ANALYSIS OF A FAMILY OF MARKOV MODELS RELATING CONSUMER DEMAND TO TRANSACTIONS DATA

by

Philip E. Paré

Submitted to Brigham Young University in partial fulfillment

of graduation requirements for University Honors

Department of Mathematics

Brigham Young University

June 2012

Advisor: Sean Warnick

Signature: ____________________

Honors Representative: Joseph Parry

Signature: ____________________

ABSTRACT

ANALYSIS OF A FAMILY OF MARKOV MODELS RELATING CONSUMER DEMAND TO TRANSACTIONS DATA

Philip E. Paré

Department of Mathematics

Bachelor of Arts

Mathematical analysis tools are changing how businesses function; improving the use of these tools can result in significant profit gains. There are many challenges when implementing these tools, one of the largest being data issues. For example, real daily sales data is often not well conditioned enough to yield good estimates of the underlying demand function.

Daily sales data, however, is typically aggregated from transactions data that details all the products sold in each individual transaction. This data clearly reveals more information than daily sales data, since daily sales data does not reveal which products are frequently or rarely sold together. If the relationship between consumer demand and transactions data was understood, this data may provide a more reliable source to estimate underlying consumer demand than daily sales data.

This thesis hypothesizes potential models relating transactions and daily

sales data. The problem considered is given a linear demand function, construct a model that generates transactions data consistent with this demand function. A family of Markov models is developed that attempt to address this problem. The basic idea driving these models is that substitutes should rarely appear in the same transaction while complements should frequently appear together. Simulation experiments indicate, however, that the particular Markov models developed here are insufficient and more work needs to be done to construct a model with the desired consistency properties.

ACKNOWLEDGMENTS

First, I want to thank my wife, Annette, for all her support with editing and in general. I am grateful to all my professors that have helped me get to where I am today. I especially want to acknowledge the outstanding guidance of my advisor, Dr. Sean Warnick, for all the support and guidance he has given me over the years. I am also grateful for the help of my referee, Dr. Joe Price. I am grateful for my parents and siblings and the support they have always been as well as my friends in IDeA Labs, for their insights and friendship, especially the help of Brigham Wilson and Anurag Rai with this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Retail and Analysis Tools

In the retail world, a long tradition of relying on manager's intuition and experience for pricing and inventory decisions is being replaced with mathematical analysis. Mathematical analysis can be very informative and useful in making pricing decisions, inventory decisions, and other managerial decisions. However, even though retailers have data readily available to them, without the right models and tools the data can be useless. Moreover, if they have the right models and tools but the data does not have enough information in it, the results can be useless.

There are many tools of mathematical analysis that can be used in a retail setting. This thesis analyzes some of these tools and how they can be useful with various types of data. The most basic tool in analyzing retail data is a demand model. A demand model is a mathematical representation of how the sales of a product are dependent on its price and other explanatory variables. Estimation of the demand model is important because it is useful in many ways:

1. Demand models drive forecasting, predicting future sales, which is essential for

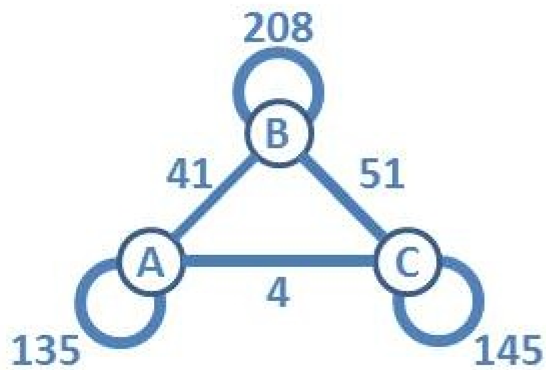

Figure 1.1 Transactions Graph: Products A and B were sold together 41 times, A and C were sold together 4 times, and B and C are were sold together 51 times. Products A, B, and C were purchased multiple times in a single transaction 135, 208, and 145 times respectively.

inventory management.

2. They are essential for optimal pricing.

3. A demand model for a group of products is a significant part of offer design.

4. Analysis of the demand model can indicate whether or not a retailer has market power. A company that has market power has monopolistic characteristics, meaning by changing its own prices it can alter the prices of the whole market, making the market unfair.

Another useful tool of mathematical analysis is a transactions graph, where the nodes are the products and the edges are weighted by the number of times the two products are sold together. This helps to identify the relationships between the different products sold. Figure 1.1 gives a simple example of a transactions graph for products A, B, and C. This graphical representation can give insights that a demand model may not reveal, like existence or lack of relationships between products, and is often easier to generate. These models are explained in further detail in Chapter 2.

1.2 Main Modeling Challenges

When implementing tools of mathematical analysis there are many challenges that can arise. Data preparation and conditioning, or richness, of the data are among these challenges.

1.2.1 Data Preparation

One of the biggest challenges associated with implementing any tool of mathematical analysis using real data is preparing the data for use. Each mechanism that records data can do it in a different way. Sometimes different machines store data to the same place in different ways, resulting in inconsistent representations of the data in the same data set. Even if all the data is uniform, the format required for input to the analysis tools is usually different than the storage format, unless the tools were programmed specifically for that data set.

In retail, there are often several cash registers in a store. Most registers record every transaction that occurs into a central database. This *transaction data* can be used in different ways but must be manipulated in order to be used by most analytical tools. Total sales of each product for each day can be added up to acquire *daily sales data*, which is necessary for performing linear regression, explained in Section 2.2. Transactions data can also be used to produce transactions graphs, by adding up the number of times two products were sold together and making that number the weighted edges between the products.

Depending on the data set, there can be a range of issues that must be addressed before it can be used. The following are some of these problems:

- Abnormal numbers can be used to indicate something special, such as an extremely large price indicating the use of a special coupon.

- Data recording discrepancies can arise. For example, some registers record the price of a product being a multiple of the actual price when multiple products are sold together and others record the individual product price all the time.

- Data recorders will only record an event when it occurs and will not record anything if it does not occur on a given day. This can cause a problem with performing a regression on the data. For example, if the sale of a certain product does not occur on a given day the registers usually will not record a zero nor the product's price for that day. This causes a vital part of the data to be lost, which will drastically affect the regression results. Therefore, in order to perform a dependable regression on the data set, zeros must be added to the data when no products are sold and their prices also must be added appropriately.

- The lack of data could also be lack of inventory or discontinuance of a product. This situation must be identified and a decision must be made whether or not to use the data.

All of the data issues must be dealt with before implementing the analysis tools in order for the results to be dependable. If these issues are overlooked or ignored, the results will be unreliable.

1.2.2 Conditioning Problem

After the data is prepared, solving systems of linear equations is required for estimation of some models. Systems of linear equations can be solved using matrix analysis. Many of these solution techniques require computing the inverse of a matrix. Depending on the structure of a matrix, its inverse can be very sensitive to perturbation. To illustrate this point a simple 2x2 matrix will be used [2]. Consider,

$$A = \begin{bmatrix} 50 & 50 \\ 50.2 & 50 \end{bmatrix}. \tag{1.1}$$

The inverse of A is given by,

$$A^{-1} = \begin{bmatrix} -5 & 5 \\ 5.02 & -5 \end{bmatrix}. \tag{1.2}$$

Now if A is perturbed a small amount,

$$A + \Delta = \begin{bmatrix} 50 & 50 \\ 50.1 & 50 \end{bmatrix}, \tag{1.3}$$

the inverse becomes,

$$(A + \Delta)^{-1} = \begin{bmatrix} -10 & 10 \\ 10.04 & -10 \end{bmatrix}. \tag{1.4}$$

A small perturbation on one entry of the matrix drastically changed the inverse. The reason this occurred is because one column is almost a multiple of another column, i.e., the matrix is almost not invertible. This is known as ill-conditioning [5].

