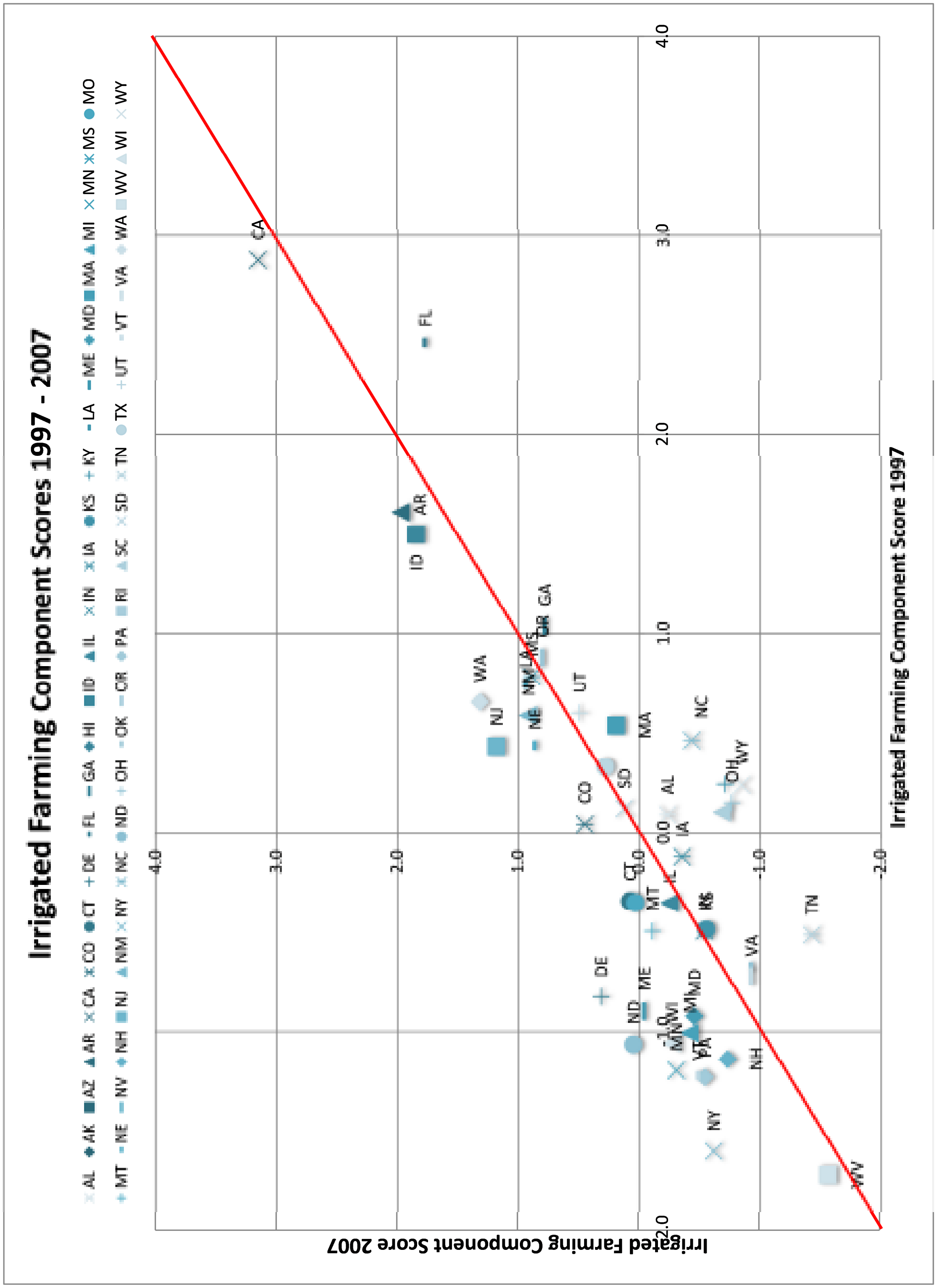


Figure 5.8: Irrigated Farming Component Score Comparison for 1997 to 2007

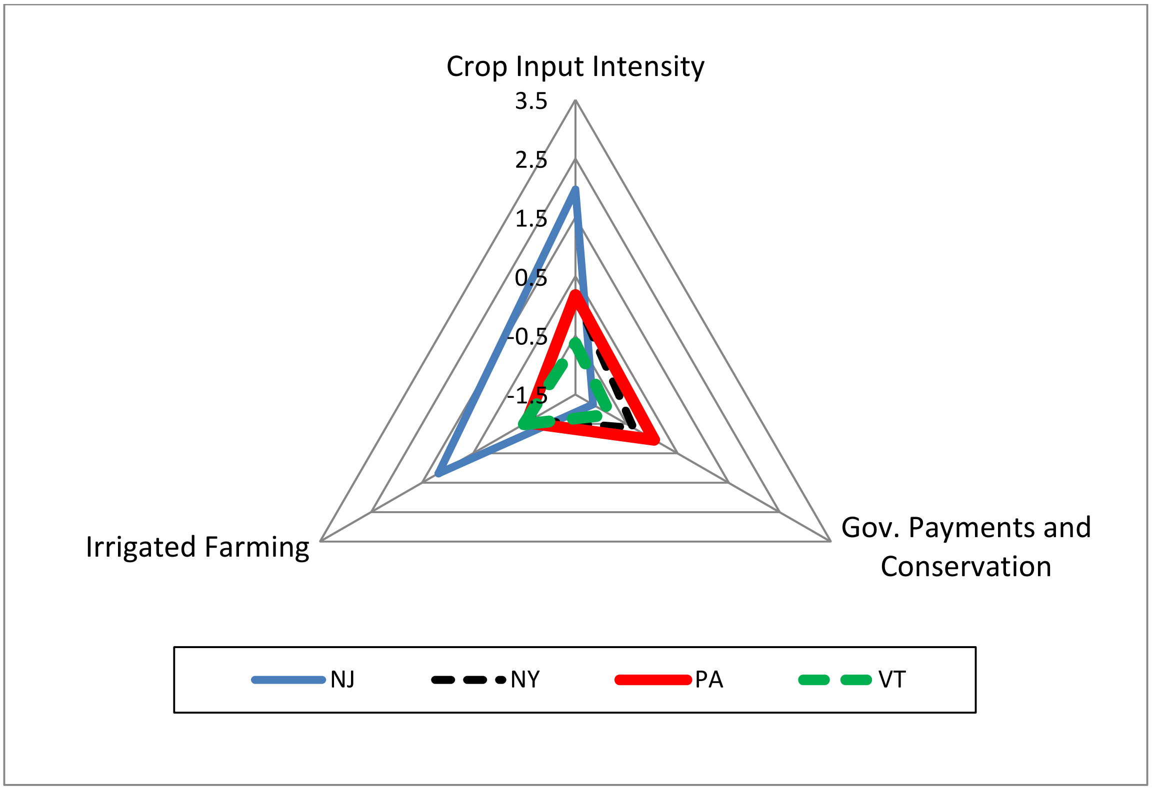
Regional Comparisons.

Northeast Region. Figure 5.9 Panel a shows the radar chart for the Northeast Region. There is no clear trend in the shape of the state triangles. New Jersey has the largest triangle. It receives few government payments, but uses very input intensive farming practices. In contrast Vermont has the smallest triangle. It receives few government payments, and does not heavily use irrigation or input intensive farming practices. New York and Pennsylvania have very similar shaped triangles. The triangles are skewed towards input intensity and government payments and conservation.

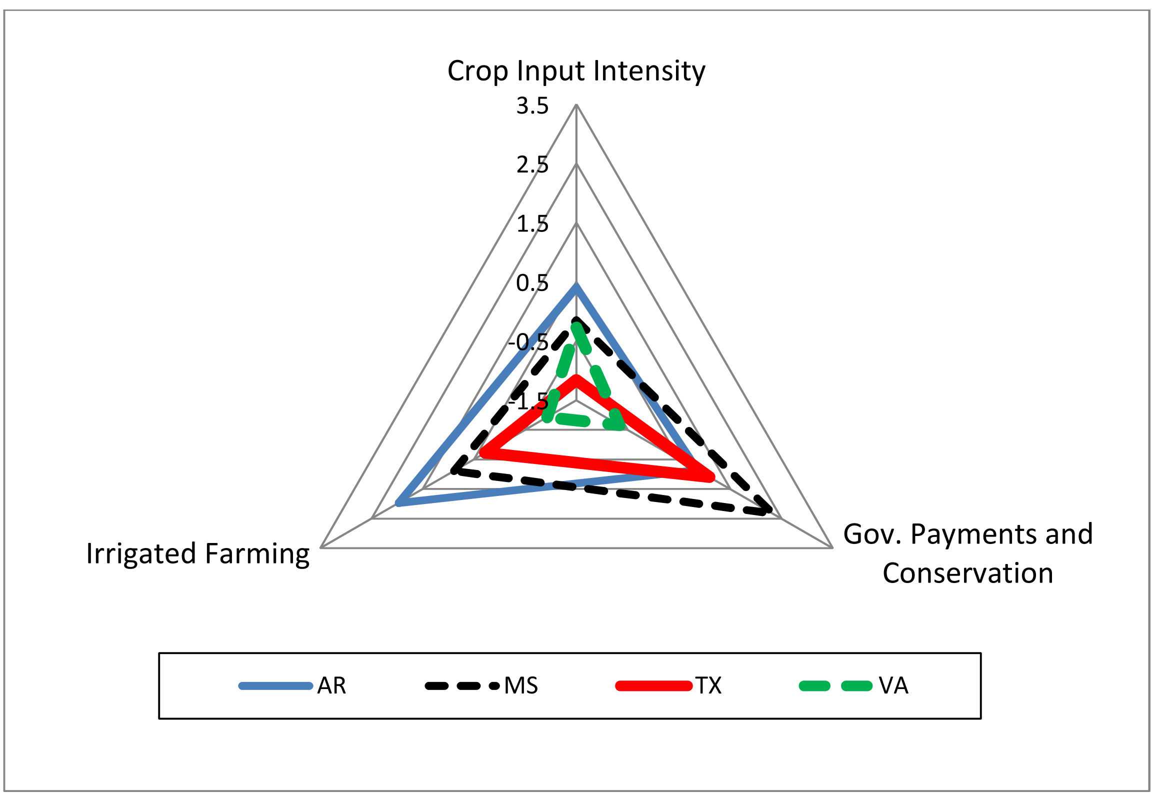
South Region. Figure 5.9 Panel b shows the radar chart for the South Region. Overall the South Region has triangles that have very broad bases. They have larger component scores in irrigated farming and government payments, than input intensity. Commonly irrigated crops are cotton and rice, which are both heavily produced in southern states. These two crops also receive a large amount of government payments (Effland and Stout 2011). In addition the area also receives government payments for disaster relief, when the gulf area is affected by hurricanes and tropical storms. Virginia's triangle is much smaller than the other southern states and also does not have a broad base. Virginia, as was also the case for the economic structure of the food system, is different from other southern states in agricultural production intensity.

Midwest Region. Figure 5.9 Panel c shows the radar chart for the Midwest Region. Overall the Midwest has triangles heavily skewed towards the government payments and conservation component and the crop input intensity component. The government payments and conservation is partially driven by the payments to corn production, especially in Iowa and Minnesota, but is also due to large amounts of payments for conservation (Effland and Stout 2011). In the Midwest high winds and sometimes dry conditions make the land exceptionally susceptible to wind erosion. In order to counteract this erosion more payments to conservation are given. The crop input intensity component is also driven by these two factors. Corn production quickly depletes the soil, so high amounts of input materials are needed to maintain high yield levels. In addition, wind erosion removes nutritious top soil causing more input materials to be needed.

West Region. Figure 5.9 Panel d shows the radar chart for the Midwest Region. Each of the states compared in the West Region, but California have small involvement in crop production. Colorado, Oregon, and Utah all have small triangles. California, however, has a very large triangle, especially in the crop input intensity and irrigated farming directions. Southern California produces cotton, rice, and lettuce as well as a number of other crops, but receives little rain. Therefore, these crops are heavily irrigated. California also uses a large amount of input materials to increase yields. The large component scores for these two components show the result of turning arid land into crop production land.

