Performance Measurement in the eCommerce Industry

by

Simon Donkor

A Professional Masters Project
Submitted to the Faculty
of
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirements for the
Degree of Professional Master of Science
in
Financial Mathematics
by

May 2003

APPROVED:

Professor Balgobin Nandram, Major Advisor

Professor J. J. Mistry, Advisor

Professor Bogdan Vernescu, Department Head

Abstract

The eCommerce industry introduced new business principles, as well as new strategies for achieving these principles, and as a result some traditional measures of success are no longer valid. We classified and ranked the performance of twenty business-to-consumer eCommerce companies by developing critical benchmarks using the Balanced scorecard methodology. We applied a Latent class model, a statistical model along the Bayesian framework, to facilitate the determination of the best and worst performing companies.

An eCommerce site's greatest asset is its customers, which is why some of the most valued and sophisticated metrics used today evolve around customer behavior. The results from our classification and ranking procedure showed that companies that ranked high overall also ranked comparatively well in the customer analysis ranking, For example, Amazon.com, one of the highest rated eCommerce companies with a large customer base ranked second in the critical benchmark developed towards measuring customer analysis. The results from our simulation also showed that the Latent class model is a good fit for the classification procedure, and it has a high classification rate for the worst and best performing companies. The resulting work offers a practical tool with the ability to identify profitable investment opportunities for financial managers and analysts.

Acknowledgements

My greatest appreciation beyond words goes to my first advisor, Professor Balgobin Nandram, with his great help, patience, and guidance I have produced a work that exceeds any of my prior expectations and that I take great pride in. Thanks to Professor J. J. Mistry for his vision and guidance in this project, and introducing me to the Balanced scorecard. My sincere thanks to my mentor, Professor Domokos Vermes, for his guidance and insight in this field of study.

Also I'd like to thank my friend, Nordia Thomas, for her kindness in providing the LaTeXtemplate and her help as an editor of this project.

Finally, I extend a special thanks to my wife, Sophia, for her understanding, support, and humor in helping me through this program.

To my mother, father, brothers and sister

thank You for all the love

you have given me.

Contents

Chapter 1

Balanced scorecard

1.1 Introduction

The goal of this project is to generate measures from the qualitative idea Balanced scorecard (BSC) approach to quantify performance in the individual perspectives of the BSC. With these measures we apply a statistical model to classify and rank twenty eCommerce companies.

In this Chapter 1 we begin with the introduction of the BSC, discuss the four critical perspectives from which we intend to generate our relevant metrics, and illustrate with an example from a recent paper *The Managers Online Reference* from CEOReview.com [2]. This paper presents measures under the four perspectives currently in use by some organizations.

In Chapter 2 we describe the data obtained from the eCommerce Almanac, a data summary - a compilation of the data under the four critical perspectives, and a table of the twenty business-to-consumer (B2C) eCommerce companies we intend to classify.

In Chapter 3 we describe the methodology and computational approach of the latent class model (LCM), a statistical model along the Bayesian framework,

to facilitate the determination of the best and worst performing companies. In Chapter 4 we describe the analysis of the LCM using the Bayesian cross-validation analysis, discuss the sensitivity to different transformations and its importance to the choice of the transformation that best fits our LCM. And finally present a brief result from a simulation study of the LCM.
In Chapter 5 we present our conclusion, a discussion and comparison of the overall classification and ranking process from the LCM.

1.2 A brief overview of the Balanced Scorecard (BSC)

"The Balanced Scorecard has long been thought of as the premier tool for measuring corporate strategy," said Stan Smith president and CEO of Open Ratings. The idea of the BSC was first introduced by Kaplan and Norton in the February 1992 issue of the *Harvard Business Review*. The BSC is a formal management technique built on the premise that the main prerequisite to effective management is measurement [8]. It provides a realistic framework that links measurement, on both quantitative and qualitative criteria, to strategic objects [9].
Therefore, in the framework of the BSC, a balanced view of organizational performance must include measures that indicate performance in at least four areas: Financial, Customer, Internal Business Processes, and Learning and Growth.

1. **Financial** - success in achieving mission from the perspective of the shareholder.

2. **Customer** - strategy for creating value in service

3. **Internal Business Processes** - strategic priorities for various business processes that create customer and shareholder satisfaction.

4. **Learning and Growth** - the urge for innovation consistent with vision and business strategy.

The implementation of the BSC begins with the setting of goals, and then the strategies to achieve them, in four critical perspectives [10]. Figure 1.1 shows how these four critical perspectives are linked to the company mission. These common objectives (metrics) once chosen will facilitate comparative analysis and benchmarking in the classification and ranking of the eCommerce companies.

A recent paper (CEOReview.com) [2] on measures in each of the four perspectives of the BSC gives a comprehensive view of the four perspectives of the BSC and some very relevant measures:

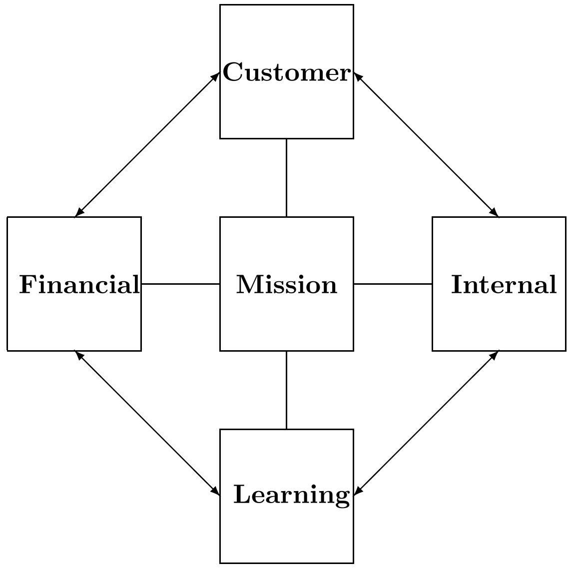


Figure 1.1: The Balanced Scorecard

1.3 Financial Measures

There are three general objectives or themes that are typically reflected in the financial perspective of a Balanced Scorecard: Revenue Growth, Cost Management, and Asset Utilization. We can identify measures for each of these objectives by answering the question, "How can this objective be achieved?"

Revenue Growth

1. Sales and market share
2. Number of new products, or new applications of existing products and services
3. Number of customers and markets
4. Number of new market channels, differentiating on service, delivery

mode and price

5. Number of new pricing strategies

Cost Management

1. Revenue per employee
2. Unit cost reduction
3. Percent use of low cost business processes. (e.g. increase use of EDI to replace costly manual purchasing approaches)
4. Percentage of expenses measured by Activity Based Costing

Asset Utilization

1. Inventory reduction,and increased turns
2. Cash-to-cash cycle
3. Return on capital
4. Productivity /efficiency

1.4 Customer Measures

Before establishing customer measures, organizations must identify the market segments they are serving or wish to serve. Organizations may select market segments that are most profitable, or that are under-served. For each segment consider customizing the following set of widely-used measures to the specific characteristics of your business: market share, customer retention, customer acquisition, customer satisfaction, and customer profitability.

Market Share

1. Percent of market segment captured by your organization

2. Percent of each customer's total requirement served by your company (e.g. for customer's purchasing clothing at your apparel store, what portion of their total annual clothing budget do they spend with you?)