Products in many retail situations rarely change price and when changes do occur they usually occur at the same time for several products. The matrix of prices, where the columns are different products' prices, will be ill-conditioned because their prices are similar with simultaneous price changes. This can cause serious problems when trying to perform a regression on this data. In the field of Econometrics, this conditioning problem is called multicollinearity. Models derived from systems with this property are unstable with large confidence intervals, increased probability of incorrect sign, and reduced precision [4].

There is no perfect solution for this problem but there are several solutions of varying cost and outcomes:

- The best way to fix this problem in a retail setting would be to start changing prices on different days for different products. This solution enriches the data, but it is a high cost solution because changing prices is not a facile task.
- Another lower cost solution is dropping one of the variables that is collinear.
- Also, imposing constraints on the regression could fix this problem. For example, in the retail setting, there could be constraints imposed that force the elasticities of demand for the own products' prices to be negative.

1.3 Problem Formulation

These challenges can make it difficult to estimate a demand model from sales data. Avoiding these problems is quite simple for companies like Walmart with their "rollback polka-dot man" because he keeps the price data rich by lowering and raising prices. Nonetheless, most retailers, unlike Walmart, do not have the resources to change prices sufficiently in order to avoid the conditioning problem. This leaves these companies at a disadvantage for estimating market demand from data.

Nevertheless, even though the underlying demand model may be difficult to estimate from sales data because of conditioning issues, there might be another way to compute a parsimonious estimate. The work in this thesis explores the idea of using transactions data, which typically get aggregated into daily sales data, as a mechanism for understanding the underlying demand function.

The idea is to hypothesize a transactions model that is consistent with a given demand function. Understanding how the demand function then contributes to the generation of transactions data can help generate algorithms that unravel the demand function information from transactions data instead of daily sales data. The fundamental problem addressed in this thesis then becomes:

Given a linear demand function generating daily sales data, develop a model that generates transactions data, which can be used to obtain a transactions graph. This generated transactions data should be consistent with the original demand function in that aggregating the generated transactions data into daily sales data and performing a subsequent regression to estimate an underlying demand function should yield one that is close to the original. A few key metrics are used to determine the closeness of the resulting estimate of the demand function to the original, and these are detailed in Section 3.4.

1.4 Contributions of this Work

The main contribution of this work is the construction of a Markov chain model that is derived from the given demand function and generates transactions data. Although the resulting data is not completely consistent with the original demand function, some aspects of consistency, such as sign symmetry, are generally observed.

1.5 Literature Review

Similar research has been done in the Marketing Science, Data Mining, and Economics literature. Here is a review of the literature divided by topic.

I. Pricing Decisions

In [9], Oxenfeldt discusses the different reasons for making pricing decisions and how executives can do it. He explains that these decisions are affected by the organization's policies and also by prior decisions made by others.

Lazear develops a mathematical model in [6] for maximizing expected profits by adjusting price appropriately over time. He explores products that are being introduced to the market and how to direct their price changes over time. He uses probability that the product will be bought given a particular price and sets up the problem as a dynamic programming problem.

II. Demand Forecasting

This work was inspired by the interaction between IDeA Labs and the BYU Bookstore. In [12], the author explores different ways to perform data denoising in order to forecast demand in the retail setting inspired by and illustrating his point with the BYU Bookstore data. In [4] and [7] the authors explain the econometric skills of regression and statistical analysis. For this thesis a linear demand model is used where the dependent variable is the sales of a products and the independent variables are the price of the product and the prices of the other products related to that product.

III. Markov Transactions Models

Markov models are used in many different disciplines. Even though the area of research explored in this work is new, there are many people who have used Markov models to simulate systems that involve transactions.

In [10], the author uses a collections Markov processes to model a brokered foreign exchange auction. The processes adapt depending on the arrival of buyers and sellers and the bidding transactions that occur. In [11], the author models online retail per transaction using a hidden Markov model.

1.6 Thesis Organization

The rest of the thesis is organized as follows: In Chapter 2, linear demand models and transactions graph models are explored in detail. In Chapter 3, an explanation is given of the Markov chain model developed to connect these two models with a three product example. In Chapter 4, results from simulations with 3, 5, and 10 products cases are presented. Conclusions are drawn and future research ideas are introduced in Chapter 5.

Chapter 2

Linear Models and Transactions Graphs

2.1 Overview

Linear regression models have been researched extensively in many disciplines. In Section 2.2, this well known field will be described briefly, outlining the required underlining assumptions of the linear regression solution. Also, in Section 2.2.3, the form of the linear demand model used to derive the Markov model in Chapter 3 will be explained.

Transactions graphs, unlike linear regression models, are unique. The idea of representing systems as graphs with nodes and edges is not new to any field but this particular representation of products is not common among retailers or retail researchers. Nonetheless, they can be quite useful. The derivation and importance of transactions graphs will be expounded upon in Section 2.3.

2.2 Linear Regression Model

The general form of a linear regression model is,

$$y = Ax + \epsilon, \tag{2.1}$$

where y is a dependent variable, A is a matrix consisting of columns of independent explanatory variables, x is a column of parameters, and ϵ is a vector of errors. The parameters in x are assumed to be some unknown constants that describe the system. There are several ways to approximate these parameters.

2.2.1 Ordinary Least Squares

Ordinary least squares (OLS) is a straightforward way to find estimations of the x parameters in the linear regression model. The derivation of OLS is found by minimizing the sum of the squared errors, SSE. Let $\hat{x}$ be the vector of estimated parameters and the error be defined as $e = y - A\hat{x}$. Then,

$$SSE = (y - A\hat{x})^T(y - A\hat{x}) = y^T y - 2\hat{x}^T A^T y - \hat{x}^T A^T A\hat{x}. \tag{2.2}$$

In order to minimize SSE, the derivative is taken with respect to $\hat{x}$ and set equal to zero,

$$\frac{dSSE}{d\hat{x}} = 2A^T y - 2A^T A\hat{x} = 0. \tag{2.3}$$

Therefore,

$$A^T A\hat{x} = A^T y. \tag{2.4}$$

Solving for $\hat{x}$ gives,

$$\hat{x}^* = (A^T A)^{-1} A^T y. \tag{2.5}$$

This solution is the set of best linear unbiased estimators (BLUE) if some basic required assumptions hold. An unbiased estimator means that the expected value of the estimator is equal to the actual parameter, $E(\hat{x}) = x$.

2.2.2 Basic Required Assumptions

There are five basic assumptions required for the OLS solution to the linear regression model in Equation 2.5 to be the optimal solution.

1. *The errors are normally distributed.* The columns of A do not explain every factor that affects y therefore the other factors that are not included are assumed to be in the error term. By the central limit theorem of statistics, the mean of a sufficiently large number of independent random variables with finite mean and variance is approximately normal. If this assumption does not hold the estimators may not be minimum variance.

2. *The expected value of the error is zero*, that is, $E(\epsilon_t) = 0$ for all t. If this assumption does not hold the intercept terms can be biased.

3. *The errors are homoskedastic*, that is, the variance of the error is constant over time, $var(\epsilon_t) = \sigma^2$ for all t.

4. *There is no autocorrelation between the errors*, that is, the covariance between errors over time are zero, $cov(\epsilon_t, \epsilon_s) = 0$ for all $s \neq t$. If assumptions 3 and 4 do not hold, the estimators are not minimum variance and the standard test statistics are invalid.

5. *The columns of A are non-stochastic.* This means the columns of A are not correlated with the errors and that there is no feedback between the y and the columns of A, that is, no endogenous regressors. If assumption 5 does not hold then the estimators are biased, $E(\hat{x}) \neq x$, and inconsistent, $variance(\hat{x}) \nrightarrow 0$ as $n \to \infty$.

2.2.3 Linear Demand Model

The linear demand model used for the rest of this thesis can be written as,

$$q_{ij} = d_i^T p_j + b_i, \tag{2.6}$$

where q_{ij} is the daily sales of product i on day j, d_i^T is a row of elasticities for product i, p_j is a vector of all products' prices for day j, and b_i is the intercept or the amount of product i sold if all prices were zero.