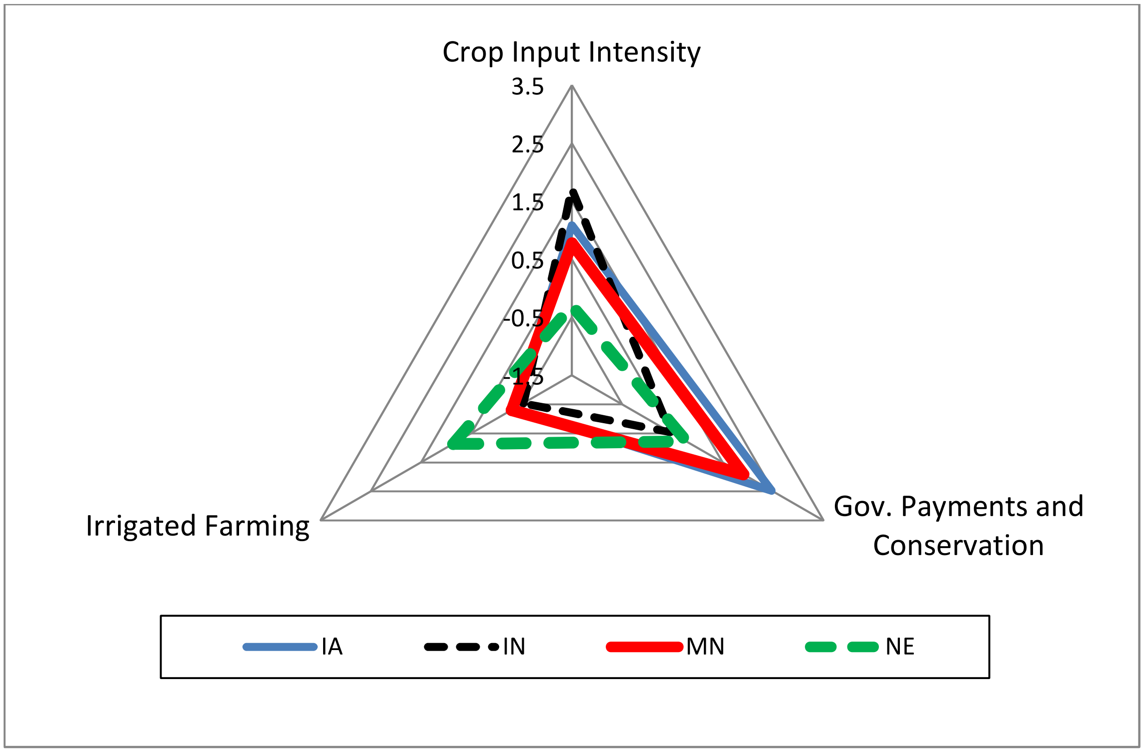


Panel a: Northeast Component Score Comparison

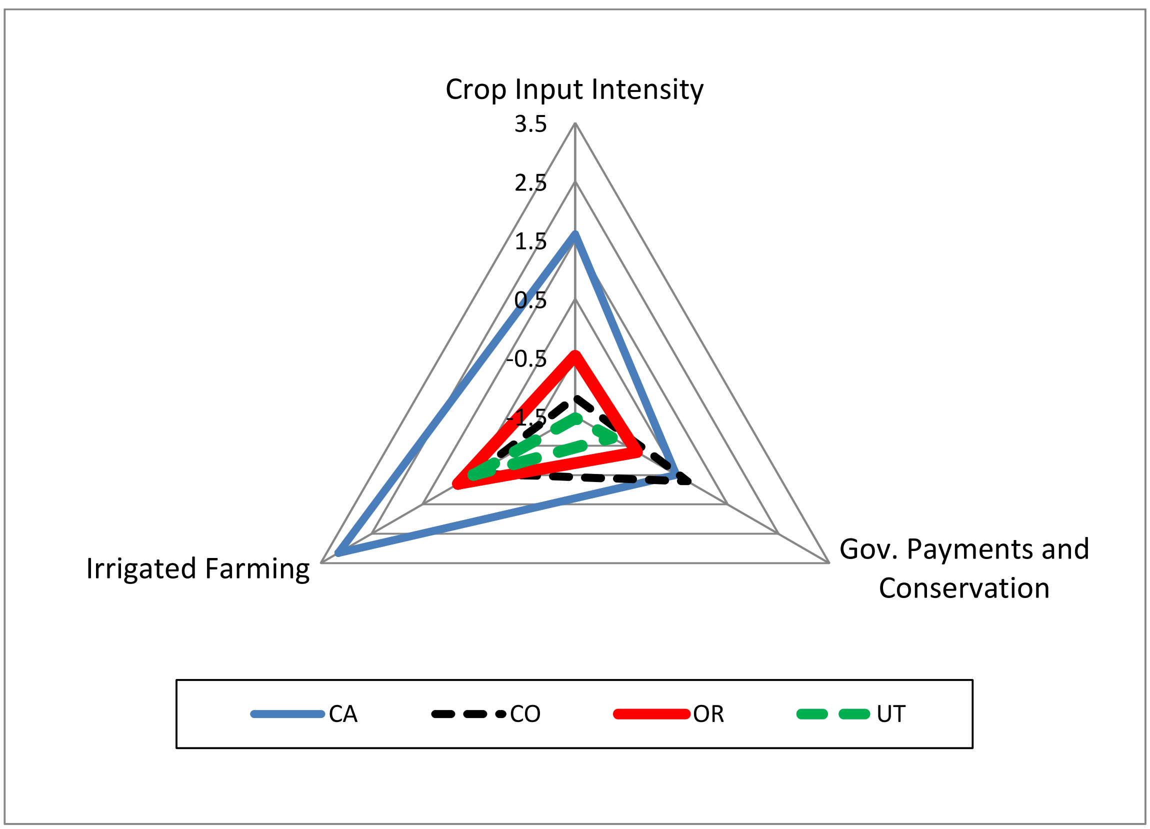


Panel b: South Component Score Comparison

Figure 5.9: Radar Charts of Four States in each Census Region



Panel c: Midwest Component Score Comparison



Panel d: West Component Score Comparison
Figure 5.9 cont.: Radar Charts of Four States in each Census Region

State Comparison

Figure 5.10 shows the radar chart for Minnesota for 1997, 2002 and 2007. The size and shape of the 1997 and 2002 triangles are very similar. However, the 2007 triangle is larger in the direction of both irrigated farming and government payments and conservation. The growth in the irrigated farming component score for 2007 can be caused by irrigated farming increasing in Minnesota or by irrigated farming decreasing in other states because the scores are comparative. The crop input intensity leg of the triangle is consistent for all three years. Minnesota's farming practices did not change much over the ten years.

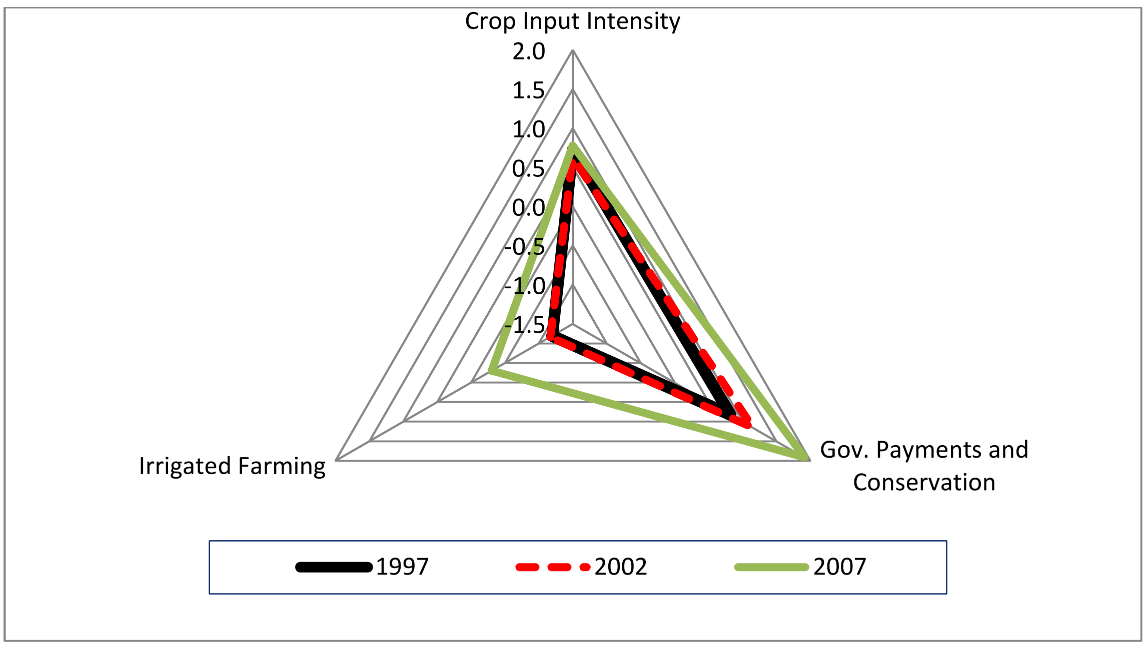


Figure 5.10: Radar chart for Minnesota for 1997, 2002 and 2007

Consumption and Health

Table 5.11 reproduces the variable definitions and abbreviations for the Consumption and Health group given in Chapter 2.

Table 5.11: Variable Definitions and Names for Consumption and Health Group

Variable Definition	Variable Abbreviation
Percent of adult population who are obese	Obese
Percent of adult population who have diabetes	Diabetes
Percent of population eligible for SNAP benefits	SNAP
Percent of food expenditures at grocery stores or supercenters	Grocery
Percent of food expenditures at convenience stores	Convenient
Percent of food expenditures at full service restaurants	Full Service
Percent of food expenditures at limited service restaurants	Limited Service
Percent of food expenditures at food service establishments	Food Service

PCA Results. After accounting for outliers there were 44 observations in the 1997 group, 46 observations in the 2002 group, and 48 observations in the 2007 group.

The communalities for each variable, accumulated percent variance explained, and component loadings are presented for 1997, 2002, and 2007 in Table 5.12. The component structures for the Consumption and Health group are not consistently strong and interpretable across all three years for each component. Only the first component is interpretable, as nutrition and health, and is consistent across the years. However, the second and third components are not clearly interpretable and change between years.

Although not consistent for all years, three retained components explain a majority of the variance of the variables and of the overall variance of the group. The communalities for all groups are very high. The lowest communality for any variable across all three years is 0.57 for PercentLSR in 2002, which is still considered high. The accumulated variance explained by the three components is 76% for 1997, 76% for 2002, and 80% for 2007.

Component 1: Health and Nutrition

The variables for the percent of state population who are obese, percent of the state population that has diabetes, the percent of the state population eligible for SNAP benefits, and the percent of food expenditures at limited service restaurants heavily load on the first component. This demonstrates the clear connection between consumption habits and major health problems. Low food security and consistent consumption at limited service restaurants are connected with serious nutrition-caused illnesses. Multiple studies already associate these variables (Periera 2005).

Component 2 and Component 3: Not Interpretable

In 1997, the variable for percent of food expenditures at food service establishments positively loads on the second component and the variable for percent of food expenditures at convenient stores negatively loads on the second component. These two variables vary in opposite directions, but there is no clear reason for the occurrence. The variable for the percent of food expenditures at grocery stores positively loads on the third component, while the variable for the percent of food expenditure spent at full service restaurants negatively loads on the third component. There is a clear reason for these variables varying in opposite directions. If there is a greater amount of expenditure on food at the grocery store, there is food consumption in the home, decreasing food consumption at restaurants.

In 2002, the variable loading on the second and third components switched from the 1997 component structure. The variable for the percent of food expenditures at grocery stores positively loads on the second component, while the variable for the percent of food expenditure spent at full service restaurants negatively loads on the second component. In addition, the variable for percent of food expenditures at food service establishments positively loads on the third component and the variable for percent of food expenditures at convenient stores negatively loads on the third component.

The second and third components are again different in 2007. The variables for the percent of food expenditures at food service establishments and full service restaurants positively load on the second component, while the variable for food expenditures at grocery stores negatively loads on the second component. Only the variable for the percent of food expenditures at convenient stores loads on the third component.

Table 5.12: PCA results for Consumption and Health Group

Variable	Communalities	Factor 1: Nutrition and Health	Factor 2: Not Interpretable	Factor 3: Not Interpretable
		1997		
Diabetes	0.85	0.88		
Obesity	0.67	0.78		
SNAP	0.61	0.77		
Limit Service	0.66	0.62		
Food Service	0.82		0.86	
Convenience	0.78		-0.83	
Grocery	0.92			0.96
Full Service	0.79			-0.67
Total Variance Explained		0.38	0.61	0.76
		2002		
Diabetes	0.84	0.89		
Obesity	0.79	0.83		
SNAP	0.70	0.78		
Limit Service	0.57	0.55		
Food Service	0.82			0.87
Convenience	0.64			-0.68
Grocery	0.88		0.93	
Full Service	0.82		-0.78	
Total Variance Explained		0.40	0.63	0.76

Table 5.12 continued: PCA results for Consumption and Health Group

Variable	Communalities	Factor 1: Nutrition and Health	Factor 2: Not Interpretable	Factor 3: Not Interpretable
		2007		
Diabetes	0.83	0.91		
Obesity	0.87	0.79		
SNAP	0.70	0.76		
Limit Service	0.71	0.83		
Food Service	0.66		0.75	
Convenience	0.95			0.95
Grocery	0.89		-0.88	
Full Service	0.79		0.82	
Total Variance Explained		0.42	0.67	0.80

Partial CPCA results. Table 5.13 presents the results of the partial CPCA. The partial CPCA is restricted to 3 similar components coinciding with the PCA results. The 4 strictest models tested are rejected because their p-value are less than 0.1. The partial CPCA model with 2 similar components and the partial CPCA model with 1 similar component failed to be rejected. The variability between the component structures of the individual years is represented in the rejection of the stricter models. The instability is caused by the large variability in consumption patterns over time.