Customer Retention

1. Number of defections (customers who take their business elsewhere)

2. Increase in sales to current customers

3. Frequency of orders/visits/contacts with current customers

Customer Acquisition

1. Number of new customers, or total sales to new customers

2. Ratio of sales to inquiries

3. Average cost to acquire a new customer

4. Average order size, or average revenue per customer interaction

Customer Satisfaction

1. Number of complaints

2. Number of unsolicited thank you letters

3. Number of individuals indicating that they are extremely satisfied with their experience with your organization on a satisfaction survey

Customer Profitability

1. Total profit per customer

2. Total cost per customer or per transaction

Perhaps more than any other perspective, the customer dimension of a Balanced Scorecard affords opportunities to learn about and transform a business. We have summarized typical quantitative measures to assess performance with customers. However, the customer perspective also provides rich opportunities to obtain qualitative data. For example, comments and complaints by customers on satisfaction surveys may be more important than the satisfaction level they express. Analysis of this information may lead to identification of new market segments, or new product/service opportunities, or many other transformations in a business.
Indeed the customer perspective of the Balanced Scorecard provides opportunities to go beyond core measures to those that are even more strategic, reflecting the value proposition offered to each market segment. By value proposition, we mean the unique combination of product attributes, image and relationship characteristics that define the interaction with customers.

1.5 Internal Business Measures

There are many internal processes in the typical organization that deserve attention and measurement. But measuring and managing these processes can only drive incremental improvements, and do not contribute to the strategic management of that organization. It may be appropriate to include measures about the accounts receivable process in a Balanced Scorecard for the accounts receivable department. A Balanced Scorecard for the strategic business unit, on the other hand, needs to reflect the entire value chain. We need measures of organizational performance all the way from the identification of a customer need to the satisfaction of that customer need.

Identify or Make the Market

1. Profitability by market segment

2. Percent of revenue from new products

3. Percent of revenue from new customers

Design

1. Time to market

2. Break-even time

Build

1. Number of defects

2. Process time

3. Process cost

Deliver

1. Percent on-time delivery

2. Percent defects

3. Stock-out

Service (post-sales)

1. Average satisfaction rating
2. Number of customers re-ordering within a three-month period
3. Number of customers who do not re-order again within a year
4. Number of deliveries during which a related product or service is cross-sold

1.6 Learning and Growth Measures

The learning and growth perspective of the Balanced Scorecard focuses in the organizational infrastructure that is required in order to achieve objectives in the other areas. These are the three common categories for learning and growth measures: employee capabilities, information technology, and motivation, empowerment and alignment. Here are a few examples of measures for the learning and growth perspective.

Employee Capabilities

1. Employee satisfaction (involvement, recognition, access to information, support from staff functions, etc.)
2. Staff turnover
3. Productivity (revenue per employee, return on compensation, profit per employee, etc.)

4. Number of employees qualified for key jobs relative to anticipated requirement

Information Technology

1. Information coverage ratio - number of processes having adequate information on quality, cycle time, and cost
2. Percent of customer information available during front-line interactions
3. Return on data - new revenue per database, etc.

Motivation and Alignment

1. Suggestions received
2. Suggestions implemented
3. Rewards provided
4. Length of time required to improve a key measure such as on-time deliveries by 50% (half-time metric)
5. Percentage of employees with objectives aligned by key Balanced Scorecard measures

1.7 Application of the BSC to eCommerce

Much of corporate wealth is now being accumulated through non-financial means, this implies that it is becoming increasingly important to include intangibles, often referred to as intellectual, human, social or relational capital, in company reports [6].

A constraining feature when classifying the performance of the eCommerce industry by performance indicators is that there is no correct number of metrics to establish. The measuring process of these individual indicators is currently a subject of research and that obtaining a universally acceptable formulation for measurement may prove difficult. Despite these difficulties, we hope to extract all the relevant figures from the profile of a company in the eCommerce Almanac as the base of our data set.

We begin by generating the most important or relevant metrics that can expose the overall operational performance of the eCommerce industry under discussion.

The challenge in the next chapter is to generate a set of acceptable and transparent metrics that can quantify performance in the individual perspectives, and also reflect well on our classification and ranking process.

Chapter 2

The Data Set

2.1 Description of the Data

The *eCommerce Almanac*, published by the Intermarket Group, compiles information on most of the sophisticated online retailers. It profiles both leading business-to-consumer (B2C) and business-to business (B2B) companies. Although, data on both B2C and B2B companies was available and could be considered to obtain a larger data set, comparing their operational performance is pointless since their business practices are completely different. Therefore, in this project, we focus only on the B2C companies.

From the *eCommerce Almanac*, the profile of a particular B2C company was subdivided into sections that classified a particular B2C company under principal headings: *organization, website overview, marketing, management, internet infrastructure, and operating benchmarks.* Companies with inadequate information under these subsections were eliminated.

In Table 2.1 we present the data for a list of twenty companies in alphabetical order, with data gathered under sections outlined as: unique visitors, revenue, marketing expenditure, development expenditure, financing expenditure, reach of the company, number of employees, and profitability.

2.2 Definitions

Unique visitors - the number of visitors who visit a site more than once.

Revenue - the dollar amount of sales during the period considered.

Marketing expenditure - amount used to attract prospects, which includes market research on target groups, sales support, mass advertising, etc.

Development expenditure - the expenditure geared towards investing in strategies for superior performance: building strong customer loyalty, good internal-business-processes and excellent employees, systems, and organizational alignment.

Financing expenditure - the overall income invested in a company from the capital markets, including venture capitalist and initial public offerings (IPO).

Number of employees - full time employee counts as of January 2000.

Profitability - revenue generated less tax.

Reach - the reach is more or less the penetration level. The reach was obtained by assessing the proportion of users of a particular site to the entire profiled sites.

Number	Company	Number	Company
1	Amazon.com	11	E*TRADE Group Inc
2	Autobytel.com Inc	12	Fogdog Inc
3	Autoweb.com	13	FTD.com
4	Bolt Inc	14	Furniture.com Inc
5	CarsDirect.com Inc	15	iOwn
6	CDnow Inc	16	NetB@nk
7	800.com Inc	17	Nextcard Inc
8	drugstore.com Inc	18	Peapod Inc
9	E-Loan Inc	19	PlanetRx.com Inc
10	eToys Inc	20	Webvan Group Inc