There is a quantity, q_i, for each product and they can be stacked to be written as,

$$q = Dp_j + b, \tag{2.7}$$

where q is a vector whose ith entry is daily sales of the ith product, D is the matrix of the elasticities whose rows are the d_i^T's from Equation 2.6, that is, d_{ii} is the elasticity of demand for the ith product and the off diagonals are the cross elasticities, p_j is the same as in Equation 2.6, and b is a vector whose ith entry is the intercept of the ith product. Since the cross elasticity should have the same sign in both directions, the D matrix should always be sign symmetric, that is $sign(d_{ij}) = sign(d_{ji})$. Equation 2.7 is a vital part of the Markov model presented in Chapter 3 and sign symmetry is one of the validation mechanisms used to decide if an estimated demand function is sensible.

If the parameters of the model in Equation 2.7 are known, it can be used to calculate the optimal price for maximizing revenue. This is accomplished as follows:

$$\max_p q^T p = \max_p (Dp + b)^T p. \tag{2.8}$$

Taking the derivative of the revenue function with respect to the prices and setting it equal to zero gives,

$$\frac{dq^T p}{dp} = 2Dp + b = 0. \tag{2.9}$$

Solving for the price vector then yields,

$$p^* = -.5D^{-1}b. \tag{2.10}$$

This optimal price is another important part of the model introduced in Chapter 3.

A Simple Example

The demand model is the summation of the individual consumer demand curves [8], meaning the demand model represents a scaled version of the behavior of a single average consumer. Therefore, understanding how a single average consumer buys products can give insight into the overall demand curve. Consider a simple demand

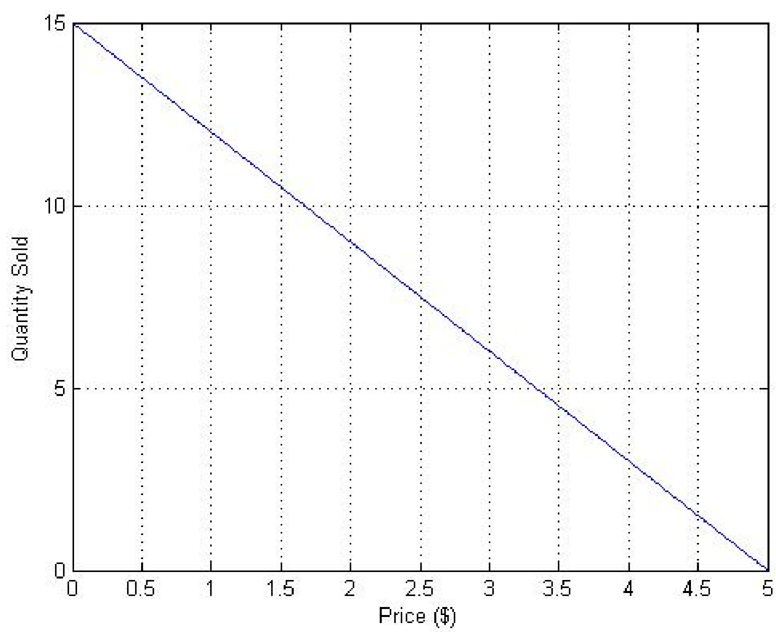

Figure 2.1 Demand Curve for Individual Consumer: If the price is zero this consumer will take 15 products. If the price $\geq$ \$5 the consumer will not buy any.

curve of the form, $q = -3p + 15$, depicted in Figure 2.1. It is clear the consumer will buy 9 products if the price is \$2. If the price is around \$4.50 the consumer will likely only buy a single product. So the demand curve determines the quantity the consumers will buy, on average, given the price. The demand curve gives insight into buying patterns, perhaps suggesting whether or not a product will be bought more than once.

Substitutes and Complements

Consider the demand equation of the form,

$$q_1 = -d_{11}p_1 + d_{21}p_2 - d_{31}p_3 + b_1, \tag{2.11}$$

where q_1 is the sales of product 1, p_i is the prices of product i, d_{11} is the elasticity of demand, d_{21} and d_{31} are the cross elasticities, and b_1 is the intercept. Since the sign of d_{21} is positive, product 2 is a substitute for product 1. If the price of product 2 increases, its sales will decrease since d_{22} will always be negative, yet the sales of product 1 will increase since d_{21} is positive. Being substitutes, there is little chance they will sell together in the same transaction. On the other hand, the sign of d_{31} implies that products 1 and 3 are complements, so there is a large chance that they will sell together; the larger the coefficient, the larger the chance will be that the products will be bought together.

2.3 Transactions Graphs

As explained in Section 1.1, a transactions graph is where the nodes are products and the edges are the number of times the products are sold together. The transactions graph can be normalized, dividing each edge by the total number of sales of each product resulting in a probability transactions graph. This produces directed edges because the total number of each product sold is different. Normalizing the exam-

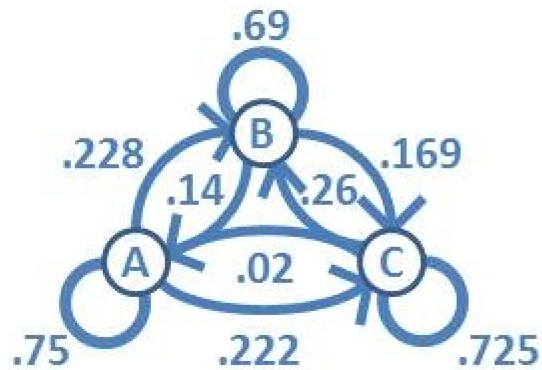

Figure 2.2 Probability Transactions Graph: The top left arrow indicates the probability of product B being in a transaction given that product A is in the transaction, i.e. $P(B|A) = .228$, and was derived by dividing that edge, weighted 41, by the total number of sales of product A, 180.

ple depicted in Figure 1.1 gives an example of a probability transactions graph for products A, B, and C shown if Figure 2.2.

Both forms of the transactions graph are important. The transactions graph, like Figure 1.1, is very helpful in visualizing the relationships between the products and the structure of the retail system. It is also important because the probability transactions graph can be computed from it, which also shows the relationship of the products. However, the probability transactions graph in addition to that, is important for the Markov model presented in Chapter 3.

2.4 Relationship between Linear Demand Models and Transactions Graphs

Transactions data as discussed in Section 1.2.1 is used to estimate both the linear demand model and the transactions graph. However both these tools require the transactions data to be manipulated very differently.

Transaction ID	Date	Products Sold	Products Prices ($)
1	Monday	A, A, A, B, B	$p_A = .50$, $p_B = 1.45$
2	Monday	B, B, C, C, C	$p_B = 1.45$, $p_C = .45$
3	Monday	A, A	$p_A = .50$
4	Tuesday	C, C, C, C	$p_C = .40$
5	Tuesday	A, A, A, A, B, B	$p_A = .50$, $p_B = 1.45$
6	Tuesday	A, B, C	$p_A = .50$, $p_B = 1.45$, $p_C = .40$
7	Tuesday	B, B, B, C, C, C	$p_B = 1.45$, $p_C = .40$
8	Wednesday	B, B, C, C	$p_B = 1.48$, $p_C = .40$
9	Wednesday	A, A	$p_A = .50$

Table 2.1 Transactions Data from Three Days: p_i is the price of product i.