With two models failing to be rejected, the model with the smallest reduced chi squared statistic is the best fitting model. The partial CPCA with 1 similar component has the smallest reduced chi squared statistic. It is appropriate for the Consumption and Health group to be pooled together for PCA, as long as only the first component is analyzed.

Table 5.13: Similarity Hierarchy Results for Consumption and Health Group

Higher Model	Lower Model	P-Value	Reduced x^2
Equality	Proportionality	0.0061	5.094
Proportionality	CPCA	0.0078	1.292
CPCA	Partial CPCA (3)	0.0103	1.573
Partial CPCA (3)	Partial CPCA (2)	0.0414	1.156
Partial CPCA (2)	Partial CPCA (1)	0.1128	0.855
Partial CPCA (1)	Unrelated	0.6991	0.674

Pooled PCA Results. The pooled PCA is performed with a data set consisting of the 1997, 2002 and 2007 data sets, giving the pooled PCA a total of 140 observations.

Table 5.14 shows the communalities, accumulated percent of total variance explained, and the component structure for the Crop Production Intensity group. The overall pooled PCA component structure is similar to the structure of 1997 PCA. However, only the first component is analyzed in accordance with the partial CPCA results. It is also the only component that is able to be consistently interpreted across all three years. The same variables that load on the first component for each individual PCA load on the first component for the pooled PCA, allowing for the same interpretation for the pooled PCA.

Table 5.14: PCA results for pooled Health and Consumption Group

Variable	Communalities	Factor 1: Nutrition and Health	Factor 2: Not Interpretable	Factor 3: Not Interpretable
Diabetes	0.90	0.93		
Obesity	0.80	0.90		
SNAP	0.66	0.75		
Limit Service	0.48	0.56		
Food Service	0.72		-0.74	
Convenience	0.75		0.83	
Grocery	0.95			0.95
Full Service	0.80			-0.75
Accumulated Percent Variance Explained		0.36	0.63	0.76

Table 5.15 shows the component scores for the first component for each state for all three years. The component loadings for the first component are used to create component scores, which can be compared across states and time. Usually standardized data causes component scores to be between values of -3 and 3.

Table 5.15: Component Scores for Nutrition and Health Component

State	1997	2002	2007	State	1997	2002	2007
AL	0.125	1.184	2.078	MT	-1.647	-0.882	-0.351
AK	-0.631	0.099	1.005	NE	-1.021	-0.252	0.323
AZ	-1.686	0.461	0.939	NV	-1.541	-0.481	0.310
AR	-0.583	0.576	1.337	NH	-1.828	-1.152	-0.127
CA	-0.608	-0.170	0.530	NJ	-0.935	-0.483	0.432
CO	-1.785	-1.243	-0.705	NM	-0.626	0.047	1.007
CT	-1.041	-0.690	0.038	NY	-0.585	0.057	0.859
DE	-0.556	-0.212	0.840	NC	-0.518	0.418	1.583
FL	-1.096	-0.369	0.318	ND	-1.564	-0.617	0.044
GA	-0.747	0.537	1.777	OH	-0.698	0.393	1.223
HI	-0.795	-0.452	0.203	OK	-0.478	0.714	1.813
ID	-1.459	-0.585	0.212	OR	-0.785	-0.014	0.703
IL	-0.194	0.281	1.178	PA	-0.705	0.134	0.934
IN	-0.408	0.356	1.182	RI	-1.044	-0.514	0.161
IA	-0.974	-0.399	0.397	SC	-0.616	0.929	1.749
KS	-1.382	-0.229	0.571	SD	-1.413	-0.285	0.251
KY	0.056	0.932	2.032	TN	-0.468	1.121	2.391
LA	0.351	1.363	2.403	TX	-0.318	0.739	1.702
ME	-1.115	-0.174	0.624	UT	-1.157	-0.766	0.023
MD	-0.579	-0.325	0.689	VT	-1.223	-0.876	-0.268
MA	-1.181	-0.756	0.118	VA	-0.870	-0.038	0.567
MI	-0.287	0.520	1.427	WA	-1.217	-0.294	0.527
MN	-1.341	-0.692	-0.013	WV	0.468	1.546	2.323
MS	0.652	1.793	2.815	WI	-1.421	-0.658	0.086
MO	-0.487	0.491	1.586	WY	-1.663	-1.003	-0.302

Scores for Component 1: Nutrition and Health

The component scores for the nutrition and health component range from 0.65 to -1.83 in 1997, 1.79 to -1.24 in 2002 and 2.81 to -0.70 in 2007. The range shifting to higher values over time shows that overall, in the United States, nutrition and health is becoming poorer over time. The same four states, Mississippi, West Virginia, Louisiana, and Alabama, have the highest component scores in each year. The south has higher poverty and poorer nutrition and health compared to other regions. The 4 states with lowest components scores, Montana, Wyoming, Colorado, and New Hampshire, also remain the same over the three years. Vermont is also in the lowest 5 for 2002 and 2007. The Midwest and the Northeast have higher food security and better nutrition and health than the other regions.

Figure 5.11 compares the component scores for 1997 and 2007for the health and nutrition component. Each state has a significantly higher component score in 2007 than in 1997. Considering the variables forming the health and nutrition component, a higher score is a negative aspect. The increase in diabetes rates, obesity rates, and consumption of low nutrition food is a national problem. Although each state's 2007 component score is much higher than their 1997 score, many of the states with the highest component scores and greatest growth in component score are in the south region. Tennessee has the largest difference between the two years. It has a component score of -0.468 in 1997, but has a component score of 2.391 in 2007.

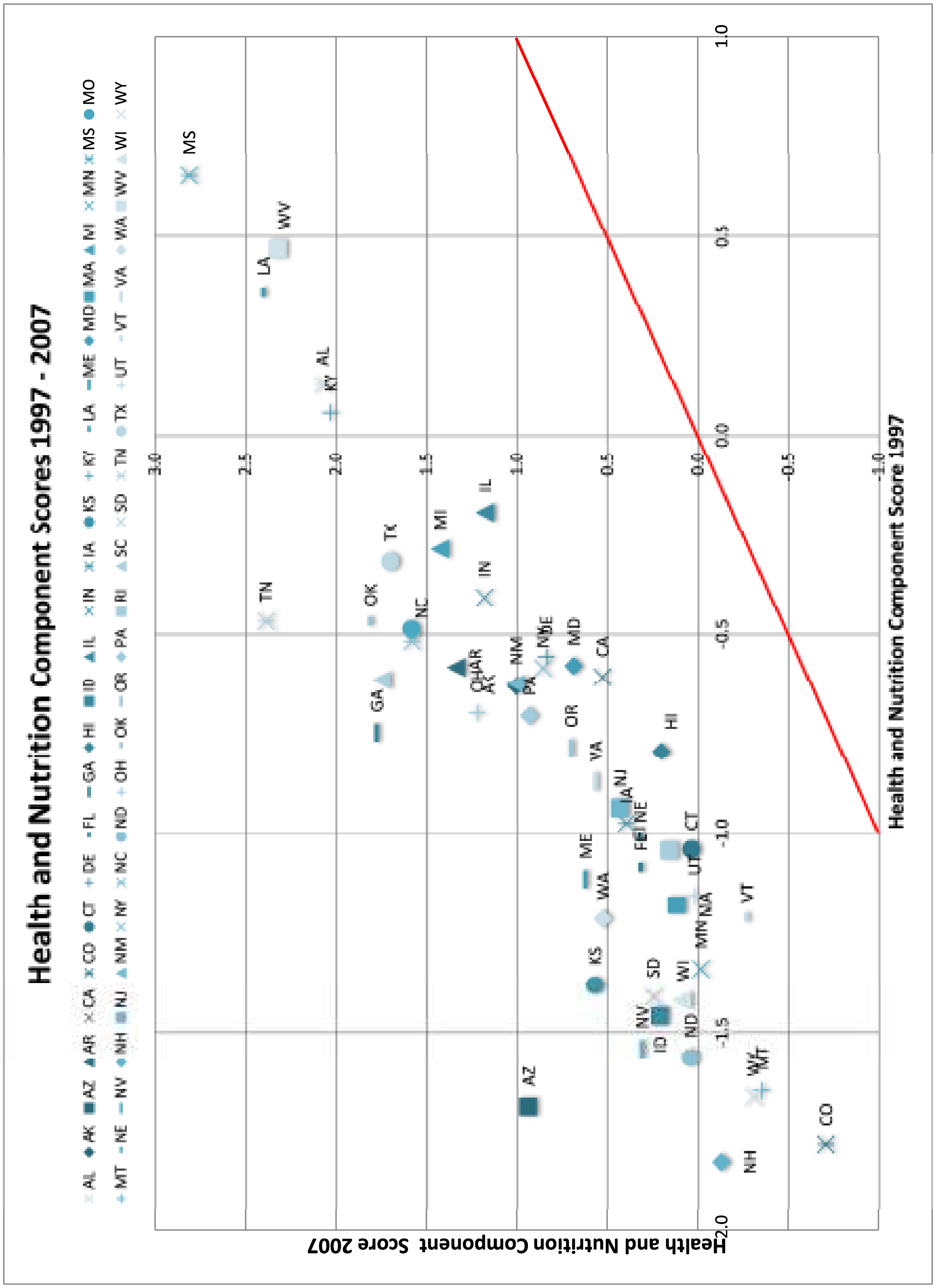


Figure 5.11: Comparison of Health and Nutrition Component Scores for 1997 and 2007

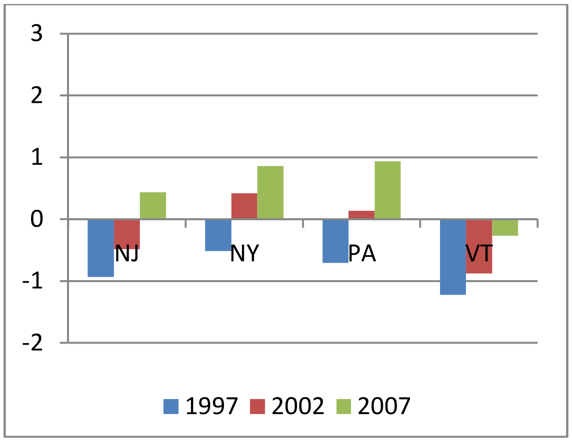
Regional Comparisons.

Northeast Region. Figure 5.12 Panel a shows the bar chart for the first component scores for all three years for the Northeast Region. There is a clear increasing trend in the size of each bar for every state. Each of the four states start with negative values, but all the states, except for Vermont, have positive component scores by 2007. While Vermont's component score is increasing, it remained negative for all three years. It also is one of the lowest component scores nationally for all three years. Compared to the other three regions of the United States the Northeast is in the middle. Overall it has higher component scores than the Midwest and West, but has lower component scores than the South. The overall food security, health, and nutrition are better than in the South, but worse than the other two regions.

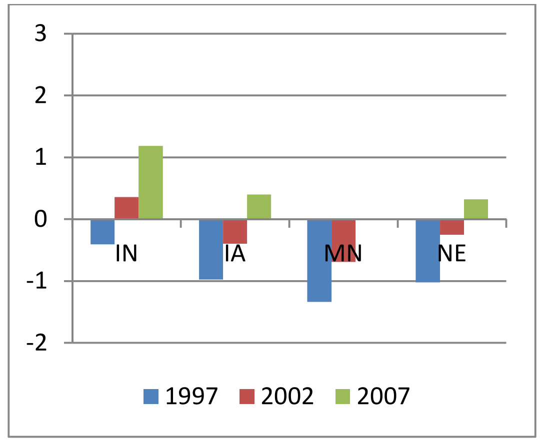
South Region. Figure 5.12 Panel b shows the bar chart for the South Region. Overall the South Region has large component scores across all three years. The component scores are also increasing for each state over time. Of the 4 states in the South Region, three have the highest values in 2007 of all 16 states used in the regional comparisons. Overall the South Region has lower food security, health, and nutrition. Mississippi has very large values over all three years. In contrast Virginia has lower values that are comparable to those for states in other regions. Although Virginia is in the South Region it has better food security, health and nutrition, than other states in the region.