Table 2.1: Representation of companies by numbers

	Unique	Total	Marketing	Development				
Mill.	Visitors	Revenue	Expdt.	Expdt	Financing	Reach	Employees	Profit
COM	U.V	TR	ME	DE	F	RH (%)	E	P
1	14.81	1,640.00	413.20	159.70	2,680.00	21.9	0.0076	-719.97
2	1.00	40.30	44.18	14.26	141.96	1.5	0.0002	-23.32
3	2.20	32.80	33.20	5.10	104.20	3.3	0.0002	-18.15
4	1.16	4.40	9.08	3.52	56.80	1.7	0.0002	-12.92
5	1.30	15.18	14.57	2.23	488.08	1.9	0.0007	-72.33
6	6.65	147.19	89.73	23.42	260.30	9.9	0.0005	-119.23
7	0.89	3.00	8.90	1.20	83.30	1.3	0.0001	-42.81
8	1.59	34.80	61.50	14.90	230.00	2.4	0.0004	-115.80
9	0.53	22.10	30.29	3.60	249.74	0.8	0.0004	-72.98
10	1.16	29.96	20.72	3.61	724.20	1.5	0.0009	-189.63
11	2.46	621.40	301.70	76.90	1,862.00	3.6	0.0024	-27.98
12	1.01	6.99	21.45	3.45	145.50	1.5	0.0001	-29.61
13	0.99	49.60	11.99	2.16	49.00	1.5	0.0001	-23.56
14	0.89	10.90	33.95	6.69	84.00	1.3	0.0002	-46.46
15	0.33	14.77	19.13	10.39	59.27	0.5	0.0003	-49.83
16	0.86	56.43	7.36	1.40	238.42	1.3	0.0001	3.05
17	3.90	26.56	24.65	22.05	384.47	5.8	0.0004	-77.20
18	0.17	73.13	7.17	3.54	145.00	0.3	0.0010	-28.45
19	1.43	8.99	55.18	12.95	144.50	2.1	0.0004	-98.01
20	0.14	13.31	11.75	15.24	966.03	0.2	0.0010	-144.60

Table 2.2: A data summary of 20 Pure-play B2C eCommerce companies

2.3 Metrics

To maintain some consistency in our calculations and provide meaningful comparison in our later analysis, we generate performance indicators that are most relevant to our data set. For each perspective we generate no less than two measures. The critical measures for each perspective are not interchangeable. Our criteriion for assessing this is basically from the performance measure we expect to quantify. Therefore, in obtaining these metrics in the framework of the BSC we put together an overall performance measurement under one framework.

2.3.1 Financial perspective

In the financial perspective the most traditionally used performance indicator, includes assessment of measures such as operating costs and return-on-investment [10]. In this project, and from our data set our main targets are returns, financing and revenue from sales. Our goal is to keep to the old measures of financial ratios, but we observed that not all the ratios are common to all the twenty companies, because the *eCommerce Almanac*, from which we obtained our data set is not enough to generate all the well known financial ratios, resulting in only three measures under the financial perspective.

Metrics

TR/F	Financing (Investment) ratio - Total revenue per amount of financing
AP/TR	Return on sales - Profitability per total revenue (from sales)
AP/F	Return on financing (investment) - Profitability per amount of financing

Description

Return on Financing (TR/F) - Like a return on investment ratio, this ratio targets the profitability from the financing expenditure, this is a data set relevant metric, not a conventional metric.

Return on Sales (AP/TR) - This ratio compares after tax profit to sales. If a company is experiencing a cash flow crunch, it could be because its mark-up is not enough to cover expenses. This metric, is very much relevant to eCommerce companies due to their pertinent cash flow problems. It helps determine if a company is making enough of a return on the sales effort.

Financing Ratio (TR/F) - Like the return on investment ratio, this ratio also targets amount of sales from financing expenditure, this is a data set relevant metric, not a conventional metric.

2.3.2 Customer Perspective - metrics and description

In the customer perspective our target is to quantify customer satisfaction and retention. The metric, *MC*, defined as the marketing expenditure per unique visitors, quantifies customer retention. Similarly, the other two metrics, *TR/UV* and *MS*, quantify customer satisfaction.

Metrics

TR/UV	Revenue generated by unique visitors (UV) - Total revenue per UV
MC	Marketing coverage - Marketing expenditure per unique visitors
MS	Penetration (market share) - Reach (% of users captured by a company)

Description

Revenue generated by unique visitors (TR/UV) - This is the total revenue generated by persons who visit a web site more than once within a specified period of time. This is a data set relevant metric, not a conventional metric.

Marketing coverage (MC) - The expenditure geared towards attracting traffic or visitors.This is a data set relevant metric, not a conventional metric.

Penetration (MS) - Percentage of users captured by a company.

2.3.3 Business Processes Perspective - metrics and description

In the internal business processes perspective we target production and innovation. The purpose is to analyze the sources of productivity and to find ways of generating revenues from effective business practices. We have two measures, *EP1* and *EP2* targeting employee productivity. Despite their overall similarity, by definition the two measures lead to substantially different conclusions. A company can generate substantial revenues from sales yet make a loss because of other negative items from the profit and loss account. Therefore, we set these measures in the internal business process as a way to target employee productivity obtained from quality business practices.

Metrics

TR/ME	Revenue generated by Mktg. Expdt. - Total Rev. per Mktg. Expdt.
EP1	Employee Productivity - 1) Revenue per Employee
EP2	Employee Productivity - 2) Profitability per Employee

Description

Revenue generated by marketing expenditure (TR/ME) - This is the total revenue generated from all marketing processes. This is a data set relevant metric, not a conventional metric.

Revenue per employee (EP1) - Total revenue divided by number of employees.This is a conventional metric.

Profitability per employee (EP2) - A productivity indicator, is the income generated divided by number of employees. This is a conventional metric.

2.3.4 Learning & Growth Perspective - metrics and description

In the learning and growth perspective we generate two measures, *EDC*, and *TR/DE*. From these measures we hope to quantify the effectiveness of management in terms of employee satisfaction and retention.

Metrics

EDC	Employee Development Coverage - Dev. Expdt. per Employee
TR/DE	Revenue generated by Dev. expdt - Total Rev. per Devpt. Expdt

Description

Employee development coverage - Workshop and individual consultations to enhance employee skills. It is the total development expenditure divided by number of employees. This is a conventional metric.

Revenue generated from development expenditure - The total revenue obtained as a result of the development projects. Assumed to be the same as the total revenue divided by the development expenditure. This is not a conventional metric.

	Financial			Customer			Internal Business			L.&Growth	
	TR/F	AP/TR	AP/F	TR/UV	MC	MS	TR/ME	EP1	EP2	EDC	TR/DE
1	2.367	0.001	0.006	1.082	0.943	4.400	1.505	1.095	0.004	1.22	1.039
2	1.098	0.302	1.174	0.394	1.493	0.305	0.346	0.909	1.176	3.679	0.286
3	1.218	0.373	1.61	0.146	0.510	0.671	0.375	0.763	1.222	1.358	0.651
4	0.300	2.801	2.974	0.037	0.265	0.345	0.184	0.131	1.577	1.202	0.127
5	0.120	0.751	0.32	0.114	0.378	0.386	0.395	0.11	0.353	0.184	0.689
6	2.187	0.072	0.561	0.216	0.456	2.012	0.622	1.487	0.462	2.708	0.636
7	0.139	3.952	1.951	0.033	0.339	0.264	0.128	0.122	2.064	0.557	0.253
8	0.585	0.308	0.638	0.214	1.306	0.488	0.215	0.433	0.571	2.120	0.236
9	0.342	0.515	0.625	0.410	1.942	0.163	0.277	0.320	0.708	0.597	0.621
10	0.160	0.319	0.181	0.253	0.605	0.305	0.548	0.162	0.221	0.223	0.840
11	1.291	0.019	0.089	2.468	4.143	0.732	0.781	1.314	0.110	1.860	0.818
12	0.186	1.726	1.136	0.068	0.721	0.305	0.124	0.259	1.916	1.462	0.205
13	3.916	0.245	3.400	0.491	0.410	0.305	1.568	3.355	3.527	1.672	2.324
14	0.502	1.082	1.925	0.120	1.291	0.264	0.122	0.260	1.205	1.822	0.165
15	0.964	0.795	2.716	0.444	1.989	0.102	0.293	0.260	0.887	2.094	0.144
16	0.916	0.223	0.723	0.644	0.290	0.264	2.907	3.491	3.336	0.988	4.094
17	0.267	0.426	0.403	0.067	0.213	1.179	0.408	0.369	0.675	3.507	0.122
18	1.951	0.165	1.142	4.132	1.401	0.061	3.866	0.364	0.258	0.201	2.091
19	0.241	1.223	1.042	0.062	1.307	0.427	0.062	0.117	0.613	1.928	0.070
20	0.053	0.771	0.146	0.929	2.836	0.041	0.429	0.068	0.223	0.885	0.088