Consider the transactions data in Table 2.1. This data can be aggregated into daily sales data in order to perform a regression to estimate a linear demand model. Given this data, the equation for product A would be,

$$\begin{bmatrix} 5 \\ 5 \\ 2 \\ \vdots \end{bmatrix} = \begin{bmatrix} .50 & 1.45 & .45 \\ .50 & 1.45 & .40 \\ .50 & 1.48 & .40 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} d_{AA} \\ d_{AB} \\ d_{AC} \end{bmatrix} + \begin{bmatrix} b_A \\ b_A \\ b_A \\ \vdots \end{bmatrix}. \tag{2.12}$$

When the data is aggregated into daily sales data the information about which products were sold together is lost. The transaction graph derived from the transactions

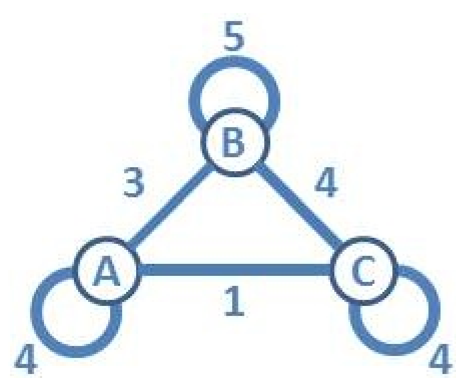

Figure 2.3 Transactions Graph for the Three Days of Data in Table 2.1.

data is depicted in Figure 2.3. This model, opposed to the demand model, maintains the relationship of the products by summing up all the times they are sold together but completely discards the price information. These two tools model the same system and are derived from the same data: therefore no matter how different they are, they are related. A model developed to show the relationship between the two is presented in Chapter 3.

Chapter 3

A Markov Chain Model for Generating Transactions Data from the Linear Demand Model

3.1 Background

Markov chains are dynamic systems that describe the evolution of a probability distribution. In this work only discrete time stationary Markov chains with a finite number of states are considered. The stationarity of the Markov chain means no matter when an initial condition is given to the system it will always produce the same output. A Markov model is described in terms of its transition probabilities, p_{ij}, which can be represented in a transition probability matrix,

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}. \tag{3.1}$$

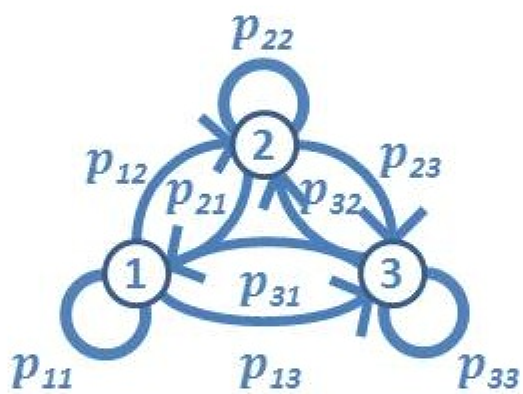

Figure 3.1 Graphical Representation of the Transition Probability Matrix for Equation 3.2: The nodes represents the state and the edges represent the transition probabilities.

The columns of P are stochastic, meaning the entries are non-negative and sum to one. At each time step, k, the state of the chain, x_k, is determined by the previous state and the transition probabilities associated with that state. The evolution of the system is determined by multiplying the transition matrix by the previous state vector, which is a stochastic vector representing the probabilities of the system being in any one of the given states. For example, in a three state system, x_k is determined by the previous state:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_k = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{k-1} . \tag{3.2}$$

Since x_{k-1} is stochastic and P is column stochastic, x_k is stochastic. The graphical representation for the transition probability for Equation 3.2 is found in Figure 3.1.

Depending on the transition probabilities, the state distribution characterized in a Markov chain continues to evolve until a steady state distribution is reached. Equation 3.2 describes the evolution of the probability distribution that characterizes the system. To understand what state the system is in at a given time, the probability distribution must be sampled.

Markov models are used in many disciplines for many different applications, from thermodynamic modeling in physics to the population modeling in biology. They can be used to model almost "any dynamical system whose evolution over time involves uncertainty" [1]. Whether or not it is the right model for a particular system is debatable. Due to the uncertainty and randomness of transactions it seemed appropriate to use a Markov model to produce transactions data from a demand model.

3.2 Markov Chain Model

The goal is to use the linear demand model to develop an associated Markov chain model that generates transactions data. A Markov chain looks promising because it offers a stochastic process that could describe which products appear in each transaction. In order to implement this, methodologies for developing transition probabilities, for choosing a starting state, and a criterion to stop generating sales transactions for each day are required. Methods of verification are also necessary. The next section describes the relationship between consumer demand and transactions, while the following sections are about how to build the Markov chain model.

3.2.1 Relating Consumer Demand to Transactions

The linear demand model, explained extensively in Section 2.2.3, can be related to transactions by the following reasoning. The demand curve can be thought of as representing the average daily sales that will occur given particular prices. A probability distribution can be constructed by positioning a distribution over the demand curve, where the demand curve is the mean, in order to give probability to the daily sales. Consider the example shown in Figure 2.1. By positioning a distribution over the curve we get something like Figure 3.2. Although the distribution in Figure 3.2

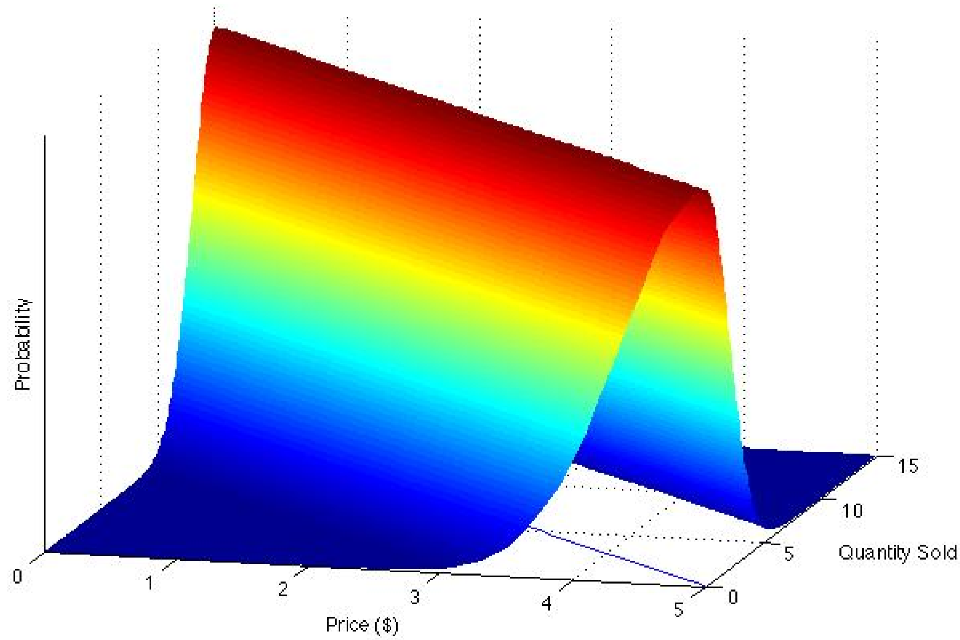

Figure 3.2 Demand Curve for Individual Consumer with Normal Distributions.

suggests the probability of daily sales for the average consumer, one might imagine it as a probability of purchasing more than one of the product in a single transaction. As explained in Section 2.2.3, if the price is around \$4.50 the probability of buying another of that product is low. Extending this idea to multi-dimensions for the multi-product case complicates the visualization but it is the same idea. The demand model produces some multi-dimensional hyperplane and a distribution can be associated with the model in a similar fashion.

Following this reasoning, in the Markov model, the elasticity of the product, d_{ii}, is used to calculate the probability that more than one of the ith product will be purchased in a single transaction. It is also used to determine the probability that a transaction will end. The cross elasticities, d_{ij}, which determine to what degree products are complements or substitutes, are used to calculate the probability that the average consumer will include these products in a single transaction. This leads to the construction of the Markov chain model.

3.2.2 Markov Model Formation

The transactions model developed here uses the linear demand model and its associated set of optimal prices to produce transactions data using a Markov model. The idea is that the Markov chain begins with a random product, which represents the first product in the average consumer's shopping cart. Transition probabilities then determine which other products are added to the shopping cart. The transaction ends when the Markov chain reaches a special stopping state. As the process continues this stopping state is augmented to the other states associated with the products over time resulting in an eventual end to the transaction. In this way the Markov chain, with the transition probabilities derived from the demand model, can generate a single transaction. This section describes how the transition probabilities and stop probabilities are constructed from a linear demand model.