Midwest Region. Figure 5.12 Panel c shows the bar chart for the Midwest Region. Again the states in this region follow the national trend, with component scores increasing over time. However, the scores are much lower than the South and Northeast Regions. Although all 4 states have positive component scores by 2007 only 1 scores is greater than 1. Iowa, Minnesota, and Nebraska have low values for all 3 years showing these states have relatively high food security, good health, and nutrition. Overall the Midwest Region has the second lowest component values of any region.

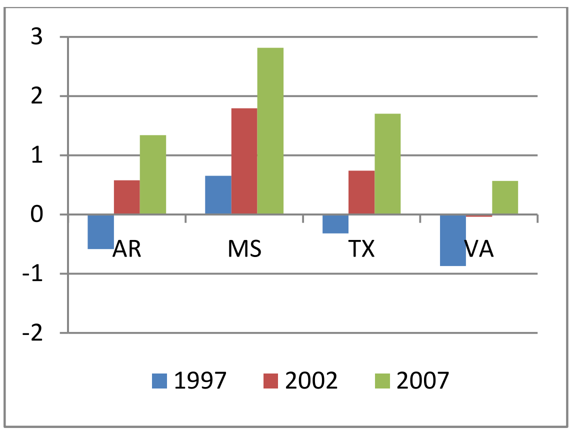
West Region. Figure 5.12 Panel d shows the bar chart for the West Region. Overall the West Region has the lowest component scores of any of the other regions. When the lowest component scores across all three years are compared 3 of the 4 four states with the lowest scores are from the West Region. In addition, all the states in the comparison have negative component scores prior to 2007. Colorado has a negative component score for all three years, and has the lowest component score of any state for 2002 and 2007. Colorado has the highest food security, and best nutrition and health of any state. Of the states compared from the West Region, California and Oregon have the highest component scores, however when compared to other regions they are still low. Despite the relatively low scores, each of the states still follows the national trend of consistently increasing component values. The West may have the highest food security, and best nutrition and health of any region, but is moving towards lower food security, poorer nutrition, and health.



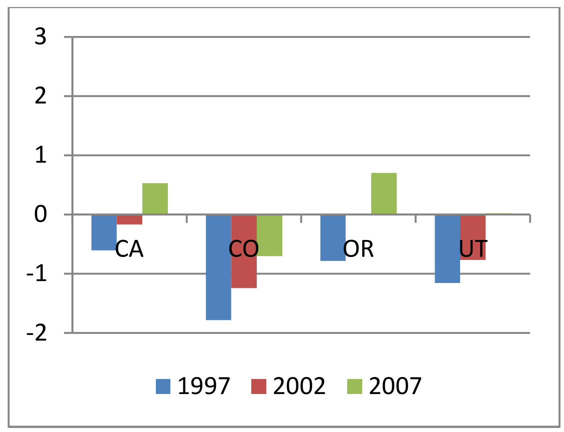
Panel a: Northeast Component Score Comparison



Panel c: Midwest Component Score Comparison



Panel b: South Component Score Comparison



Panel d: West Component Score Comparison

Figure 5.12: Bar charts comparing health and nutrition component scores

Chapter 6: Summary and Conclusion

Group Summaries

The first group analyzed, the Economic Structure of the Food System, has three retained components. All three have clear interpretations. This reduces the information in nine indicators into three scores. The component structures between 2002 and 2007 are shown to be similar to three components. Comparing components scores using the pooled component structure shows that the Northeast Region has larger employment in retail activities, while the Midwest Region has larger employment in upstream activities. Both the South and West Regions have lower overall levels of employment in the food system. However, the South Region is very evenly distributed between upstream, retail, and waste activities, while the West has higher upstream activities component scores. For the states most involved in upstream activities employment increased over time, whereas for the states most involved in retail employment decreased over time. There is no clear trend for employment in waste activities.

The analysis of the Economic Structure of the Food System showed that the component structure of the group was very stable over time. Although the partial CPCA model with three components was the best fitting model, the model for equality of covariance matrices failed to be rejected. In addition, when comparing the component scores between 2002 and 2007, there were only minimal differences. This shows that the data values remained similar over the 5 years as well. The stability was particularly true for the upstream production activities component scores. Upstream food system activities are not transient activities. Once well-established, they will remain in their location and not growing extremely quickly.

The second group analyzed, the Agricultural Production Intensity Group, also has three retained components, each with a clear interpretation. Again, this reduces the information in nine indicators to three scores. The component structures between 1997, 2002, and 2007 are shown to be similar to three retained components. The South and Midwest Regions receive a large amount of government payments to agriculture, as well as participate in conservation programs. Also, the West and South are more heavily involved in irrigated farming than the other regions. Finally, while each region has a state with an exceptionally high crop input intensity, the Midwest uses the most crop input materials. For the states with the highest crop input intensity, the amount of inputs used has increased over time. In addition, the states receiving the most government payments, also received more over time.

The Agricultural Production Intensity Group allows for a partial picture of agriculture practices in each state. The component scores for the Crop Input Intensity and Irrigated Farming component can be used in conjunction with other indicators, such as crop yield rates, water quality and availability, or erosion rates to assess correlations between input intensity and irrigation and other areas of interest. In addition, the component scores for the Government Payment and Conservation component can be used in conjunction with other indicators to assess the flow of government payments to agriculture.

The final group analyzed, the Consumption and Health Group, has three retained components for each individual PCA. However, only the first component in each PCA has a clear and consistent interpretation. In addition, the component structures for 1997, 2002, and 2007 are only similar up to the first component. Considering the first component the information in four variables are reduced to one component score. Comparing component scores for the first component only, shows that South Region has the lowest food security and the poorest health and nutrition of the four regions. In contrast, the West Region has the highest food security and best health and nutrition of the four regions. Every one of the 50 states' component score increases over time. Although certain regions of the country are better than others, this implies that the entire U.S. has a negative trend in regard to food security, health, and nutrition.

The Health and Consumption group is different from the other two groups. The component structure for the group is much less stable over time and the data values producing the component scores increase for every single state. The unstable component structure is caused by the group attempting measure consumption patterns. There are many factors that affect consumption patterns causing them to change more rapidly than the structure of the food system or farming practices. In addition, the nation-wide rise in component score puts health and nutrition further in the spotlight. Multiple studies have shown health and nutrition to be decreasing over time in the United States and the component scores further highlight those negative trends.

Analysis Challenges

Although this thesis shows that it is possible to achieve meaningful data reduction there are multiple complications that can occur during the process. The first major complication is that a PCA's component structure is very sensitive to the variables that are included. With 50 subjects, the subjects to variable ratio requires no more than ten variables be included in a PCA at one time. Consequently, many group combinations must be analyzed, possibly differing by only one variable, before a meaningful and interpretable component structure is achieved.

In addition to the procedural complications caused by the subjects to variable ratio, there is a conceptual complication. The holistic nature of the food system is a main reason for the complexity of the indicators set. However, by separating the indicator set into groups for the PCA analysis the holistic nature is decreased. It is possible for meaningful connections to be missed because the number of variables able to be included in a single PCA is limited.

A second major complicating factor is partially shown by the Health and Consumption Group. Although each year's component structure for a single group may have meaningful interpretations, they may not be consistent over time. Food system comparisons across time cannot be made with varying component structures. The component scores are a summed product of the standardized data value and the variables component loading. Any changes in component scores are attributable to changes in the data value if a single component structure is used across all years. However, if the component structure changes over time the component scores are not comparable because the change in the score may be due to the data value, the variables component loading, or likely both. Consequently, only groups that are shown to have consistent component structures over time can be used to compare the food system over time.

Despite its limitations, the requirement to have consistent component structures over time has a major benefit. With the component structures shown to be stable over time, the pooled component structure can be used with future indicator sets to create component scores without performing the entire analysis. This allows for comparisons of food systems to continue using fewer resources. In addition, a known consistent set of indicators can be collected, eliminating the lengthy task of creating the indicator list itself.

Conclusion

The food systems' interconnectivity with almost every aspect of society makes accurately characterizing it very important. This same interconnectivity also makes the problem of accurately characterizing the food system very complex. Indicator sets that attempt to capture the holistic nature of the food system and are repeated across location and time to allow for comparisons and stability testing are inevitably very large. In addition, the results of this thesis show that it is possible to characterize the information presented by groups of individual indicators by component scores, although the process is very difficult. Through PCA and partial CPCA techniques, selected groups of indicators for each state over three years are reduced in dimensionality and shown to be stable over time. This then allows for states to be compared nationally and regionally on specific aspects of their food systems. It also allows for the comparison of food system changes over time.

The ability to compare aspects of the food system through component scores can help to inform policy decisions in multiple ways. The most basic way is by using the component scores to paint a picture of the current food system. A few scores can allow a policy maker to assess how one state compares with other states. In this way, particular issues can be highlighted and addressed quickly. The component scores also allow for states to assess policy effectiveness over time. A policy is put in place with a particular expected outcome. Tracking and comparing the component scores over time will allow policy analysts to see if the intended outcome was achieved, was not achieved, or had a different outcome than expected. In addition, if similar policies are enacted in multiple states or a federal policy is enacted for all states, the component scores allow for a comparison across states and across time to analyze the relative effectiveness of the policy. The differences in the states can then be studied to address the differences in the policy's outcomes.

Works Cited

Anderson, Molly. 2009. "Charting Growth to Good Food." *Wallace Center at Winrock International*.

Antsey, Chris. 2010. “Measure What Matters: The search for farmer’s triple bottom line.” Pulse Canada.

Aramyan, Lusine, and Christien Ondersteijn. 2006. "Performance Indicators in Agri-food Production Food Chains" *Springer-Netherlands.*

Booth, Mary. April 2006. “Dead in the Water.” Environmental Working Group. http://www.ewg.org/reports/deadzone

Borgognone, María G. Bussi, Javier. Hough, Guillermo. 2001. “Principal component analysis in sensory analysis: covariance or correlation matrix?” *Food Quality and Preference*, Volume 12.5. 323-326.

Bryant and Yarnold. 1995. “Principal Component Analysis and Exploratory and Confirmatory Factor Analysis.” American Psychology Association Books.

Bureau of Labor Statistics. 2009. “Consumer Expenditures-2010.” United States Department of Labor Web. http://www.bls.gov/news.release/ cesan.nr0.htm

Casagrande, Sarah Stark. 2007. “Have Americans Increased Their Fruit and Vegetable Intake?” *American Journal of Preventative Medicine*. 32.4

Csomós, Elemér., K. Héberger, L. Simon-Sarkadi, 2002. “Principal Component Analysis of Biogenic Amines and Polyphenols in Hungarian Wines.” *Journal of Agriculture and Food Chemistry*, 50 (13). 3768–3774.

"Environmental Indicators for Agriculture: Methods and Results." 2001. *Organization for Economic Cooperation and Development* Print.

Effland, Anne and Stout, James. 2011. “Government Payments and the Farm Sector”. *Farm and Commodity Policy Briefing Rooms.* http://www.ers.usda.gov/Briefing/ FarmPolicy/gov-pay.htm

Energy Information Agency. *U.S. Census Regions Map.* 2000. http://www.eia.gov /emeu/reps/maps/us_census.html

Feenstra, Gail. 2005. "Proposed Indicators for Sustainable Food System." *Ecotrust and Roots of Change Council*.

Finkelstein, EA, Trogdon, JG, Cohen, JW, and Dietz, W. 2009. “Annual medical spending attributable to obesity: Payer- and service-specific estimates.” *Health Affairs* 28(5): 822-831.

Flury, Bernhard. 1988. *Common Principal Components and Related Multivariate Models*. 1st ed. New York: Wiley and Sons Inc.

Food and Agriculture Organization of the United Nations. 2006. “Livestock’s Long Shadow: Environmental Issues and Options.”

Garson, G. David. 2011. *Factor Analysis: Statnote.* 5th ed. North Carolina State Publishing.

Gillespie, Ardyth H. and Gilbert W. Gillespie Jr. 2000. “Community Food Systems: Toward a Common Language for Building Productive Partnerships.” *Cornell Community Nutrition Program, Division of Nutritional Sciences, Cornell University*.

Gorsuch, R.L. 1983. “Factor Analysis” Lawrence Erlbaum Publishing. Hillsdale, NJ

Gradwell, Shelly, Matt Russell, and Wendy VanDyke Evans. 2002. "Audobon County Food System Atlas: Exploring Community Food Systems." *Leopold Center for Sustainable Agriculture*.

Guttenstein, Elizabeth. Sciaballa, Nadia. Loh, Jonathan. and Courville, Sasha. 2010. “A Conceptual Framework for Progressing toward Sustainability in the Agriculture and Food Sector.” United Nations Food and Agriculture Organization.

Hardle, Wolfgang, and Leopold Simar. 2007. “Applied Multivariate Statistical Analysis.” Second Edition. Berlin: Springer, 215-269.