Table 2.3: “Standardized” data summary with metrics under the four perspectives

As an example of how the methodology might work, an organization might include in its mission statement a goal of maintaining employee satisfaction. This would be the organization’s vision. Strategies for achieving this vision might include approaches such as increasing employee-management communication. Tactical activities undertaken to implement the strategy could include, for example, regular scheduled meetings with employees. Finally, metrics could include quantifications of employee suggestions or employee surveys [10].

With the data set obtained from Table 2.2 we incorporate our relevant metrics by computing the ratios as shown in Table 2.3. We let y_{ij} be the component that represents the ith company under the jth perspective, i goes from 1 to ℓ populations (in our case, 20 companies) with m performance indicators (in our case, 11 measures). We divide each entry y_{ij} by the standard deviation of each particular metric to make our y_{ij} comparable. To correct for skewness present in our data, we transform the data set by taking the logarithm of y_{ij} to obtain a symmetric spread of values.

The structure of many data sets is too complex to be represented by a single parametric model, for example, the normal regression model. Nonparametric analysis is one way of circumventing the problems raised by the complexity of observed structures.

In the next chapter we will introduce a Latent class model to retain the framework of parametric densities and approximate the underlying density as a mixture model with latent variables.

Chapter 3

The Latent Class Model

3.1 The Objective of the Model

In this chapter we will consider a model that can help classify or rank a set of B2C eCommerce companies into relative categories of winners, losers and neutrals, a task that is prohibitively expensive considering $n!$ orderings if n companies are considered. Therefore, in the case of classifying 20 B2C eCommerce companies, we have to compute 20! permutations, which is quite a large number. To avoid this we introduce a clustering procedure facilitated by a sampling based Monte Carlo method. First, we cluster the twenty eCommerce companies, then once we have these clusters we rank them into relative categories of winners and losers, resulting in a reduction of the huge expected computations.

One major contribution to this project is the inclusion of latent variables in a normal regression model. This innovative application makes the resulting classification and ranking procedure simple and robust.

The statistical problem is to use the mean performance across different perspectives for each of the twenty companies in estimating and selecting our clusters of winners, losers, and neutrals.

The expected final result of this work is a computerized methodology to classify and categorize companies into winners, losers and neutrals. The goal is to offer a practical tool for financial managers and analyst.

3.2 Earlier Related Research

Gupta and Panchapakesan [5], provided a comprehensive discussion of methods, techniques, and approaches to ranking and selection problems mainly within the non-Bayesian framework. The general method is to find an appropriate parameter (e.g. population mean, variance) which is to be used as a measure to compare the populations.
Goldstein and Spiegelhalter [4], described statistical issues in ranking institutions in the areas of health and education based on outcome data by using certain performance indicators. They obtained interval estimates of the ranks of these indicators for the different institutions, using both Bayesian and non-Bayesian methods.
Morris and Christiansen's Bayesian approach on selection and ranking [7] used a simple two-level empirical Bayes model to select the best mean. They generated samples from the product normal posterior distribution of the means and obtained posterior probabilities that each of the means is the largest. The sampling based approach of these authors is akin to what we consider.

3.3 Description of the Model

We represent our data by the vector $\underset{\sim}{y} = \{y_{ij},\ i = 1, \ldots, \ell,\ j = 1, \ldots, m\}$.

Our model has three parts:

Part (a)

$$y_{ij}|\mu_i, \nu_j, \sigma_1^2 \overset{iid}{\sim} \text{Normal}\left(\mu_i + \nu_j, \sigma_1^2\right), \quad i = 1, \ldots, \ell, \quad j = 1, \ldots, m \tag{3.1}$$

where

$$\begin{aligned} \mu_i &= \text{Effect for the } i^{th} \text{ eCommerce company} \\ \nu_j &= \text{Effect for the } j^{th} \text{ measure} \\ \sigma_1^2 &= \text{Variation} \end{aligned}$$

and

$$\nu_j|\sigma_2^2 \overset{iid}{\sim} \text{Normal}\left(0, \sigma_2^2\right), \quad j = 1, \ldots, m \tag{3.2}$$

Part (b)

We assume μ_i to be independent, and the density function a weighted average of normals.

$$\pi(\mu_i|\underset{\sim}{\theta^*}, \sigma_3^2, \underset{\sim}{\omega}) = \sum_{k=1}^{c} \omega_k \frac{1}{\sqrt{2\pi\sigma_3^2}} e^{-\frac{1}{2\sigma_3^2}(\mu_i - \log(\frac{\theta_k^*}{1-\theta_k^*}))^2} \tag{3.3}$$

where, θ_k^* is the mean for cluster k, and c the number of clusters (groups) of eCommerce companies. $0 < \theta_1^* < \theta_2^* < \ldots < \theta_c^* < 1$, and $\theta_k^* = \frac{(\theta_{k-1} + \theta_k)}{2}$, for $\theta_0 \equiv 0$ and $\theta_c \equiv 1$, $k = 1, \ldots, c-1$.

We defined the joint density of the order statistics, the cut points of the clusters as:

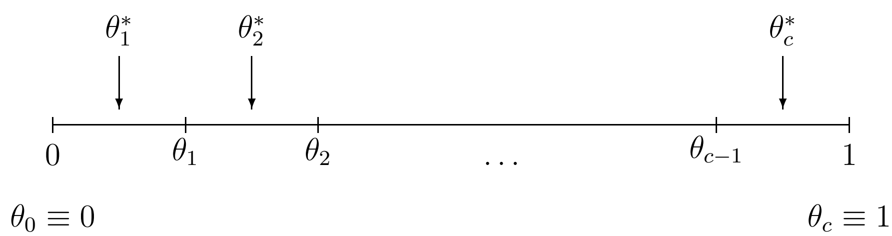
$$\pi(\underset{\sim}{\theta}) \propto \begin{cases} (c-1)!, & \theta_1 < \theta_2 < \ldots < \theta_{c-1} \\ 0, & \text{otherwise.} \end{cases}$$

For computational reasons it's preferable to work with θ_k^* or θ_k, hence the introduction of the logit scale.

Note that these θ_k's are the "boundaries" of the clusters.

Therefore, we draw samples of $\theta_1, \ldots, \theta_{c-1}$ using the grid method, where θ_k^*, also goes from $k = 1, \ldots, c$. The reason for the grid method is to include the upper and the lower bounds of the cut points.