Recall the demand model given in Equation 2.7 in Section 2.2.3. By the explanation of this equation it is clear that the D matrix is of the form,

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix}, \tag{3.3}$$

where d_{ii} is the elasticity of demand of the ith product and d_{ij} for all $i \neq j$ is the cross elasticity between the ith and jth product. From Section 2.2.3 it is clear that if the off diagonal entries are positive then the ith and jth products are substitutes and should have a low probability of being purchased together. On the other hand, if they are negative the ith and jth products are complements and should have a high probability of being purchased together.

In order to produce a Markov model that generates transactions data, a transition probability matrix must be constructed. Negative probabilities do not exist, therefore there can be no negative entries in the matrix. In order to ensure this, a linear transformation is used. Let $\bar{D}$ be defined as,

$$\bar{D} = -D + (-\max_{ij} d_{ij}). \tag{3.4}$$

Since $-\max_{ij} d_{ij}$ is the smallest entry in $-D$, $\bar{D}$ has all non-negative entries by the linear transformation. In this way substitutes will have a very low probability of being purchased together while complements will have a very high probability of being purchased together.

The reason that D was used is because it means the probability that product i will be added to the transaction, if product j is in already in the transaction, depends on the cross elasticity d_{ij}. This cross elasticity shows how the price of product j affects the sales of product i, that is, whether or not product i is sold. So it seems natural to have it in this form. The following two forms of $\bar{D}$ will be used to validate the construction presented in Equation 3.4.

$\bar{D}$ Equation with D^T

Similar to Equation 3.4,

$$\bar{D} = -D^T + (-\max_{ij} d_{ij}). \tag{3.5}$$

The reasoning to use D^T seems less natural than using D because using D^T says if product i is included in a transaction, the probability that product j will be bought is dependent on the how the price of product j affects the sales of product i. This reasoning seems backwards.

$\bar{D}$ Equation with $\frac{D+D^T}{2}$

In order to reconcile the Equations 3.4 and 3.5 an average was used:

$$\bar{D} = -\frac{D + D^T}{2} + (-\max_{ij} \frac{d_{ij} + d_{ji}}{2}). \tag{3.6}$$

These three forms of $\bar{D}$ were used for the simulations presented in Chapter 4.

Transaction End

Another issue to be addressed is that there must be a way to end a transaction: a stopping state. Just as the elasticity of demand, d_{ii}, is correlated with the probability that more than one of a certain product will be purchased in a single transaction it will also be used to determine the stop probability for the transaction. The stop probability is also a function of the price. If the price is too high, it is less likely any other products will be purchased and therefore the stop probability increases. So given a starting state of the ith product, the stop function is defined as,

$$\bar{d}_{ii} \frac{p_i}{p_i^*}, \tag{3.7}$$

where p_i is the price of the product that day and p_i^* is the optimal price from Equation 2.10. Augmenting this state to $\bar{D}$ gives

$$\bar{P} = \begin{bmatrix} 1 & \bar{d}_{11}\frac{p_1}{p_1^*} & \cdots & \bar{d}_{nn}\frac{p_n}{p_n^*} \\ 0 & \bar{d}_{11} & \cdots & \bar{d}_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \bar{d}_{n1} & \cdots & \bar{d}_{nn} \end{bmatrix}, \tag{3.8}$$

where the terms in the first row are from Equation 3.7 and $\bar{d}_{ij}$ is the i,jth entry of the matrix $\bar{D}$ from Equations 3.4, 3.5, or 3.6 depending on the experiment. The transition probability matrix, P, equals $\bar{P}$ with normalized columns, that is, dividing each entry by the column sum. Using the transition probability matrix, P, and setting

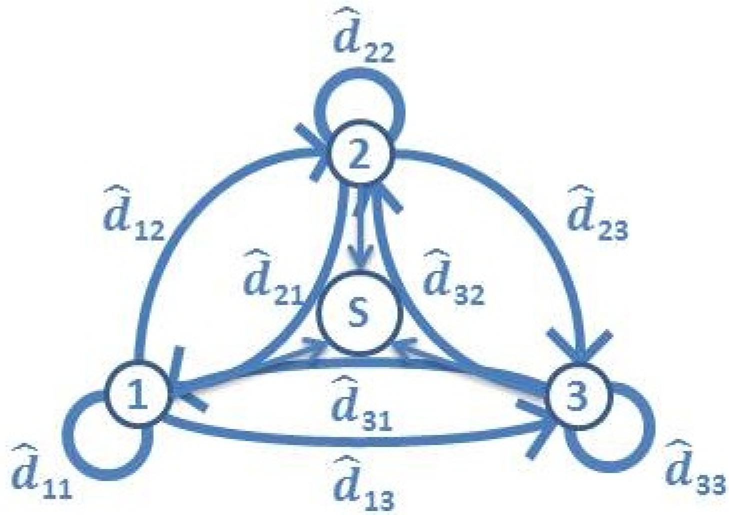

Figure 3.3 Graphical Representation of the Transition Probability Matrix: The "S" node represents the stop state and the $\hat{d}_{ij}$ is the $i+1, j+1$th entry of the transition probability matrix, P.

$n = 3$ gives a transition probability matrix that is represented in a graphical form in Figure 3.3, where $\hat{d}_{ij}$ is the $i+1, j+1$th entry of P. The weights on the edges going into the stop state are the normalized form of Equation 3.7. Note that there are no edges coming out of the stop state; it is a sink, meaning that the model, over time, converges to the stop state with 100% probability.

3.3 Implementing the Markov Model to Generate Transactions Data

The transition probability matrix requires an initial state to simulate a transaction, which is chosen from a probability distribution determined by the vector of projected sales, $\hat{q}$, for that day divided by the total projected sales of the day, $\sum_i \hat{q}_i$. Once an initial state is chosen, it is multiplied by the transition probability matrix and a random number generator is used to sample the resulting state, and this process is continued until the stop state is sampled.

Another important part of simulating transactions data is determining when the day ends, that is, how many transactions occur on a given day. A projected average daily sales amount, DS, for the system is fairly simple to calculate:

$$DS = \frac{1}{n}\sum_{i=1}^{n} \hat{q}_i. \tag{3.9}$$

Adding noise to this average gives a number, DSn, that will make the simulation more realistic. Transactions for that day continue to be generated until the average sales for the day is greater than or equal to DSn.

3.4 A Simple Example

Now a simple three product example will be presented. The example assumes the demand model depicted in the D matrix (Equation 3.3):

$$D_3 = \begin{bmatrix} -174.28 & -23.88 & 42.22 \\ -74.61 & -103.30 & 82.12 \\ 38.38 & 30.36 & -198.85 \end{bmatrix}, \tag{3.10}$$

and,

$$b_3 = \begin{bmatrix} 269.6564 & 228.6452 & 245.0445 \end{bmatrix}^T. \tag{3.11}$$

The demand function, A, is defined by

$$A = \begin{bmatrix} D_3 \\ b_3^T \end{bmatrix} \tag{3.12}$$

$$= \begin{bmatrix} -174.28 & -23.88 & 42.22 \\ -74.61 & -103.30 & 82.12 \\ 38.38 & 30.36 & -198.85 \\ 269.6564 & 228.6452 & 245.0445 \end{bmatrix} \tag{3.13}$$

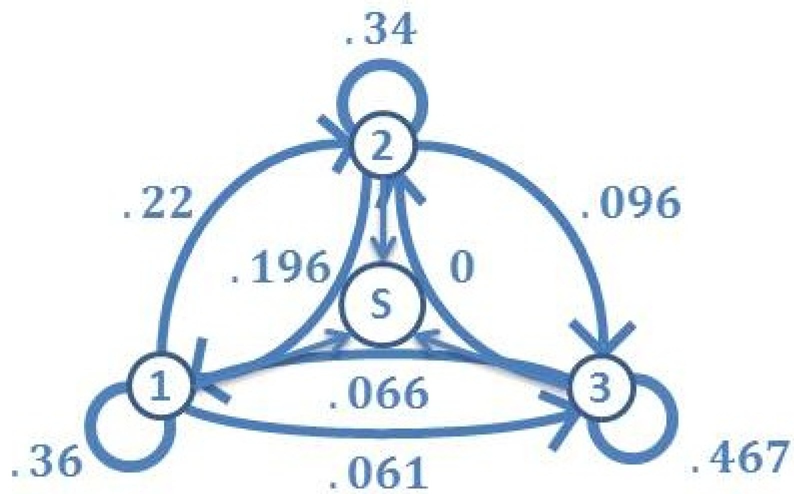

Figure 3.4 Graphical Representation of the Transition Probability Matrix for the 3 Product Example.