Hatcher, Larry. 1994. “A step by step approach to using SAS system for factor analysis and structural equation modeling.” SAS Institute. Cary, NC.

Hoang, Nguyen Huy. 2009. "Regional Welfare Disparities and Regional Economic Growth in Vietnam." *Wageningen University*.

Hotelling, H., 1933. “Analysis of a Complex of Statistical Variables Into Principal Components.” *Journal of Educational Psychology*, v. 24. 417-441 and 498-520.

Hutcheson, Graeme. Sofroniou, Nick. 1999. “The multivariate social scientist: Introductory statistics using generalized linear models.” Sage Publications. Thousand Oakes, CA.

Jackson, J. Edward. 1991. “A User's Guide to Principal Component Analysis” New York: Wiley and Sons: Wiley-Interscience.

Jerardo, Alberto. 2004. “The U.S. Ag Trade Balance.” Economic Research Service USDA. *Amber Waves.* http://www.ers.usda.gov/AmberWaves/February04/Features/USTradeBalance.htm.

Joliffe, I. T. 2002. “Principal Component Analysis.” 2nd ed. New York: Springer Statistics.

Katchova, Ani L., and Timothy A. Woods. 2011. "The Effectiveness of Local Food and Marketing Strategies of Food Cooperatives." *Journal of Economic Literature* Q13.

Katsuura, Masaki, and Allan P. Layton. 2001. "Examining Similarities among Phases of Business Cycles-Is the 1990's US Expansion Similar to the 1960s'?" *J.Japan Statist. Society* 31.2 129-52.

King, Robert P. & Jacobson, Elaine M., 2001."A Factor Analysis Of Supermarket Management Practices," 2001 Annual meeting, August 5-8, Chicago, IL. American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association).

King, Shawn and Feenstra, Gail. 2001. “Placer County Food Shed Report.” *UC Sustainable Agriculture Research and Education Program.*

Mark Pereira, Alex I. Kartashov, Cara B. Ebbeling, Linda Van Horn, Martha L. Slattery,David R. Jacobs, Jr., David S. Ludwig. 2005. “Fast-food Habits, Weight Gain, and Insulin Resistance (The CARDIA Study): 15-Year Prospective Analysis”. The Lancet, January.

Norusis, Marija J. 2005. “SPSS 13.0 Statistical Procedures Companion.” SPSS Inc., Chicago.

Pearson, K. 1901. "On Lines and Planes of Closest Fit to Systems of Points in Space". *Philosophical Magazine* 2 (6): 559–572.

Reig-Martinez, Ernest, Jose Gomez-Limon, and Andres Picazo-Tadeo. 2011. "Ranking Farms with a Composite Indicator of Sustainability." Agricultural Economics 42 561-75.

Schaible, Glenn. 2004. “Irrigation, Water Conservation, and Farm size in the Western United States.” Amber Waves Economic Research Service. http://www.ers.usda.gov/amberwave/june04/findings/IrrigationWestern.htm

Schott, James R. 1999. "Partial Common Principal Component Subspaces." *Biometrika* 86.4 899-908.

Schott, James R. 1988. "Partial Common Principal Component Subspaces in Two Groups."*Biometrika* 75.2 229-36.

Sobal, Jeffery, Laura Kettel Kahn, and Carole Bisogne. 1988. "A Conceptual Model of the Food and Nutrition System." *Social Science.* 47.7 853-63.

Wu, Mae. Sass, Jennifer. 2010. “Still Poisoning the Well.” National Resource Defense Council.

Appendix A: Standardized data for Employment Structure Group

A.1: Employment Structure Group 2007

State	Dist.	Input Supply	Prim. Prod.	Process	Retail	Waste	Agland	Groce ry	Met. pop
AL	0.249	-0.069	-0.213	0.299	-0.376	-0.584	0.073	-0.468	-0.047
AK	0.485	-0.082	-1.064	2.187	-0.062	-0.899	-0.693	2.031	-0.259
AZ	-0.707	-0.439	-0.767	-1.115	-0.062	-0.629	-1.444	-1.389	0.161
AR	0.540	-0.048	0.608	3.045	-0.690	-0.354	-0.044	-0.366	-0.220
CA	1.311	-0.336	-0.301	-0.252	-0.376	0.524	0.012	-0.040	0.520
CO	-0.210	-0.445	-0.405	-0.647	-0.062	0.315	-0.470	-1.141	0.181
CT	0.417	-0.514	-1.024	-1.089	-1.005	0.999	0.396	0.163	0.438
DE	-2.254	-0.660	-0.988	0.101	-1.005	-0.521	-1.075	-0.718	-0.764
FL	0.084	-0.353	-0.760	-1.127	-0.376	-0.057	-0.510	-0.188	0.340
GA	0.092	-0.249	-0.682	0.250	-0.376	-0.435	-0.708	-0.513	-0.015
HI	2.963	-0.527	-0.481	-0.614	0.881	0.024	-0.432	0.087	-0.078
ID	1.366	1.426	1.657	1.152	0.881	-0.472	-0.611	-0.704	-0.320
IL	0.596	0.164	-0.712	-0.070	-0.690	-0.226	1.388	0.091	0.388
IN	-1.019	0.027	-0.377	0.056	-0.376	-0.205	0.939	-0.805	0.354
IA	-0.634	4.621	1.574	2.322	-0.376	-0.111	1.850	-0.143	-0.263
KS	-0.171	1.077	0.924	1.087	-0.062	3.535	2.004	-0.565	0.015
KY	-0.215	-0.305	1.019	0.327	-0.062	-0.104	0.371	-0.049	-0.378
LA	-0.021	-0.123	-0.491	-0.264	-0.376	0.999	-0.437	0.422	0.103
ME	1.051	-0.623	-0.260	-0.044	-0.062	2.973	-1.311	1.013	-0.368
MD	0.953	-0.543	-0.983	-0.868	-0.690	0.212	-0.412	-0.325	0.488
MA	0.000	-0.602	-1.115	-0.959	-0.062	-0.158	-1.195	-0.273	0.753
MI	0.028	-0.439	-0.448	-0.693	-0.376	-0.047	-0.477	0.183	0.248
MN	0.176	0.279	0.111	0.202	-0.690	0.048	0.528	-0.555	-0.616
MS	-0.312	0.252	0.485	0.851	0.253	-0.084	-0.349	-0.341	-0.558
MO	-0.695	-0.226	0.253	0.212	-0.690	-0.614	0.920	-0.571	-0.337
MT	-0.637	-0.276	2.184	-0.514	0.881	-0.947	1.124	0.624	-0.943
NE	0.595	3.162	1.398	2.499	-0.376	0.333	2.179	0.549	-1.440
NV	-1.110	-0.635	-1.126	-1.209	-0.690	0.577	-1.125	-1.194	0.084
NH	-0.516	-0.629	-0.896	-1.145	-0.062	-0.649	-1.311	-0.056	-0.287
NJ	2.323	-0.505	-1.064	-0.645	-1.005	0.664	-0.940	1.407	2.318
NM	-0.156	-0.530	0.302	-0.622	0.881	-0.361	0.700	-1.346	-0.292
NY	0.717	-0.564	-0.988	-0.945	-0.690	-0.452	-0.698	3.526	-0.373
NC	-0.144	0.132	-0.496	-0.035	-0.376	-0.844	-0.622	-0.204	-0.154
ND	0.850	2.793	2.895	1.875	4.025	-0.715	2.077	1.437	-0.871
OH	0.397	-0.310	-0.629	-0.407	-0.376	-0.057	0.471	-0.482	-0.026
OK	-0.523	-0.287	1.152	-0.147	-0.062	-0.671	1.629	-0.491	-0.245
OR	0.869	-0.046	0.833	0.167	0.253	1.875	-0.437	-0.262	-0.403
PA	0.189	-0.366	-0.734	-0.298	-0.690	0.315	-0.594	-0.177	0.451
RI	-0.462	-0.618	-1.124	-0.921	0.567	-0.357	-1.208	-0.347	0.099
SC	-0.419	-0.452	-0.608	-0.495	-0.062	0.102	-0.781	-0.339	0.281
SD	-1.874	0.961	2.547	0.999	1.510	-0.099	2.087	1.124	-0.657
TN	-0.391	-0.393	-0.107	-0.014	0.253	-1.061	-0.005	-0.332	-0.398
TX	-0.250	-0.371	-0.216	-0.416	-0.376	-0.624	1.579	-1.245	0.150
UT	-0.288	-0.606	-0.526	0.326	-1.005	-1.053	-0.632	-1.385	0.085
VT	0.999	-0.348	0.042	-0.183	2.139	-0.263	-0.938	2.852	-0.966
VA	-0.794	-0.464	-0.652	-0.578	-0.376	-0.141	-0.475	-0.183	-0.438
WA	1.775	-0.129	0.958	-0.110	-0.376	0.109	-0.110	0.079	0.557
WV	-1.343	-0.630	0.079	-0.390	0.253	2.686	-0.764	0.277	5.681
WI	-0.206	0.244	0.161	0.847	-0.690	-0.146	0.002	-0.510	-0.921
WY	-1.820	0.181	1.052	-0.368	-0.376	0.176	0.458	-0.192	-1.060

A.2: Employment Structure Group 2002

State	Dist.	Input Supply	Prim. Prod.	Process	Retail	Waste	Agland	Grocery	Met. pop
AL	-0.175	-0.004	-0.252	0.499	-0.180	-0.914	0.068	-0.184	-0.089
AK	-0.589	0.022	-1.021	1.919	0.411	0.134	-0.693	1.650	-0.194
AZ	0.373	-0.508	-0.717	-1.103	-0.771	-0.495	-1.474	-1.560	0.641
AR	-0.463	0.058	0.518	3.075	-0.475	-0.675	-0.047	0.367	-0.209
CA	0.905	-0.395	-0.251	-0.273	-0.475	0.434	0.055	-0.313	0.486
CO	0.228	-0.481	-0.433	-0.533	-0.180	0.163	-0.417	-1.161	0.289
CT	-0.131	-0.550	-1.018	-0.986	-1.066	1.467	0.337	-0.237	0.280
DE	-2.419	-0.744	-0.950	-0.177	-1.066	-0.187	-1.166	-0.998	-0.734
FL	0.185	-0.309	-0.702	-1.009	-0.180	-0.187	-0.403	-0.583	0.530
GA	-0.102	-0.169	-0.621	0.002	-0.475	-0.887	-0.704	-0.450	0.217
HI	2.863	-0.468	-0.479	-0.587	1.002	-0.723	-0.312	0.309	-0.154
ID	1.808	1.799	1.868	1.832	0.707	-0.424	-0.621	-0.306	-0.094
IL	0.150	0.161	-0.714	-0.007	-0.771	0.111	1.406	-0.336	0.276
IN	-0.857	-0.172	-0.347	-0.159	-0.475	-0.399	0.936	-0.945	0.307
IA	-0.573	3.653	1.404	2.168	-0.180	-0.011	1.918	-0.091	-0.318
KS	0.114	1.546	0.697	1.129	-0.180	1.337	2.037	-0.481	-0.038
KY	-0.389	-0.273	1.292	0.201	0.116	0.060	0.316	0.222	-0.410
LA	-0.068	-0.038	-0.549	-0.361	-0.180	0.389	-0.480	0.841	-0.166
ME	1.657	-0.663	-0.326	-0.020	0.116	1.488	-1.343	1.017	-0.467
MD	0.425	-0.588	-0.943	-0.804	-0.180	0.580	-0.381	-0.641	0.440
MA	-0.171	-0.656	-1.089	-0.885	-0.180	-0.111	-1.224	-0.740	0.559
MI	-0.321	-0.522	-0.554	-0.761	-0.475	0.679	-0.501	0.024	0.063
MN	-0.105	0.216	0.143	0.219	-0.475	0.188	0.536	-0.715	-0.727
MS	-0.065	0.397	0.370	1.328	0.116	-0.053	-0.352	0.368	-0.635
MO	-0.709	-0.032	0.245	0.126	-0.475	-0.904	0.957	-0.633	-0.361
MT	-0.013	-0.157	2.399	-0.564	1.002	-0.405	1.011	1.441	-0.975
NE	0.258	3.301	1.416	2.836	-0.475	0.792	2.172	0.716	-1.491
NV	-1.068	-0.712	-1.045	-1.258	-0.475	-0.013	-1.131	-1.213	0.601
NH	-0.813	-0.707	-0.943	-1.161	-0.180	0.123	-1.346	-0.127	-0.348
NJ	1.341	-0.607	-1.045	-0.654	-1.066	0.626	-0.927	0.626	2.118
NM	-0.495	-0.619	0.155	-0.696	0.707	-0.376	0.744	-1.389	-0.222
NY	0.433	-0.615	-0.943	-0.929	-0.771	-0.482	-0.670	2.567	-0.493
NC	-0.232	-0.005	-0.407	0.038	-0.475	-0.702	-0.597	-0.095	-0.027
ND	1.656	3.325	2.844	0.934	2.184	-1.163	2.006	1.307	-0.929
OH	0.097	-0.367	-0.605	-0.394	-0.180	-0.193	0.502	-0.582	-0.176
OK	-0.102	-0.316	0.959	-0.031	-0.180	-1.075	1.471	-0.452	-0.280
OR	0.635	-0.068	0.981	0.022	-0.180	1.550	-0.430	0.044	-0.367
PA	-0.107	-0.456	-0.737	-0.236	-0.475	0.155	-0.608	-0.206	0.291
RI	-0.505	-0.692	-1.114	-1.040	0.411	-0.816	-1.258	-0.732	-0.113
SC	-0.820	-0.473	-0.597	-0.424	-0.180	-0.308	-0.795	-0.164	0.437
SD	-1.079	0.910	2.498	0.871	1.298	-0.545	2.055	1.424	-0.636
TN	-0.239	-0.239	-0.033	0.132	0.116	-0.185	0.036	-0.020	-0.356
TX	-0.376	-0.439	-0.231	-0.373	-0.475	-0.378	1.530	-0.979	0.357
UT	0.051	-0.662	-0.457	0.166	-0.475	-0.922	-0.615	-1.486	0.446
VT	2.577	-0.583	-0.113	-0.112	1.593	2.171	-0.968	3.394	-1.056
VA	-0.862	-0.518	-0.594	-0.422	-0.475	-0.721	-0.460	-0.420	-0.440
WA	1.765	-0.233	1.081	-0.029	-0.180	1.081	-0.108	0.085	0.635
WV	-0.892	-0.708	-0.134	-0.224	-0.180	4.013	-0.821	0.761	5.603
WI	0.639	0.183	0.033	0.905	-0.475	0.255	0.015	-0.549	-0.979
WY	-1.615	0.518	1.061	-0.601	0.116	-0.887	0.684	-0.294	-1.093