For example, if we consider c clusters, then there must be $c-1$ cut points.



That is, $0 \leq \theta_1 \leq \theta_2 \leq \ldots \leq \theta_{c-1} \leq 1$.

We also assume a proper noninformative prior on the weights, $\underset{\sim}{\omega}$, such that,

$$\underset{\sim}{\omega} \sim Dirichlet(1, \ldots, 1) \tag{3.4}$$

i.e.,

$$\pi(\underset{\sim}{\omega}) = \begin{cases} 1, & \sum_{k=1}^{c} \omega_k = 1, \quad \omega_k \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Our goal is to make the ω_k's bounded, and also to keep the number of eCommerce companies in each cluster an equal a priori.
We take $d \leq \omega_k \leq e$, where d and e are known.

Part (c)

The variances are distributed as:

$$\sigma_1^{-2}, \sigma_2^{-2}, \sigma_3^{-2} \overset{iid}{\sim} \text{Gamma}\left(\frac{a}{2}, \frac{b}{2}\right), \quad a = b = 0.002 \tag{3.5}$$

The choice of a is to provide a proper but noninformative prior for σ_s^{-2}, $s = 1, 2, 3$.

3.4 Introduction of the Latent Variables

We now provide a simplification of the model by the introduction of latent variables. We define the vector $\underset{\sim}{z_i}$ as:

$$\underset{\sim}{z_i} = (z_{i1}, \ldots, z_{ic}).$$

From the above vector, we are able to ascertain the particular cluster a to which a company belongs. For example, if $\underset{\sim}{z_1}$ generates a vector of the form:

$$\underset{\sim}{z_1} = (0, 1, 0, \ldots, 0).$$

we can conclude that company 1 is in cluster 2, from the position of the number 1 in the vector $(0, 1, 0, \ldots, 0)$.

Thus, this allows us to simplify assumption (3.3) as follows:

$$\pi(\mu_i|(z_{ik}=1,\ z_{i\acute{k}}=0,\ k\neq\acute{k}),\underset{\sim}{\theta}^*,\sigma_3^2)$$
$$=\frac{1}{\sqrt{2\pi\sigma_3^2}}\exp\left(-\frac{1}{2\sigma_3^2}(\mu_i-\text{logit }(\theta_k^*))^2\right) \quad (3.7)$$

$$\Pr(z_{ik}=1,\ z_{i\acute{k}}=0,\ k\neq\acute{k}|\ \underset{\sim}{\omega})=\omega_k,\quad k=1,\ldots,c \quad (3.8)$$

$$\pi(\mu_i,(z_{i_k}=1,\ z_{i\acute{k}}=0,\ k\neq\acute{k})|\ \underset{\sim}{\theta}^*,\sigma_3^2,\underset{\sim}{\omega})$$
$$=\prod_{k=1}^{c}\left[\frac{\omega_k}{\sqrt{2\pi\sigma_3^2}}\exp\left(-\frac{1}{2\sigma_3^2}(\mu_i-\text{logit }(\theta_k^*))^2\right)\right]^{z_{ik}} \quad (3.9)$$

so that the latent structure of our model can be represented as

$$\pi(\underset{\sim}{\mu},\underset{\sim}{z}|\underset{\sim}{\theta}^*,\sigma_3^2,\underset{\sim}{\omega})=\prod_{i=1}^{\ell}\prod_{k=1}^{c}\left[\frac{\omega_k}{\sqrt{2\pi\sigma_3^2}}\exp\left(-\frac{1}{2\sigma_3^2}(\mu_i-\text{logit }(\theta_k^*))^2\right)\right]^{z_{ik}} \quad (3.10)$$

3.5 Evaluation of the Joint Posterior Density

Let the set of all parameters (including the latent variables) be denoted by

$\Omega=\{\underset{\sim}{\mu},\underset{\sim}{\nu},\underset{\sim}{\theta},\underset{\sim}{\omega},\underset{\sim}{z},\underset{\sim}{\sigma}^2\}$, and our data, the vector $\underset{\sim}{y}=\{y_{ij},\ i=1,\ldots,\ell,\ j=1,\ldots,m\}$.

Using Bayes' theorem, the joint posterior density of all the parameters is:

$$\begin{aligned}\pi(\Omega|\underset{\sim}{y}) &\propto \pi(\underset{\sim}{y}|\underset{\sim}{\mu},\underset{\sim}{\nu},\underset{\sim}{\theta},\underset{\sim}{\omega},\underset{\sim}{z},\underset{\sim}{\sigma}^2)\pi(\underset{\sim}{\mu},\underset{\sim}{\nu},\underset{\sim}{\theta},\underset{\sim}{\omega},\underset{\sim}{z},\underset{\sim}{\sigma}^2)\\ &= \pi(\underset{\sim}{y}|\underset{\sim}{\mu},\underset{\sim}{\nu},\sigma_1^2)\pi(\underset{\sim}{\mu},\underset{\sim}{z}|\underset{\sim}{\nu},\underset{\sim}{\theta},\underset{\sim}{\omega},\underset{\sim}{\sigma}^2)\pi(\underset{\sim}{\nu},\underset{\sim}{\theta},\underset{\sim}{\omega},\underset{\sim}{\sigma}^2)\\ &= \pi(\underset{\sim}{y}|\underset{\sim}{\mu},\underset{\sim}{\nu},\sigma_1^2)\pi(\underset{\sim}{\mu},\underset{\sim}{z}|\underset{\sim}{\theta},\underset{\sim}{\omega},\sigma_3^2)\pi(\underset{\sim}{\nu}|\sigma_2^2)\pi(\underset{\sim}{\theta})\pi(\underset{\sim}{\omega})\pi(\underset{\sim}{\sigma}^2).\end{aligned}$$

Thus,

$$\begin{aligned}\pi(\Omega|\underset{\sim}{y}) \;\propto\; & \prod_{i=1}^{\ell}\prod_{j=1}^{m}\Big[\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2\sigma_1^2}(y_{ij}-\mu_i-\nu_j)^2}\Big] \times \prod_{i=1}^{\ell}\prod_{k=1}^{c}\Big[\frac{\omega_k}{\sqrt{2\pi\sigma_3^2}}e^{-\frac{1}{2\sigma_3^2}(\mu_i-\text{logit}(\theta_k^*))^2}\Big]^{z_{ik}} \\ & \times \prod_{j=1}^{m}\Big[\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2\sigma_2^2}\nu_j^2}\Big] \times \prod_{s=1}^{3}\Big[(\frac{1}{\sigma_s^2})^{\frac{a}{2}+1}e^{-\frac{b}{2\sigma_s^2}}\Big]. \end{aligned} \tag{3.11}$$

As shown above, the joint density function does not have a clear recognizable form, meaning it will be difficult to make analytical inferences. Therefore, we introduce Markov chain Monte Carlo algorithm to obtain estimates of the posterior distribution of the parameters. We use the Gibbs sampler to draw samples and then use these samples to make posterior inferences.

3.6 Model Fitting and Computations

The Gibbs sampler is an iterative simulation scheme for generating samples that converge to a target distribution. It constructs a Markov chain with the target distribution as its equilibrium distribution.