Clearly products 1 and 2 are complements, products 1 and 3 are substitutes, and products 2 and 3 are substitutes. Both D_3 and b_3 were used to produce sales data with and without noise and to calculate the set of optimal prices, which was used to help develop the Markov models. The Markov models used for this example followed the form of Equation 3.4.

The transition probability matrix for the first simulated day was,

$$P_3 = \begin{bmatrix} 1.0000 & 0.360 & 0.365 & 0.467 \\ 0 & 0.360 & 0.196 & .066 \\ 0 & 0.22 & 0.34 & 0 \\ 0 & 0.061 & 0.096 & 0.467 \end{bmatrix}. \tag{3.14}$$

The graphical representation of this Markov model is in Figure 3.4. The initial state probability distribution, determined by the projected daily sales from Equation 3.10 and 3.11, for the first day was,

$$I_3 = \begin{bmatrix} 0.1685 & 0.2753 & 0.5562 \end{bmatrix}^T. \tag{3.15}$$

Daily transactions were generated using this information and the approach presented in the last section. After the simulation produced 30 days of data, regressions were performed on the daily sales from the simulated transactions data, $\hat{A}_{Markov}$, and the daily sales with noise, $\hat{A}_{Control}$, as a control:

$$\hat{A}_{Markov} = \begin{bmatrix} -160.9623 & -27.6099 & 38.2550 \\ -69.9421 & -104.3277 & 72.6203 \\ 40.3840 & 36.7952 & -205.1631 \\ 267.8698 & 220.8330 & 255.4532 \end{bmatrix} \tag{3.16}$$

$$\hat{A}_{Control} = \begin{bmatrix} -171.8789 & -42.3822 & 66.0958 \\ -75.2023 & -80.9468 & 68.5338 \\ 43.2317 & 29.3244 & -198.2942 \\ 264.4060 & 216.1419 & 241.8008 \end{bmatrix}. \tag{3.17}$$

Both sets of results were sign symmetric, which is necessary for linear demand models of this form. Both were also consistent with the sign symmetry of the original demand equations.

Another check for accuracy used was calculating the Frobenius norm (FN) of the difference between the original linear demand model and the estimated demand model and dividing it by the FN of the original A. The FN of a matrix, A, is calculated by,

$$FN(A) = \sum_{i,j} a_{ij}^2. \tag{3.18}$$

The FN, referring to the normalized error, for the Markov Model was .0452, much better than the FN for the control regression, .0803. The code used for this experiment is in Appendix A.

Chapter 4

Simulations and Results

4.1 Simulation Methodology

A number of simulation experiments were conducted to test the consistency of the Markov model with the demand model it was generated from. To accomplish this, each experiment began with a random demand function, and the process for generating a transactions model was implemented to develop a Markov model for transactions. This Markov model was then used to generate transactions data for a specified number of days. This data was then aggregated into daily sales data and regression was used to estimate a demand function from this sales data. The estimated demand model and the original demand function were then compared to understand the consistency of the transactions Markov model with the original demand function.

These simulation experiments were conducted under different conditions to understand the performance of the transactions Markov model in various situations. Specifically the following experiments were conducted:

- Demand functions with three, five, and ten distinct products were compared.

- For three product markets, a single experiment generated enough transactions to yield 300 days of sales data, and 1000 distinct experiments were conducted.

- For five product markets, a single experiment generated enough transactions to yield 500 days of sales data, and 500 distinct experiments were conducted.
- For ten product markets, a single experiment generated enough transactions to yield 1000 days of sales data, and 100 distinct experiments were conducted.
- Transactions models generated from D, D^T, and the average D, explained in Section 3.2.2, were compared using all the experiments detailed above.

The metrics for success included calculating the Frobenius norm (FN) of the difference between the original A and the $\hat{A}$ estimated from the simulated transactions data and dividing it by the FN of the original A. The result of this calculation is referred to as the FN. Also, since every reasonable demand function is sign symmetric, sign symmetry of the estimated $\hat{D}$ matrix was checked. The third measure was to see if the sign symmetry of the estimated $\hat{D}$ was consistent with the sign symmetry of the original demand model, D. The results report, for each market size, the average FN of the error over the total number of experiments, the fraction of experiments yielding sign symmetry, and the fraction of experiments yielding sign symmetry consistent with the original demand function.

4.2 Control

For each experiment, a control experiment was conducted as a basis for comparison. In this control experiment, the same random demand function was used to directly generate daily sales data by adding white noise, ϵ, to demand as follows:

$$q = Dp + b + \epsilon. \tag{4.1}$$

Linear regression was then performed on this sales data (as opposed to the sales data produced by aggregating transactions data generated by the Markov model), yielding a presumably much closer estimate of the original demand function. The same metrics were then used to evaluate the performance of the control, and these results can be compared in Tables 4.1, 4.2, and 4.3.

The controls yielded expected results. As the number of days increased, the FN of the error decreased, that is, the estimated linear demand model became more accurate. This is depicted in Figure 4.1. Nevertheless, as the number of products

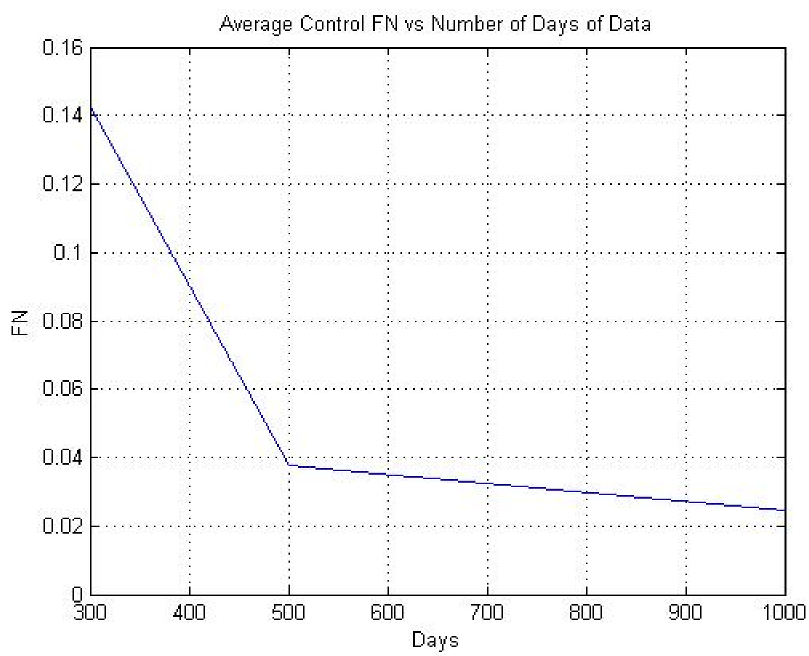

Figure 4.1 The Development of the Accuracy of the Estimation of the Controls over the Number of Days.

increased the sign symmetry suffered. For three products, only nine parameters are estimated but for ten there are 100 parameters. This makes it more difficult to obtain accurate estimates and therefore the sign symmetry suffered. This is depicted in Figure 4.2. There is practically no change from three to five products markets in the FN performance because the increased number of days is being made up for