Appendix B: Standardized data for Agricultural Production Intensity Group

B.1: Agricultural Production Intensity Group 2007

State	Income	Pay	Chem	Fert	Conserve	Crop sales	VLfarm	Irrigat	Crop land
AL	0.167	0.234	-0.496	-0.449	1.288	-1.507	-0.002	-0.655	-0.187
AK	-0.977	0.273	-1.252	-1.367	-0.256	-0.150	-1.010	-0.815	-1.820
AZ	1.064	-0.468	-0.852	-1.156	.	0.611	-0.870	-0.473	-2.048
AR	0.641	0.724	0.699	0.285	-0.183	-0.381	0.978	3.359	0.540
CA	1.369	-1.123	3.240	1.448	-1.045	1.020	2.681	2.912	-0.686
CO	-0.089	0.073	-1.009	-1.154	1.319	-0.668	-0.385	0.214	-0.732
CT	0.299	-1.098	0.313	0.619	-1.288	1.270	-0.958	-0.477	0.004
DE	0.182	-1.051	2.519	2.484	-0.721	-1.308	-0.918	1.763	1.546
FL	0.348	-1.204	2.137	1.205	-0.463	1.637	-0.129	1.350	-0.770
GA	1.011	0.447	0.790	0.473	0.221	-0.792	0.467	0.848	0.385
HI	0.667	-1.281	0.008	0.003	.	1.799	-0.900	-0.221	-1.556
ID	0.703	-0.456	-0.114	0.325	1.116	-0.273	-0.300	2.555	-0.103
IL	-0.133	0.767	1.010	1.647	-0.220	1.697	2.170	-0.649	1.535
IN	-0.100	0.445	0.840	1.924	-0.502	0.861	0.807	-0.532	1.487
IA	-0.251	0.641	0.405	0.985	0.928	0.201	3.192	-0.792	1.343
KS	-1.562	0.323	-0.632	-0.481	1.031	-0.609	0.765	-0.161	0.269
KY	-0.139	-0.207	-0.675	-0.256	-0.305	-0.841	-0.321	-0.808	0.299
LA	1.232	2.571	0.570	0.096	1.101	0.715	-0.385	0.702	0.444
ME	0.826	-0.662	-0.023	-0.481	0.167	0.310	-0.921	-0.521	0.787
MD	-0.254	-0.411	0.796	1.225	0.007	-0.590	-0.631	-0.338	0.543
MA	-0.144	-0.975	0.740	0.116	-1.331	1.349	-0.917	-0.038	0.097
MI	-0.275	-0.252	0.633	1.071	-0.249	0.550	0.041	-0.212	1.311
MN	-0.140	0.592	0.120	0.607	1.163	0.337	1.829	-0.634	1.259
MS	0.283	1.466	0.328	-0.200	3.071	-0.595	0.147	1.042	0.444
MO	0.343	1.150	-0.555	-0.195	0.773	0.001	0.424	-0.335	0.288
MT	-0.295	3.502	-1.115	-1.309	0.590	-0.051	-0.513	-0.479	-1.015
NE	-0.525	0.025	-0.534	-0.368	-0.347	-0.115	1.616	1.349	-0.304
NV	-0.330	-1.077	-1.111	-1.314	.	-0.182	-0.962	0.517	-1.711
NH	-0.578	-0.781	-0.859	-0.957	-1.289	0.338	-0.978	-0.726	0.297
NJ	1.526	-1.124	1.950	1.908	-1.061	1.924	-0.892	0.954	1.008
NM	2.176	-0.299	-1.221	-1.442	-0.918	-1.019	-0.768	-0.637	-2.012
NY	0.217	-0.668	-0.069	-0.139	-0.714	-0.539	-0.257	-0.729	0.805
NC	0.112	-0.662	0.978	1.025	-0.546	-1.027	0.680	-0.452	0.804
ND	0.126	2.213	-0.376	-0.545	1.550	1.756	0.637	-0.796	0.599
OH	-0.216	0.529	0.249	1.170	-0.368	0.562	0.339	-0.832	1.222
OK	-1.533	0.736	-1.064	-1.051	-0.256	-1.260	-0.215	-0.684	-0.692
OR	-0.421	-0.459	-0.432	-0.662	-0.081	1.032	-0.300	0.499	-0.947
PA	0.464	-0.739	-0.082	-0.012	-0.129	-0.692	-0.044	-0.795	0.943
RI	1.018	-0.855	0.500	1.162	.	1.831	-1.006	0.359	0.183
SC	-1.196	0.256	-0.013	0.161	1.343	-0.608	-0.598	-0.393	0.434
SD	0.673	1.065	-0.785	-0.831	-0.148	0.241	0.254	-0.766	-0.451
TN	-2.327	0.769	-0.313	0.055	-0.381	-0.128	-0.470	-0.768	0.305
TX	-0.067	0.624	-0.963	-1.110	-0.293	-0.737	1.744	-0.410	-1.169
UT	-0.018	-0.547	-1.173	-1.340	-0.750	-0.977	-0.832	0.347	-1.552
VT	1.418	-0.933	-0.896	-0.600	-1.095	-1.536	-0.914	-0.832	0.494
VA	-1.064	-0.365	-0.493	-0.181	-1.013	-0.821	-0.427	-0.713	-0.110
WA	0.600	-0.272	0.498	-0.064	2.352	1.136	0.041	0.537	-0.098
WV	-2.793	-1.260	-1.142	-1.219	-1.332	-1.609	-0.683	-0.857	-0.809
WI	0.688	-0.177	0.172	0.348	0.341	-0.809	0.467	-0.521	0.985
WY	-2.722	-0.018	-1.247	-1.458	-1.075	-1.355	-0.776	-0.263	-1.887

B.2: Agricultural Production Intensity Group 2002

State	Income	Pay	Chem	Fert	Conserve	Crop sales	VLfarm	Irrigat	Crop land
AL	1.780	-0.387	-0.299	-0.303	1.661	-1.334	0.350	-0.661	-0.036
AK	3.012	0.294	-1.276	-1.286	0.088	-0.054	-0.956	-0.830	-1.876
AZ	1.128	-0.891	-1.001	-1.157	-1.283	0.999	-0.772	-0.431	-2.141
AR	0.661	0.766	0.598	-0.025	-0.825	-0.625	1.013	3.115	0.616
CA	0.586	-1.208	3.189	1.444	-0.956	1.393	4.591	3.158	-0.701
CO	-0.403	-0.201	-1.065	-1.094	1.013	-0.908	-0.279	0.185	-0.825
CT	-0.139	-1.148	0.801	2.356	-1.177	1.160	-0.901	-0.363	0.407
DE	-0.640	-0.857	1.855	2.762	-0.724	-1.033	-0.855	1.606	1.403
FL	2.172	-1.353	2.166	1.206	-0.870	1.700	0.284	1.526	-0.759
GA	2.083	-0.375	0.764	0.245	1.255	-0.652	0.681	0.615	0.225
HI	0.742	-1.440	0.093	-0.141	-1.279	1.832	-0.813	-0.162	-1.639
ID	0.818	-0.380	-0.147	0.194	1.364	0.004	-0.114	2.677	-0.204
IL	-0.943	1.029	0.999	1.112	0.152	1.493	1.079	-0.683	1.380
IN	-0.942	0.708	0.627	1.318	-0.446	0.820	0.350	-0.592	1.337
IA	-0.249	0.563	0.344	0.558	0.944	0.185	2.338	-0.812	1.192
KS	-1.510	0.260	-0.824	-0.612	0.943	-0.871	0.350	-0.155	0.198
KY	0.593	-0.071	-0.669	-0.199	0.127	-0.464	-0.313	-0.830	0.571
LA	-0.169	1.708	0.982	0.098	0.282	0.632	-0.246	0.783	0.531
ME	0.399	-0.633	0.126	-0.180	0.229	0.113	-0.861	-0.520	0.723
MD	-0.705	-0.305	0.532	1.090	0.037	-0.524	-0.567	-0.297	1.050
MA	-0.409	-0.993	0.685	0.737	-1.264	1.281	-0.852	0.029	0.194
MI	-0.870	0.300	0.706	0.825	0.044	0.823	0.019	-0.241	1.198
MN	-1.004	0.420	0.130	0.268	1.248	0.366	1.609	-0.652	1.154
MS	0.185	0.695	0.821	-0.353	2.520	-0.616	0.284	0.834	0.374
MO	-0.561	0.997	-0.586	-0.312	0.825	-0.274	0.151	-0.397	0.438
MT	-0.893	3.790	-1.146	-1.264	1.148	-0.323	-0.578	-0.451	-1.087
NE	-1.045	0.179	-0.611	-0.479	-0.248	-0.519	1.013	1.217	-0.355
NV	-0.187	-1.061	-1.175	-1.269	-1.238	-0.501	-0.884	0.614	-1.732
NH	-0.606	-0.268	-0.784	-0.848	-1.051	0.570	-0.926	-0.727	0.263
NJ	0.034	-1.238	1.842	1.899	-1.129	2.035	-0.811	0.938	0.945
NM	0.946	-0.119	-1.250	-1.379	-0.782	-1.078	-0.677	-0.630	-2.096
NY	-0.915	0.157	-0.117	-0.153	0.095	-0.447	-0.180	-0.720	0.780
NC	0.247	-0.854	1.422	1.062	-0.227	-0.812	1.410	-0.400	0.793
ND	-0.378	2.778	-0.560	-0.734	1.878	1.474	-0.180	-0.805	0.385
OH	-0.045	0.676	0.191	0.683	-0.162	0.408	-0.114	-0.831	1.134
OK	1.079	0.076	-1.097	-0.971	0.074	-1.319	-0.180	-0.674	-0.527
OR	-0.450	-0.746	-0.444	-0.621	-0.091	1.116	-0.246	0.582	-1.010
PA	-0.770	-0.564	-0.172	0.229	-0.069	-0.707	-0.246	-0.787	0.888
RI	0.632	-1.068	0.620	1.446	-1.243	1.899	-0.955	0.428	0.136
SC	-0.339	-0.298	0.155	0.257	1.474	-0.281	-0.511	-0.515	0.372
SD	-0.685	1.140	-0.881	-0.926	-0.036	-0.220	-0.114	-0.755	-0.465
TN	-0.577	-0.243	-0.323	-0.043	-0.376	0.151	-0.412	-0.795	0.423
TX	1.947	0.255	-1.001	-1.040	-0.239	-0.931	1.344	-0.375	-1.123
UT	0.082	-0.386	-1.193	-1.255	-0.559	-1.091	-0.743	0.307	-1.616
VT	-0.483	0.924	-0.854	-0.613	-1.215	-1.477	-0.836	-0.832	0.605
VA	-0.062	-0.421	-0.421	-0.174	-0.622	-0.736	-0.246	-0.682	0.222
WA	-0.044	-0.329	0.623	-0.001	2.180	1.044	0.416	0.663	-0.165
WV	-2.284	-1.010	-1.125	-1.103	-1.089	-1.510	-0.662	-0.860	-0.517
WI	-0.016	0.571	0.015	0.131	0.602	-0.754	0.350	-0.505	0.921
WY	-0.800	0.562	-1.265	-1.386	-0.983	-1.437	-0.603	-0.307	-1.981