To perform the Gibbs sampling we need the conditional posterior density (cpd's) for each parameter (including the latent variables) given the others and the data. Let Ω_a be the set Ω excluding the parameter a.

$$\mu_i|\Omega_\mu, \underset{\sim}{y} \overset{iid}{\sim} \text{Normal}\left(\lambda\Big[\frac{\sum_{j=1}^{m}(y_{ij}-\nu_j)}{m}\Big] + (1-\lambda)\Big[\sum_{k=1}^{c} z_{ik}\theta_k^*\Big],\ (1-\lambda)\sigma_3^2\right)$$

where, $\lambda = \frac{\sigma_3^2}{\sigma_3^2 + \frac{\sigma_1^2}{m}}$

$$\sigma_1^2 | \Omega_{\sigma_1^2}, \underset{\sim}{y} \overset{iid}{\sim} \text{Gamma}\left(\frac{\ell m + a}{2}, \frac{b + \sum_{i=1}^{\ell}\sum_{j=1}^{m}(y_{ij} - \mu_i - \nu_j)^2}{2}\right) \tag{3.13}$$

$$\sigma_2^2 | \Omega_{\sigma_2^2}, \underset{\sim}{y} \overset{iid}{\sim} \text{Gamma}\left(\frac{m + a}{2}, \frac{b + \sum_{j=1}^{m}\nu_j^2}{2}\right) \tag{3.14}$$

$$\sigma_3^2 | \Omega_{\sigma_3^2}, \underset{\sim}{y} \overset{iid}{\sim} \text{Gamma}\left(\frac{\ell + a}{2}, \frac{b + \sum_{i=1}^{\ell}\sum_{k=1}^{c} z_{ik}(\mu_i - \text{logit}(\theta_k^*))^2}{2}\right) \tag{3.15}$$

$$\nu_j | \Omega_{\nu}, \underset{\sim}{y} \overset{iid}{\sim} \text{Normal}\left(\beta\left[\frac{1}{\ell}\sum_{i=1}^{\ell}(\bar{y_{ij}} - \mu_i)\right], (1-\beta)\sigma_2^2\right) \tag{3.16}$$

where $\beta = \frac{\sigma_2^2}{\sigma_2^2 + \frac{\sigma_1^2}{\ell}}$, and $\bar{y_{ij}} = \frac{\sum_{i=1}^{\ell} y_{ij}}{\ell}$

$$z_i | \underset{\sim}{\Omega}_z, \underset{\sim}{y} \overset{iid}{\sim} \text{Multinomial}\,(1,\ \underset{\sim}{q_i}) \tag{3.17}$$

where $q_{ik} = \frac{\omega_k \frac{1}{\sqrt{2\pi\sigma_3^2}} e^{-\frac{1}{2\sigma_3^2}(\mu_i - \text{logit}(\theta_k^*))^2}}{\sum_{k=1}^{c}\omega_k \frac{1}{\sqrt{2\pi\sigma_3^2}} e^{-\frac{1}{2\sigma_3^2}(\mu_i - \text{logit}(\theta_k^*))^2}}, \quad k = 1, \ldots, c,\ i = 1, \ldots, \ell.$

The parameters $\underset{\sim}{\theta}$ and $\underset{\sim}{\omega}$ are very important to this research. Therefore, we place more emphasis on their derivation as stated below. The conditional posterior density for $\underset{\sim}{\theta}$ is given by

$$\pi(\underset{\sim}{\theta} | \Omega_{\underset{\sim}{\theta}}, \underset{\sim}{y}) \propto \prod_{k=1}^{c}\left[\frac{\omega_k^{\sum_{k=1}^{c} z_{ik}}}{\sqrt{2\pi\tilde{\sigma}_{3k}^2}} e^{-\frac{1}{2\tilde{\sigma}_{3k}^2}\sum_{k=1}^{c} z_{ik}(\mu_i - \log(\frac{\theta_k^*}{1-\theta_k^*}))^2}\right], -\infty < \theta_1 < \theta_2 < \ldots < \theta_c < -\infty, \tag{3.18}$$

and

$$\pi(\theta_k | \Omega_{\theta k}, \underset{\sim}{y}) \propto e^{-\frac{1}{2\sigma_3^2}\sum_{i=1}^{\ell}\left[z_{ik}\{\mu_i - \log(\frac{(\theta_{k-1}+\theta_k)/2}{1-(\theta_{k-1}+\theta_k)/2})\}^2 + z_{ik}\{\mu_i - \log(\frac{(\theta_k+\theta_{k+1})/2}{1-(\theta_k+\theta_{k+1})/2})\}^2\right]}. \tag{3.19}$$

The conditional posterior density for $\underset{\sim}{\omega}$ is given by

$$(\underset{\sim}{\omega}|\Omega_\omega, \underset{\sim}{y}) \propto \prod_{k=1}^{c} \omega_k^{z_{.k}} \tag{3.20}$$

where

$$(\underset{\sim}{\omega}|\Omega_\omega, \underset{\sim}{y}) \overset{iid}{\sim} \text{Dirichlet}\,(z_{.1}+1, \ldots, z_{.c}+1) \tag{3.21}$$

subject to $d \leq \omega_k \leq e$, $\sum_{k=1}^{c} \omega_k = 1$, $\omega_k \geq 0$.

To find the conditional posterior density of $(\omega_k|\underset{\sim}{\omega_k}, \Omega_{\omega_k}, \underset{\sim}{y})$, we begin with the introduction of a theorem,

<u>Theorem 1</u>:

$$\begin{aligned} \omega_k &\sim \text{Dirichlet}\,(\alpha_1, \ldots, \alpha_c) \\ \frac{\omega_k}{(1-\sum_{j=1}^{c-1}\omega_j)} \mid \underset{\sim}{\omega_k} &\sim \text{Beta}\,(\alpha_k, \alpha_c)\,, \quad \text{in } (0,1) \end{aligned}$$

$$k = 1, \ldots, c, \quad \textit{and } \omega_c = 1 - \sum_{j=1}^{c-1} \omega_j.$$

$\underset{\sim}{\omega_k} = (\omega_1, \omega_2, \omega_{k-1}, \omega_{k+1}, \ldots, \omega_c)$

$$d \leq \omega_k \leq e, \quad k = 1, \ldots, c$$

$$\frac{\omega_k}{1-\sum_{j=1}^{c-1}\omega_j} \mid \underset{\sim}{\omega_k} \sim \text{Beta}\,(\alpha_k, \alpha_c)\,, \ \text{in } (A, B)$$

where

$$\begin{aligned} A &= \max\left[0,\ \frac{d}{1-\sum_{j=1}^{c-1}\omega_j},\ \frac{e}{1-\sum_{j=1}^{c-1}\omega_j}\right] \\ B &= \min\left[1,\ \frac{e}{1-\sum_{j=1}^{c-1}\omega_j},\ \frac{d}{1-\sum_{j=1}^{c-1}\omega_j}\right] \end{aligned}$$

For example, if k=2,

$$\omega_j \sim \text{Beta}(z_{.1}, z_{.2}),\ d < \omega_1 < e, \quad \text{and} \ \ \omega_2 = 1 - \omega_1.$$

We choose $c \cdot d = 0.6$ and $c \cdot e = 1.2$, where c is the number of clusters, so that, for c = 2, ω will be in the interval (0.3, 0.6); for c = 3, ω will be in the interval (0.2,0.4) and so on.