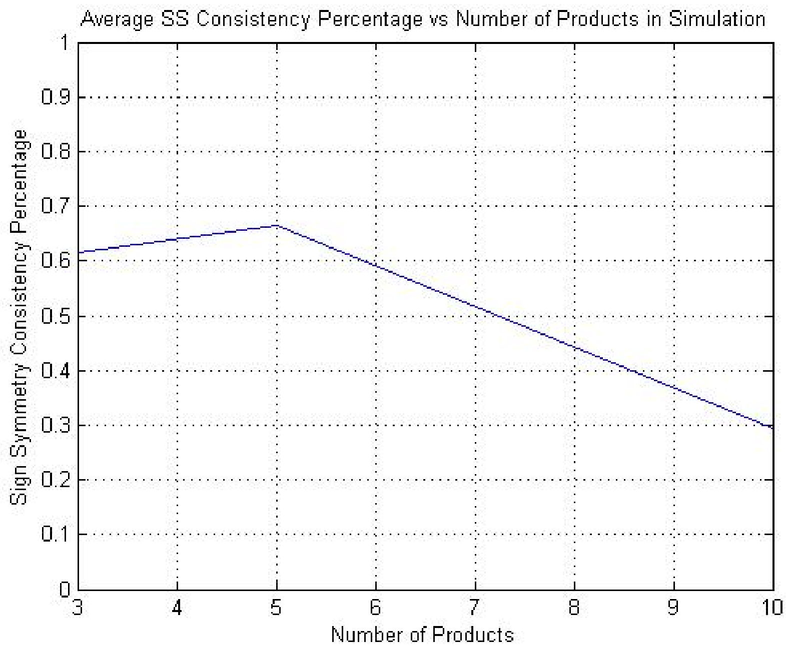

Figure 4.2 The Development of the Accuracy of the Estimation of the Controls over the Number of Days: The y-axis is the average percentage of simulations that produced results that were consistent with the sign symmetry of the original demand model for the control groups.

by the increased number of parameters. However, for ten products there were not a sufficient number of days to account for the increased number of parameters.

4.3 Results

The results of the simulations are presented in Table 4.1, Table 4.2, and Table 4.3. The Markov part of the tables show the results of comparing the original demand model to the demand model estimated by summing up the simulated transactions data into daily sales. The D column indicates the use of Equation 3.4, D^T indicates Equation 3.5, and $(D + D^T)/2$ indicates Equation 3.6.

3 Products		D	D^T	$(D+D^T)/2$
Markov	Average FN	0.2286	0.2634	0.2322
	Sign Symmetric (SS)	607/1000	457/1000	566/1000
	SS Consistency	584/1000	419/1000	539/1000
Control	Average FN	0.1391	0.1488	0.141
	Sign Symmetric (SS)	637/1000	605/1000	638/1000
	SS Consistency	624/1000	587/1000	637/1000

Table 4.1 Average of 3 Products Over 1000 Times for 300 Simulated Days.

5 Products		D	D^T	$(D+D^T)/2$
Markov	Average FN	0.2371	0.2395	0.2176
	Sign Symmetric (SS)	53/500	136/500	138/500
	SS Consistency	45/500	116/500	116/500
Control	Average FN	0.0324	0.0404	0.0404
	Sign Symmetric (SS)	324/500	337/500	338/500
	SS Consistency	323/500	333/500	340/500

Table 4.2 Average of 5 Products Over 500 Times for 500 Simulated Days.

10 Products		D	D^T	$(D+D^T)/2$
Markov	Average FN	0.2535	0.2523	0.2587
	Sign Symmetric (SS)	0/100	1/100	0/100
	SS Consistency	0/100	0/100	0/100
Control	Average FN	0.0258	0.0210	0.0265
	Sign Symmetric (SS)	29/100	29/100	30/100
	SS Consistency	29/100	29/100	30/100

Table 4.3 Average of 10 Products Over 100 Times for 1000 Simulated Days.

4.3.1 Discussion

It is clear from Table 4.1 that, unlike the example in Section 3.4, the Markov model results, on average, underperformed the control regression results. The results for the three product simulations are comparable to the control results but much worse for the five and ten product cases, shown in Figure 4.3. This is because the increased number

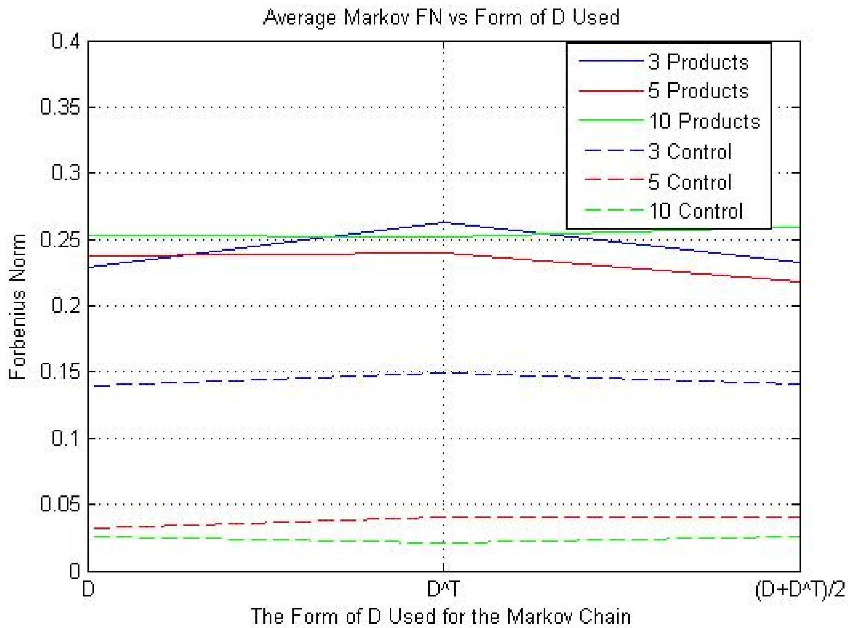

Figure 4.3 The Average Frobenius Norm of the Estimation vs the Form of the $\bar{D}$ Used: The dotted lines show the average FN for the control groups.

of days for the five and ten product simulations improved the control regression but apparently did not affect the Markov results. There is no obvious choice as for which of the three constructions of $\bar{D}$ is the correct answer.

Chapter 5

Conclusions and Future Work

The results of the simulations are inconclusive. It seems that all forms of $\bar{D}$ perform approximately the same. It maybe argued that the results from D^T were the worst but the ten product case showed that they were a little better than the D and much better than the average D. It must be noted that the simulations had rich price data in order to avoid the conditioning problem.

Although the Markov idea seems to be a promising direction for modeling transactions, the particular Markov models explored in this work do not appear to be an effective way to simulate transactions data. One change that could be made to possibly increase performance would be to multiply the $\bar{D}$ matrix by the ratio of the daily price, p_i, over the optimal price, p_i^* the same way the stopping probability is calculated. Another option to explore is another way to improve the construction of the stopping probabilities. The end of day criterion might also have room for improvement.

Although these transactions models did not yield results that were consistent with the original demand function, the idea that substitutes should only occur in the same transactions with low probability and complements should occur in the same transactions with high probability remains compelling. Future work should build off this idea to develop a transactions model that is consistent with the original demand function.