B.3: Agricultural Production Intensity Group 1997

State	Income	Pay	Chem	Fert	Conserve	Crop sales	VLfarm	Irrigat	Crop land
AL	0.792	-0.426	-0.299	-0.252	1.716	-1.454	0.393	-0.707	0.169
AK	4.413	1.271	-1.326	-1.361	-0.138	0.826	-0.970	-0.816	-1.893
AZ	1.134	-0.507	-1.057	-1.245	.	0.824	-0.760	-0.338	-2.137
AR	0.792	0.173	0.693	0.047	-0.884	-0.462	1.358	2.853	0.676
CA	0.338	-0.874	2.560	1.119	-1.125	1.295	4.529	3.263	-0.800
CO	-0.761	0.171	-1.093	-1.157	0.765	-0.983	-0.228	0.532	-1.005
CT	-0.087	-0.969	0.245	2.030	-0.806	0.735	-0.916	-0.504	0.420
DE	-1.055	-0.893	1.817	1.916	-1.315	-1.327	-0.839	1.004	1.311
FL	1.208	-1.047	2.449	1.741	-0.859	1.562	0.324	1.711	-0.853
GA	1.169	-0.484	0.990	0.694	1.431	-0.528	0.944	0.417	0.413
HI	-0.248	-1.080	-0.020	0.197	.	1.650	-0.838	-0.116	-1.492
ID	-0.357	-0.019	-0.199	0.536	1.400	0.251	-0.090	3.053	-0.210
IL	0.099	1.004	0.939	1.059	-0.224	1.428	1.151	-0.682	1.260
IN	0.107	0.674	0.837	1.411	-0.214	0.680	0.462	-0.625	1.214
IA	0.581	1.160	0.606	0.538	1.223	0.192	2.116	-0.802	1.108
KS	-0.478	0.755	-0.802	-0.709	0.965	-0.657	0.393	-0.093	0.297
KY	1.375	-0.217	-0.601	-0.114	-0.065	0.116	-0.366	-0.790	0.640
LA	0.419	0.537	1.380	0.222	-0.427	0.994	-0.228	0.792	0.523
ME	-0.931	-0.768	0.327	-0.129	0.185	0.001	-0.880	-0.465	0.728
MD	-0.411	-0.617	0.659	1.119	-0.860	-0.796	-0.463	-0.378	1.040
MA	0.412	-1.005	0.553	0.554	-1.163	1.590	-0.856	0.066	0.210
MI	-1.174	0.173	0.819	0.977	0.070	0.716	-0.021	-0.287	1.131
MN	-1.229	0.850	0.449	0.369	1.085	0.132	1.013	-0.655	1.053
MS	0.435	0.334	1.121	-0.318	2.303	-0.523	0.393	0.781	0.456
MO	-0.049	0.783	-0.432	-0.328	1.209	-0.259	0.186	-0.421	0.489
MT	-0.738	3.664	-1.205	-1.316	0.688	0.042	-0.641	-0.383	-1.102
NE	-0.349	0.496	-0.659	-0.578	-0.235	-0.499	0.806	1.172	-0.390
NV	-0.872	-0.956	-1.261	-1.353	.	-0.312	-0.889	0.709	-1.813
NH	-1.143	-0.856	-0.889	-0.821	-0.769	0.071	-0.943	-0.688	0.298
NJ	-0.009	-0.931	1.502	1.921	-1.376	1.864	-0.818	0.828	0.937
NM	0.631	-0.175	-1.307	-1.452	-1.048	-0.975	-0.641	-0.612	-2.134
NY	-1.552	-0.585	0.011	-0.131	-0.828	-0.637	-0.297	-0.705	0.717
NC	1.485	-0.808	1.228	1.321	-0.546	-0.801	1.634	-0.571	0.785
ND	-1.116	3.559	-0.724	-0.762	1.659	1.439	-0.469	-0.796	0.382
OH	1.312	0.393	0.458	0.886	-0.130	0.606	-0.021	-0.823	1.083
OK	-0.431	0.444	-1.117	-1.019	-0.107	-1.359	-0.297	-0.657	-0.515
OR	-0.215	-0.326	-0.514	-0.661	-0.042	1.188	-0.159	0.653	-1.073
PA	-0.765	-0.763	-0.013	0.090	-0.695	-0.749	-0.021	-0.776	0.924
RI	-0.524	-1.023	0.379	1.299	.	1.667	-0.969	0.105	0.233
SC	0.482	-0.225	0.647	0.578	1.713	0.008	-0.504	-0.526	0.568
SD	0.609	1.353	-0.891	-1.112	-0.012	-0.059	-0.228	-0.748	-0.580
TN	-0.016	-0.025	-0.356	-0.061	0.276	0.161	-0.435	-0.796	0.551
TX	0.171	0.373	-1.044	-1.126	-0.237	-0.866	1.634	-0.286	-1.177
UT	0.050	-0.248	-1.263	-1.342	-0.585	-1.031	-0.745	0.488	-1.647
VT	-0.400	-0.833	-0.962	-0.748	-1.112	-1.825	-0.851	-0.814	0.525
VA	-0.230	-0.635	-0.360	-0.120	-0.809	-0.803	-0.228	-0.687	0.330
WA	-0.382	-0.123	0.262	0.018	1.557	1.032	0.324	0.665	-0.221
WV	-1.649	-0.711	-1.185	-1.187	-1.395	-1.738	-0.692	-0.840	-0.253
WI	-0.835	0.149	-0.036	0.218	0.984	-0.951	0.255	-0.515	0.809
WY	-0.008	-0.185	-1.315	-1.458	-1.221	-1.480	-0.612	-0.187	-1.985

Appendix C: Standardized data for Consumption and Health Group 2007

C.1: Consumption and Health Group 2007

State	Obese	Diabetes	SNAP	Grocery	Full Service	Limit Service	Food Service	Convenient
AL	1.618	1.558	0.929	0.029	-0.879	1.319	-0.749	0.850
AK	0.667	-0.624	-0.519	1.654	-1.571	-2.103	3.462	-2.206
AZ	-0.178	0.176	0.074	0.885	-0.095	0.139	0.121	-0.329
AR	1.055	-0.624	1.308	-0.547	-0.622	1.381	-0.759	0.940
CA	-1.059	0.322	-1.054	0.003	0.902	1.159	0.091	-1.462
CO	-2.468	-1.570	-1.301	0.486	0.604	-0.420	-0.415	-1.532
CT	-1.623	-0.915	-0.918	0.548	0.226	-0.927	1.590	-1.593
DE	0.667	0.176	-0.263	-1.799	0.765	-0.969	0.934	0.836
FL	-0.777	0.103	-0.521	0.196	1.634	-0.751	-0.115	-0.427
GA	0.843	1.776	0.368	0.401	-0.117	0.448	0.036	-0.086
HI	-1.623	-0.188	-0.660	-3.071	4.019	1.764	0.701	-0.145
ID	-0.425	-0.406	-0.909	1.182	-0.446	-0.362	-1.045	-0.128
IL	-0.249	0.394	0.251	-0.506	0.265	0.301	1.280	-0.709
IN	0.385	0.467	0.120	-0.428	-0.105	1.118	-0.474	0.110
IA	0.491	-0.988	-0.281	1.401	-1.222	-0.843	-0.919	1.038
KS	0.491	-0.479	-0.810	0.820	-0.903	0.217	-0.712	-0.549
KY	0.843	1.267	1.761	0.823	-1.049	0.568	-0.535	0.022
LA	1.548	1.485	1.876	-0.319	-0.606	0.864	0.528	0.771
ME	-0.390	-0.624	1.171	0.859	-0.117	-1.155	-0.803	1.302
MD	-0.002	0.176	-0.972	-0.452	0.260	0.040	0.704	-0.628
MA	-1.623	-0.915	-0.534	-1.405	1.242	-0.296	1.648	-1.181
MI	0.667	0.685	1.007	-0.367	0.092	-0.315	0.043	0.491
MN	-0.108	-1.570	-1.153	-1.701	0.333	-0.179	0.888	0.884
MS	2.217	2.213	1.881	-0.578	-1.404	1.899	-0.522	1.834
MO	0.667	-0.042	1.771	-0.625	0.172	0.453	0.679	1.108
MT	-1.305	-1.352	-0.271	-0.458	0.512	-1.161	-1.246	0.484
NE	0.068	-0.915	-0.769	1.534	-1.337	-0.210	-0.808	-0.196
NV	-0.601	0.249	-1.226	-0.887	1.468	-0.127	0.139	-0.683
NH	-0.425	-0.624	-1.433	1.384	0.028	-1.278	-0.751	-0.062
NJ	-0.777	0.103	-1.358	0.155	-0.523	-1.379	1.843	-1.124
NM	-0.425	-0.260	0.879	0.502	-0.183	1.359	-0.876	-0.382
NY	-0.284	-0.115	0.166	-0.904	1.818	-0.830	2.287	-1.792
NC	0.843	0.904	0.320	-0.631	0.197	1.239	0.391	0.451
ND	0.244	-0.988	-0.547	-0.846	-0.282	-0.876	-1.596	1.336
OH	0.632	0.467	0.191	0.248	-0.186	0.535	-0.284	0.126
OK	0.879	1.340	0.754	-0.038	-0.802	1.503	-0.868	0.125
OR	-0.002	-0.915	0.962	1.115	0.149	-0.227	-0.867	-1.565
PA	0.526	0.031	0.068	-0.159	-0.226	-1.485	1.099	0.138
RI	-1.623	-0.551	-0.441	-1.422	1.507	0.114	0.618	-1.509
SC	0.949	1.122	1.224	-0.895	0.749	0.864	-0.220	0.863
SD	0.315	-1.206	-0.446	-0.570	-0.511	-0.281	-0.643	1.515
TN	1.548	2.067	1.699	-0.205	0.400	1.131	-0.375	0.010
TX	0.808	1.413	0.345	-0.589	-0.116	1.346	-0.230	0.311
UT	-1.376	-0.697	-1.402	2.154	-1.653	0.201	-0.661	-0.685
VT	-1.552	-1.279	-0.131	0.403	-0.066	-2.480	0.274	1.753
VA	-0.354	-0.115	-0.728	-0.233	0.437	0.480	0.287	0.419
WA	-0.143	-0.624	-0.183	1.494	-0.388	-0.422	-0.346	-1.481
WV	1.407	1.995	1.898	1.190	-1.608	0.005	-0.940	1.286
WI	-0.354	-1.061	-0.621	0.079	-0.407	-0.720	-0.480	0.796
WY	-0.636	-0.842	-1.573	0.085	-0.354	-0.653	-1.407	0.653