By the application of Devroye's method [3],we draw a sample from Beta(α, β) in (a, b) and take

$$X = F^{-1}\Big[UF(a) + (1-U)F(b)\Big] \tag{3.22}$$

where, $U \sim$ uniform$[0,1]$, $F(a)$ is the Beta cdf, and $F^{-1}(.)$ the inverse cdf.

Chapter 4

Analysis of the Latent Class Model

In this chapter, we discuss the goodness of fit of the latent class model (LCM), and inference on our data set using the LCM.

4.1 Cross-validation Analysis

The assessment of this model is by a Bayesian cross-validation analysis to obtain deleted residuals on the observed values, y_{ij}.
From our model

$$y_{ij}|\mu_i, \nu_j, \sigma_1^2 \overset{iid}{\sim} \text{Normal}\left(\mu_i + \nu_j, \sigma_1^2\right), \quad i = 1, \ldots, \ell, \quad j = 1, \ldots, m.$$

Let $y_{(ij)}$ denote the vector of all observations excluding the $(ij)^{th}$ observation y_{ij}.
Then the $(ij)^{th}$ deleted residual is given by

$$DRES_{ij} = \{y_{ij} - E\left(y_{ij}|y_{(ij)}\right)\}/STD\left(y_{ij}|y_{(ij)}\right), \ i = 1, \ldots, \ell, \ j = 1, \ldots, m,$$

where

$$E\left(y_{ij}|y_{(ij)}\right) = E_{(\Omega|y_{(ij)})}\{E(y_{ij}|\Omega)\}$$

and

$$\mathrm{Var}\left(y_{ij}|y_{(ij)}\right) = E_{(\Omega|y_{(ij)})}\{\mathrm{Var}\left(y_{ij}|\Omega\right)\} + \mathrm{Var}_{(\Omega|y_{(ij)})}\{E\left(y_{ij}|\Omega\right)\}$$

We can estimate $E\left(y_{ij}|y_{(ij)}\right)$ by

$$E\left(\widehat{y_{ij}|y_{(ij)}}\right) = \sum_{h=1}^{M} w_{ij}^{(h)}(\mu_i^{(h)} + \nu_j^{(h)})$$

and Var $\left(y_{ij}|y_{(ij)}\right)$ by

$$\mathrm{Var}\left(\widehat{y_{ij}|y_{(ij)}}\right) = \sum_{h=1}^{M} w_{ij}^{(h)}\sigma_1^{2(h)} + \sum_{h=1}^{M} w_{ij}^{(h)}(t_{ij}^{(h)} - \bar{t}_{ij})^2$$

where

$$t_{ij}^{(h)} = \mu_i^{(h)} + \nu_j^{(h)}, \text{ and } \bar{t}_{ij} = \sum_{h=1}^{M} w_{ij}^{(h)} t_{ij}^{(h)}.$$

These values are obtained by performing analysis using the output of the original Gibbs sampler.

From the Gibbs sampler we obtain $\Omega^{(h)}$, h = 1,. . . ,M = 1000, where M is the sample size.

Our weights can be obtained as

$$w_{ij}^{(h)} = \frac{1/f(y_{ij}|\ \Omega^{(h)})}{1/\sum_{h=1}^{M} f(y_{ij}|\ \Omega^{(h)})}, \quad i = 1, \ldots, \ell, \quad j = 1, \ldots, m, \quad h = 1, \ldots, M$$

where

$$f(y_{ij}|\Omega^{(h)}) = \frac{1}{\sqrt{2\pi\sigma_1^{2(h)}}} e^{-\frac{1}{2\sigma_1^{2(h)}}(y_{ij}-(\mu_i^{(h)}+\nu_j^{(h)}))^2}.$$

Figure 1 shows a scatter plot of deleted residuals versus predicted values for logarithm of the original y_{ij}. Here most of the residuals fall in the (-2, +2) band with a negative slope, and a correlation of -0.192 between the deleted residuals and the predicted values.

Figure 2 is a box plot of residuals for each company. We observe a few outliers, which in fact, are expected, because in a universe of companies, a particular company, say, Amazon.com may decide on spending more, sometimes above the industry average, thereby creating an outlier.

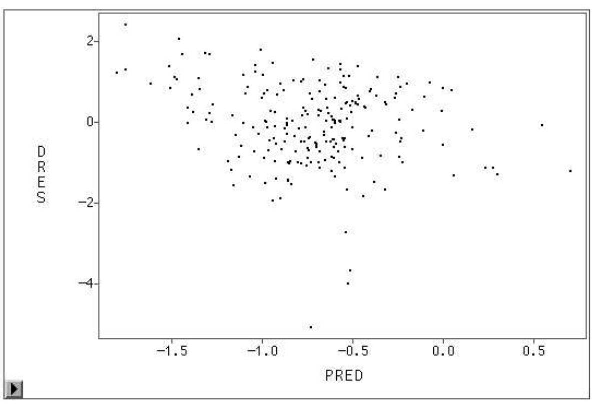
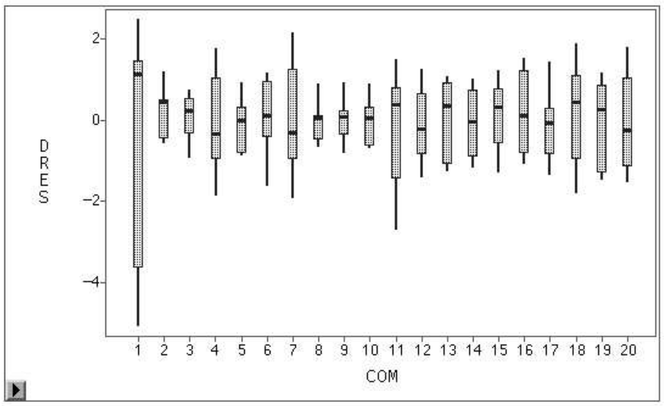


Figure 4.1: Scatter plot of the deleted residual (DRES) versus predicted value (PRED) from the cross validation. Using the logarithm of the transformation.

Figure 4.2: Box plot of the deleted residual (DRES) versus predicted value (PRED) from the cross validation. Using the logarithm of the transformation.



4.2 Sensitivity to Transformation

Our objective is to obtain an almost zero correlation between the deleted residuals (DRES) and the predicted values (PRED), and an almost zero slope, which indicates a good fit. With this in mind, we try out different transformations until we find one that best fits these specifications. As observed from the previous section, y_{ij} demonstrate some skewness, rescaling y_{ij} by taking the logarithm gave some symmetric spread of values with a correlation of -0.192 between the deleted residuals (DRES1) and the predicted values (PRED1). For now, instead of the logarithm we rescale y_{ij} by taking

		PRED1	PRED2	PRED3
Logarithm	DRES1	-0.192		
Square root	DRES2		-0.062	
Cube root	DRES3			-0.057

Table 4.1: Correlations between the DRES and PRED for the three transformations

the square root, this gives a better symmetric spread of values, and a correlation of -0.062 between the deleted residuals (DRES2) and the predicted values (PRED2). Finally, rescaling y_{ij} by taking the cube root, also gives a better symmetric spread of values, and a correlation of -0.056 between the deleted residuals (DRES3) and the predicted values (PRED3) as shown in Table 4.2.