Appendix A

Code for 3 Products

```
% trans6.m
% Creating simulated Transactions data a bunch of times
% 4/27/12
ansCount = 1;
count = 0;
clear Answ;
minFN = 1;
while (count<1000)
    a = zeros (3,1);
    for j=1:3
        r = rand;
        if r<(1/2)
            a(j) = 1;
        else
            a(j) = -1;
        end
    end
    Sign = [ -1 a(1) a(2); a(1) -1 a(3); a(2) a(3) -1];

    A = 100*rand(3,3);
    A = A+ diag(sum ( A));
    A = Sign.*A;

    b = 200 + 100*rand(3,1);

    pstar = -.5*(inv(A)*b);
```

```
    if (sum((pstar>0))==3)
        count = count+1;

        [ Trans , Q, P, Ahat, edgeCP, edgeCF, edgePF, Anoise ] = trans(A, pstar, b

        Sym = sign(Ahat(1:3,:));
        Sym2 = Sym+Sym';
        A = [A; b'];
        Symn = sign(Anoise(1:3,:));
        Symn2 = Symn+Symn';

        if (length(find((Symn2==0)))==0) || (length(find((Sym2==0)))==0)
            Answ{ansCount,1} = A;
            Answ{ansCount,2} = Ahat;
            Answ{ansCount,3} = norm(A-Ahat,'fro')/norm(A,'fro');
            Answ{ansCount,4} = sum(sum(Sym - sign(A(1:3,:))));
            Answ{ansCount,5} = Anoise;
            Answ{ansCount,6} = norm(A-Anoise,'fro')/norm(A,'fro');
            Answ{ansCount,7} = (length(find((Sym2==0)))==0);
            Answ{ansCount,8} = (length(find((Symn2==0)))==0);
            Answ{ansCount,9} = sum(sum(Symn - sign(A(1:3,:))));

            if (norm(A-Ahat,'fro')/norm(A,'fro')<minFN)
                minFN = norm(A-Ahat,'fro')/norm(A,'fro');
                minTrans = Trans;
                minA = A;
                minQ = Q;
                minP = P;
                minEdgecp = edgeCP;
                minEdgecf = edgeCF;
                minEdgepf = edgePF;
                minInd = ansCount;
            end
            ansCount = ansCount+1;
        end
    end
end
countSSS = 0;
for i = 1: length(Answ)
    if (Answ{i,4}==0)
        countSSS= countSSS+1;
    end
end
```

```
countSSSn = 0;
for i = 1: length(Answ)
    if (Answ{i,9}==0)
        countSSSn= countSSSn+1;
    end
end
countSS = 0;
for i = 1: length(Answ)
    if (Answ{i,7}==1)
        countSS= countSS+1;
    end
end
countSSn = 0;
for i = 1: length(Answ)
    if (Answ{i,8}==1)
        countSSn= countSSn+1;
    end
end
FNhat = 0;
for i = 1: length(Answ)
    FNhat= FNhat+Answ{i,3};
end
FNhat = FNhat/length(Answ);
FNnoise = 0;
for i = 1: length(Answ)
    FNnoise= FNnoise+Answ{i,6};
end
FNnoise = FNnoise/length(Answ);

%Trans3 function

function [ Trans3, Q, P, Ahat, edgeCP, edgeCF, edgePF, Anoise ] = trans(A, pstar,

c = pstar(1);
p = pstar(2);
f = pstar(3);
l = 300;
dc = A(:,1);
dp = A(:,2);
df = A(:,3);

Pc = c*ones(l,1);
Pp = p*ones(l,1);
```

```
Pf = f*ones(l,1);
inc = .05;

for i =50:100
    Pc(i) = Pc(i) + inc;
end

dec = -.03;
for i =210:300
    Pc(i) = Pc(i) + dec;
end

inc2 = .08;
for i =1:130
    Pp(i) = Pp(i) + inc2;
end

dec2 = -.01;
for i =160:300
    Pp(i) = Pp(i) + dec2;
end

dec3 = -.05;
for i =70:180
    Pf(i) = Pf(i) + dec3;
end

inc3 = .06;
for i =181:250
    Pf(i) = Pf(i) + inc3;
end

ec = round(10*randn(l,1));
qc = round(dc(1) * Pc + dc(2) * Pp + dc(3) * Pf + b(1));
qc(qc<0) = 0;

ep = round(5*randn(l,1));
qp = round(dp(1) * Pc + dp(2) * Pp + dp(3) * Pf + b(2));
qp(qp<0) = 0;

ef = round(10*randn(l,1));
qf = round(df(1) * Pc + df(2) * Pp + df(3) * Pf + b(3));
qf(qf<0) = 0;
```

```
Q = [qc qp qf];
Qnoise = [qc+ec qp+ep qf+ef];
P = [Pc Pp Pf];

%pstar = ones(k,1);

A = -A;
low = min(min(A));
A = -low + A;
elast = diag(A);
A = [zeros(length(dc),1) A];
ind = 1;
Trans3 = zeros( sum(sum(Q)), 3);

for i = 1:l
    top = elast.*P(i,:)'./pstar;
    top = [1 top'];
    Ast = [top; A];
    su = sum(Ast);
    mu = mean(Q(i,:));
    for j =1:length(Ast(1,:))
        Ast(:,j) = Ast(:,j)/su(j);
    end

    sales = Q(i,:)';
    total=sum(sales);
    salesc = 0;
    salesp = 0;
    salesf = 0;

    while (mean([salesc salesp salesf])<mu )
        r = rand;
        start = zeros(length(sales)+1,1);
        salesProb = sales/total;
        if(r<salesProb(1))
            start(2)=1;
            Trans3(ind,1)=Trans3(ind,1)+1;
            salesc = salesc + 1;
        elseif (r<salesProb(1)+salesProb(2))
            start(3)=1;
            Trans3(ind,2)=Trans3(ind,2)+1;
            salesp = salesp + 1;
```

```
        else
            start(4)=1;
            Trans3(ind,3)=Trans3(ind,3)+1;
            salesf = salesf + 1;
        end

        transProb = Ast*start;
        r = rand;
        while (1==1)
            if (r<transProb(1))
                break
            elseif(r<sum(transProb(1:2)) && (mean([salesc salesp salesf])<mu ))
                Trans3(ind,1)=Trans3(ind,1)+1;
                salesc = salesc + 1;
            elseif (r<sum(transProb(1:3)) && (mean([salesc salesp salesf])<mu ))
                Trans3(ind,2)=Trans3(ind,2)+1;
                salesp = salesp + 1;
            elseif ((mean([salesc salesp salesf])<mu ))
                Trans3(ind,3)=Trans3(ind,3)+1;
                salesf = salesf + 1;
            else
                break
            end
            transProb = Ast*transProb;
            r = rand;
        end
        ind = ind + 1;
    end

    % Row of zeros indicates the end of the day

    ind = ind+1;

end

Qsim = daily(Trans3,l);

[edgeCP, edgeCF, edgePF] = edges(Trans3);

Ahat = regress(Qsim, P);

Anoise = regress(Qnoise, P);
```

Bibliography

[1] Bertsekas, Dimitri and John Tsitsiklis. (2008). Introduction to Probability. Athena Scientific.

[2] Dahleh, Mohammed, Dahleh, Munther, and George Verghese *Lectures on Dynamic Systems and Control*, MIT OpenCourseWare: EECS 6.241 Dynamic Systems and Control, Ch 4.

[3] Donald, G. H. *Interactive system for retail transactions*, US Patent 5,053,956, 1991.

[4] Gujarati, Damodar. (2003). Basic Econometrics. McGraw-Hill.

[5] Horn, Roger and Charles Johnson. (2010). Matrix Analysis. Cambridge University Press.

[6] Lazear, Edward P. *Retail Pricing and Clearance Sales.* American Economic Review 76.1 (1986): 14-32.

[7] Makridakis, Spyros, Wheelwright, Steven, and Rob Hyndman. (1998). Forecasting: Methods and Applications. Hoboken: Wiley and Sons.

[8] Nicholson, Walter and Christopher Snyder. (2007). Intermediate Microeconomics and its Application. Thomson, p. 114115.

[9] Oxenfeldt, A.R., *Pricing decisions: how they are made and how they are influenced*, Management Review, November 1983, pp.23-5.

[10] Wada, Ryosuke, *Stochastic Structure of Brokered Foreign Exchange Auctions*, Forecasting Financial Markets: Advances for Exchange Rates and Asset Management, 2005.

[11] Wang, Shi, *Adaptive Online Retail Web Site Based on Hidden Markov Model*, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 2000.

[12] Woo, Nathaniel, *Data Denoising for Demand Forecasting in Retail*, Honors Thesis. Brigham Young University, 2010.