C.2: Consumption and Health Group 2002

State	Obese	Diabetes	SNAP	Grocery	Full Service	Limit Service	Food Service	Convenient
AL	1.434	1.954	1.105	1.139	-1.288	0.120	-0.605	0.352
AK	0.589	-1.377	-0.230	1.204	-1.126	-1.241	3.409	-2.004
AZ	-0.808	-0.359	0.259	-5.453	5.172	5.177	2.195	-3.158
AR	0.699	-0.451	1.420	-0.046	-1.021	0.625	-0.208	0.997
CA	-0.955	0.659	-0.977	0.349	0.271	0.531	-0.114	-1.306
CO	-1.947	-1.562	-1.140	0.061	0.653	-0.286	-0.459	-1.136
CT	-1.396	-0.822	-0.864	0.718	-0.062	-0.939	1.184	-1.458
DE	0.221	0.474	-0.718	-0.996	0.645	-0.836	0.636	0.541
FL	-0.881	0.659	-0.464	0.408	0.807	-0.681	-0.249	-0.242
GA	0.625	0.936	0.337	0.635	-0.391	0.100	0.003	-0.105
HI	-1.726	-0.266	0.162	-1.655	2.337	0.724	0.676	-0.133
ID	-0.587	-0.544	-0.601	0.965	-0.442	-0.632	-1.023	-0.123
IL	0.037	0.289	-0.009	-0.369	0.054	0.270	1.188	-0.586
IN	0.846	0.474	-0.029	-0.341	-0.178	0.718	-0.637	0.304
IA	0.405	-0.729	-0.871	1.159	-1.006	-0.508	-0.888	0.709
KS	0.368	-0.637	-0.632	0.509	-0.693	0.272	-1.073	-0.120
KY	0.956	0.566	1.693	-0.167	-0.655	0.903	-0.513	0.774
LA	1.360	1.121	2.537	0.445	-0.709	0.163	0.383	0.247
ME	-0.404	0.104	0.868	0.776	-0.161	-0.994	-1.187	1.165
MD	-0.881	0.381	-1.096	0.087	-0.035	-0.355	0.635	-0.675
MA	-1.285	-0.822	-1.183	-0.811	0.780	-0.339	1.307	-1.094
MI	1.324	0.936	0.233	0.012	-0.173	-0.241	-0.058	0.600
MN	0.221	-1.562	-1.072	-0.635	0.230	-0.397	0.439	0.360
MS	1.838	2.787	1.750	0.204	-1.509	0.863	-0.360	1.621
MO	0.515	0.011	0.919	-0.074	-0.221	0.234	0.772	0.874
MT	-1.138	-1.284	0.045	-0.147	0.056	-0.731	-1.340	0.147
NE	0.515	-0.914	-0.702	0.312	-0.564	0.321	-0.695	0.038
NV	-0.073	-0.451	-0.938	-0.415	0.983	-0.027	0.004	-0.353
NH	-1.432	-0.822	-1.498	1.086	-0.082	-1.081	-0.782	0.190
NJ	-1.028	0.011	-1.340	0.320	-0.431	-1.160	1.679	-0.985
NM	-0.771	-0.544	1.043	-0.798	0.500	1.119	-0.430	0.527
NY	-0.440	0.289	-0.041	-0.321	0.834	-0.741	2.390	-1.586
NC	0.625	0.751	0.027	0.392	-0.367	0.393	-0.282	0.325
ND	0.589	-1.007	-0.489	-0.378	0.132	-0.578	-1.628	0.438
OH	0.442	1.029	-0.085	-0.068	-0.061	0.450	-0.248	0.241
OK	0.405	0.381	1.158	-0.687	-0.338	1.677	-0.716	1.147
OR	-0.550	-0.729	1.350	1.115	-0.031	-0.255	-0.496	-1.447
PA	0.772	0.381	-0.365	-0.251	-0.152	-0.863	1.052	0.007
RI	-1.212	-0.544	-0.224	-1.049	1.189	-0.031	0.581	-1.249
SC	1.471	1.769	1.153	-0.118	0.269	0.013	-0.472	0.929
SD	-0.220	-0.451	-0.320	-0.221	-0.667	-0.534	0.145	1.335
TN	0.993	1.676	1.689	0.133	-0.035	0.569	-0.380	0.171
TX	1.360	1.121	0.242	0.078	-0.365	0.619	-0.413	0.096
UT	-1.579	-1.099	-1.165	0.985	-0.834	0.316	-0.510	-0.088
VT	-1.065	-1.007	-0.295	0.244	-0.075	-1.658	0.148	1.559
VA	0.736	0.011	-0.842	0.372	-0.103	-0.102	0.185	0.632
WA	-0.183	-0.451	-0.383	0.990	-0.308	-0.437	-0.339	-1.056
WV	2.132	1.954	2.297	0.091	-0.891	0.394	-0.781	2.145
WI	-0.073	-1.099	-0.810	-0.108	-0.007	-0.383	-0.570	0.588
WY	-0.844	-1.192	-0.902	0.319	0.071	-0.540	-1.556	-0.156

C.3: Consumption and Health Group 1997

State	Obese	Diabetes	SNAP	Grocery	Full Service	Limit Service	Food Service	Convenient
AL	0.609	2.413	1.130	1.231	-1.610	0.795	-0.936	0.682
AK	1.262	-1.067	-0.596	1.719	-1.453	-2.132	3.151	-2.188
AZ	-1.917	-1.879	-0.300	0.587	0.484	-0.064	-0.318	0.058
AR	0.565	-0.719	1.092	1.471	-1.779	0.297	-1.078	0.476
CA	-0.349	1.021	0.006	0.391	0.862	0.762	0.117	-1.716
CO	-2.179	-0.835	-1.014	-0.350	1.512	-0.511	-0.119	-1.135
CT	-0.916	-0.255	-0.549	0.892	-0.238	-1.256	1.631	-1.675
DE	0.870	0.673	-0.420	-1.997	1.036	-0.500	1.975	-0.547
FL	-0.306	0.325	-0.207	0.003	1.720	-0.683	-0.147	-0.343
GA	-1.046	0.557	0.458	-0.254	0.125	0.739	-0.128	0.343
HI	-1.395	0.325	1.089	-3.110	3.746	1.361	1.218	-0.311
ID	-0.219	-1.183	-0.875	1.468	-0.659	-0.928	-0.897	0.524
IL	0.130	1.717	0.222	-0.969	0.205	0.635	1.316	-0.485
IN	1.916	0.789	-0.793	-0.760	-0.186	1.337	-0.380	0.287
IA	1.132	-0.487	-0.950	0.626	-0.938	-0.095	-0.639	1.248
KS	-0.916	-1.531	-1.089	-0.248	-0.850	1.203	-0.346	0.308
KY	2.177	0.325	1.389	0.169	-1.103	1.027	-0.475	0.693
LA	1.219	1.833	2.052	0.950	-1.029	0.199	0.251	0.304
ME	-0.262	-0.835	0.855	0.991	-0.002	-1.361	-0.758	0.769
MD	0.304	1.021	-0.366	0.346	0.031	-0.393	0.719	-1.501
MA	-0.872	-0.255	-1.048	-1.346	1.184	-0.309	1.686	-1.424
MI	1.088	1.601	0.302	-0.690	0.012	0.041	-0.054	1.000
MN	-0.132	-0.487	-1.063	-1.746	0.577	-0.720	0.249	1.048
MS	2.264	2.065	2.163	0.668	-2.077	1.079	-0.619	1.665
MO	1.001	-0.139	0.154	-0.315	-0.418	0.861	0.400	0.960
MT	-0.959	-1.763	-0.081	0.245	0.537	-0.883	-1.072	-0.530
NE	0.086	-0.139	-0.701	0.495	-0.418	-0.002	-0.499	-0.033
NV	-1.177	-0.719	-1.308	-0.018	0.564	-0.171	-0.078	-0.192
NH	-1.133	-0.951	-1.612	1.438	-0.012	-1.783	-0.510	-0.040
NJ	-0.349	-0.023	-0.814	0.511	-0.558	-1.447	1.833	-1.643
NM	-0.828	0.093	1.248	-1.260	0.728	1.641	-0.442	0.251
NY	-0.349	-0.023	0.735	-0.525	0.940	-0.808	2.362	-1.717
NC	0.653	0.673	-0.090	0.032	-0.138	1.046	-0.333	0.619
ND	0.086	-1.531	-0.858	-1.122	0.082	-0.372	-1.286	0.438
OH	0.391	0.209	-0.213	-0.638	0.051	0.915	-0.145	0.304
OK	-0.741	1.137	0.566	-0.454	-0.904	2.483	-1.662	1.367
OR	1.132	-0.255	0.006	1.227	0.696	-0.265	-0.519	-1.915
PA	0.304	0.325	0.137	0.099	-0.089	-1.288	0.914	-0.539
RI	-1.308	-0.023	0.086	-1.764	1.094	0.085	0.736	-0.902
SC	0.043	0.673	0.613	-0.195	0.342	0.361	-0.482	0.775
SD	0.086	-1.531	-0.508	-1.016	-0.049	-0.216	-0.876	1.070
TN	0.391	0.209	1.126	-0.337	0.138	0.861	-0.508	0.892
TX	0.827	0.673	0.614	0.148	-0.274	0.904	-0.275	-0.096
UT	-0.698	-0.487	-1.226	1.542	-1.410	0.679	-0.464	-0.022
VT	-0.393	-0.487	0.434	0.097	0.485	-2.381	0.180	1.024
VA	-0.175	0.209	-0.524	0.534	-0.111	0.087	0.322	0.371
WA	-0.698	-0.719	-0.244	1.060	0.100	-0.613	-0.150	-0.969
WV	1.654	1.137	3.275	0.360	-1.504	0.821	-0.949	2.257
WI	-0.088	-0.371	-1.467	0.045	-0.037	-0.840	-0.541	0.436
WY	-0.785	-1.299	-0.836	-0.232	0.597	-0.198	-1.374	-0.249

Appendix D: Standardized data for Sample Variable Group 2002

State	Crop	Fruit/Nut	Vegetable	Income	Pay	Top3	Research
AL	-1.140	-0.459	-0.737	1.780	-0.387	1.023	-0.586
AK	0.058	-0.526	0.312	3.012	0.294	-1.684	5.343
AZ	-0.247	-0.017	3.489	1.128	-0.890	-0.070	-0.072
AR	-0.171	-0.505	-0.846	0.661	0.766	-0.233	-0.476
CA	-0.788	3.628	1.703	0.586	-1.208	-1.964	-0.515
CO	-0.910	-0.494	0.013	-0.403	-0.201	0.438	-0.392
CT	1.694	-0.152	-0.341	-0.139	-1.148	0.062	0.381
DE	-1.171	-0.487	0.242	-0.640	-0.857	1.189	-0.537
FL	0.241	2.637	1.370	2.172	-1.353	-1.042	-0.210
GA	-0.793	-0.231	0.186	2.083	-0.375	-0.351	-0.217
HI	0.296	3.599	0.526	0.742	-1.440	-0.144	0.862
ID	-0.538	-0.482	1.798	0.818	-0.380	0.261	-0.539
IL	2.465	-0.507	-0.732	-0.943	1.029	1.442	-0.476
IN	1.585	-0.496	-0.684	-0.942	0.708	0.450	-0.510
IA	0.889	-0.532	-0.890	-0.249	0.563	0.856	-0.569
KS	-0.450	-0.532	-0.889	-1.510	0.260	1.427	-0.665
KY	0.030	-0.513	-0.831	0.593	-0.071	-0.851	-0.565
LA	1.295	-0.466	-0.600	-0.169	1.708	-1.965	0.402
ME	-1.309	0.364	2.912	0.399	-0.633	-0.754	0.053
MD	-0.347	-0.413	-0.254	-0.705	-0.305	0.337	2.014
MA	0.792	1.239	0.489	-0.409	-0.993	0.062	1.186
MI	0.889	0.055	0.290	-0.870	0.300	-1.075	-0.177
MN	0.912	-0.518	-0.434	-1.004	0.420	-0.769	-0.413
MS	-0.222	-0.481	-0.732	0.185	0.695	-0.153	0.139
MO	0.246	-0.480	-0.808	-0.561	0.997	-0.853	-0.099
MT	0.146	-0.499	-0.703	-0.893	3.790	1.246	-0.269
NE	-0.030	-0.535	-0.827	-1.045	0.179	1.659	-0.611
NV	-0.465	-0.524	0.211	-0.187	-1.061	1.257	0.037
NH	0.635	0.254	-0.090	-0.606	-0.268	0.638	0.133
NJ	1.158	0.892	2.238	0.034	-1.238	-0.118	-0.015
NM	-1.263	-0.110	-0.110	0.946	-0.119	1.203	-0.586
NY	-0.890	0.175	0.543	-0.915	0.157	0.172	0.557
NC	-0.603	-0.438	-0.477	0.247	-0.854	-0.457	-0.434
ND	2.280	-0.536	-0.330	-0.378	2.778	-0.540	-0.373
OH	0.936	-0.437	-0.461	-0.045	0.676	-0.943	-0.490
OK	-1.049	-0.504	-0.851	1.079	0.076	0.485	-0.517
OR	0.867	0.532	0.646	-0.450	-0.746	-0.938	-0.054
PA	-0.570	-0.221	-0.496	-0.770	-0.564	-0.460	-0.293
RI	2.201	-0.015	0.487	0.632	-1.068	1.262	0.799
SC	-0.228	-0.207	-0.058	-0.339	-0.298	-0.818	-0.300
SD	0.378	-0.536	-0.893	-0.685	1.140	0.859	-0.716
TN	0.648	-0.501	-0.474	-0.577	-0.243	-1.460	-0.375
TX	-0.694	-0.437	-0.614	1.947	0.255	0.162	-0.487
UT	-0.825	-0.469	-0.740	0.082	-0.386	0.065	-0.193
VT	-1.457	-0.296	-0.611	-0.483	0.924	1.775	-0.366
VA	-0.580	-0.323	-0.439	-0.062	-0.421	-1.412	-0.210
WA	-0.496	2.586	1.213	-0.044	-0.329	-1.531	-0.420
WV	-1.452	-0.245	-0.778	-2.284	-1.010	-0.084	2.657
WI	-0.781	-0.301	-0.058	-0.016	0.571	0.635	-0.362
WY	-1.169	-0.535	-0.883	-0.800	0.562	1.707	-0.476

Appendix E: Map of U.S. Census Regions (EIA 2000)