It is from this observation that we perform a sensitivity analysis to see how our results, the overall classification and ranking procedure, vary with different transformations.

Figure 3 represents the scatter plots for the three different transformations. In fact, all three plots indicate a reasonable fit, however, there is a slight improvement in the second and third plots, obtained when we used the square and cube root of the data set.

Figure 4 represents the box plots for the three different transformations. As in the previous figure, all three plots have their observations close to the zero reference line, with only a few outliers in the case of the square and the cube root.

From Tables 4.3, we observe the sensitivity of the clusters and rankings to the misspecification of the transformation. It is observed that when the

	Company	LOG Cluster	LOG Rank	Sq. Root Cluster	Sq. Root Rank	Cube Root Cluster	Cube Root Rank
13	FTD.com	4	1	3	1	3	1
16	NetB@nk	3	2	3	2	3	2
18	Peapod Inc	3	4	3	3	3	3
11	E*TRADE Group Inc	3	8	3	4	3	4
6	CDnow Inc	3	5	3	5	3	6
1	Amazon.com	1	18	3	6	3	9
2	Autobytel.com Inc	3	3	3	7	3	5
15	iOwn	3	7	3	8	3	7
3	Autoweb.com	3	6	3	9	3	8
4	Bolt Inc	2	14	2	10	2	10
20	Webvan Group Inc	1	17	2	11	2	11
19	PlanetRx.com Inc	2	16	2	12	2	12
17	Nextcard Inc	2	13	2	13	2	13
14	Furniture.com Inc	2	9	2	14	2	14
12	Fogdog Inc	2	12	2	15	2	15
10	eToys Inc	1	20	2	16	2	16
9	E-Loan Inc	2	10	2	17	2	17
8	drugstore.com Inc	2	11	2	18	2	18
7	800.com Inc	2	15	2	19	2	19
5	CarsDirect.com Inc	1	19	2	20	2	20

Table 4.2: Sensitivity of the clusters and rankings to misspecification of the transformation (first set: logarithm, second set: square root, third set: cube root).

logarithm of the data set is used, there exist at least one company in every cluster. Where as, when the observed data is transformed using the square or cube root of the observed data, the procedure classifies the individual companies into only the middle two clusters (clusters 2 and 3). Moreover, the rankings almost always stays the same under both the square and the cube root transformations except a slight change in ranking occurs in the third cluster, between the fifth and the ninth ranks. Although, it is frequently possible to overlook the distinction between the different transformations in some instances such as dealing with daily returns of stock prices the difference cannot be ignored.

4.3 Results and Implications

We "fire up" the Gibbs sampler by drawing up the z_i's, ω's, ν_j's, μ_i's, θ's, and the σ^2's.

To compare the fit of the predictions, $\hat{y_{ij}}$, to the observed outcomes y_{ij}, we used the following statistics: Posterior mean, posterior standard deviation, numerical standard error (NSE), and confidence intervals (C.I 75% and C.I 25%).

For each analysis, we drew 75,000 iterates from the Gibbs sampler. Convergence was deemed to have occurred within the first 1,000, since our algorithm enables us to 'burn' in 5,000, we picked every seventy to remove the autocorrelation among the iterates. This rule was obtained by trial and error. There is nothing to infer from the confidence intervals (C.I 75% and C.I 25%) since they tend to overlap one another.

Appendix A presents the classification and ranking of the twenty eCommerce companies under the four perspectives. Table 4.4 represents the overall performance of the classification and ranking of the twenty B2C companies. Each of the classification tables has two rows for each company. The first row represent the estimation of probabilities that the i^{th} company belongs to the k^{th} cluster, $k = 1, \ldots, c = 4$, and $i = 1, \ldots, 20$, and the second row, the numerical standard errors (NSE). We observed that the NSE were very small, which ensured steady probabilities in the different clusters.

Note: Clusters to the right are better, i.e., $C4$ is considered the best cluster.

Table 4.3: Overall performance

	Company	C1	C2	C3	C4
1	Amazon.com	0.000	0.358	0.642	0.000
		0.000	0.008	0.008	0.000
2	Autobyel.com Inc	0.000	0.360	0.640	0.000
		0.000	0.008	0.008	0.000
3	Autoweb.com	0.000	0.468	0.532	0.000
		0.000	0.008	0.008	0.000
4	Bolt Inc.	0.000	0.613	0.387	0.000
		0.000	0.008	0.008	0.000
5	CarsDirect.com Inc.	0.000	0.873	0.126	0.000
		0.000	0.005	0.005	0.000
6	Cdnow.com Inc	0.000	0.355	0.644	0.000
		0.000	0.008	0.008	0.000
7	800.com Inc	0.000	0.665	0.3351	0.000
		0.000	0.008	0.008	0.000
8	Drugstore.com Inc	0.000	0.697	0.303	0.000
		0.000	0.008	0.008	0.000
9	E-Loan Inc	0.000	0.714	0.286	0.000
		0.000	0.008	0.008	0.000
10	eToys Inc	0.000	0.870	0.130	0.000
		0.000	0.006	0.006	0.000

Table 4.4: Overall Performance

	Company	C1	C2	C3	C4
11	E*Trade Group Inc	0.000	0.282	0.718	0.000
		0.000	0.008	0.008	0.000
12	Fogdog Inc	0.000	0.658	0.342	0.000
		0.000	0.008	0.008	0.000
13	FTD.com	0.000	0.059	0.898	0.043
		0.000	0.004	0.005	0.003
14	Furniture.com Inc	0.000	0.580	0.420	0.000
		0.000	0.008	0.008	0.000
15	iOwn	0.000	0.422	0.578	0.000
		0.000	0.008	0.008	0.000
16	NetB@nk	0.000	0.122	0.877	0.000
		0.000	0.006	0.006	0.000
17	NextCard Inc	0.000	0.727	0.273	0.000
		0.000	0.008	0.008	0.000
18	Peapod Inc	0.000	0.211	0.789	0.000
		0.000	0.007	0.007	0.000
19	PlanetRx.com Inc	0.000	0.749	0.251	0.000
		0.000	0.007	0.007	0.000
20	Webvan Group Inc	0.000	0.811	0.189	0.000
		0.000	0.007	0.007	0.000

Table 4.5: Overall performance-classification into clusters and Ranks

	Company	Cluster	Rank
1	Amazon.com	3	6
2	Autobytel.com Inc	3	7
3	Autoweb.com	3	9
4	Bolt Inc	2	10
5	CarsDirect.com Inc	2	20
6	CDnow Inc	3	5
7	800.com Inc	2	19
8	drugstore.com Inc	2	18
9	E-Loan Inc	2	17
10	eToys Inc	2	16
11	E*TRADE Group Inc	3	4
12	Fogdog Inc	2	15
13	FTD.com	3	1
14	Furniture.com Inc	2	14
15	iOwn	3	8
16	NetB@nk	3	2
17	Nextcard Inc	2	13
18	Peapod Inc	3	3
19	PlanetRx.com Inc	2	12
20	Webvan Group Inc	2	11

4.4 Simulation Study

In this section, we perform a study to assess how the residual plots should look like, and the performance of the classification of the LCM.

We obtain the σ^2's, θ's, and ω's from our original model in chapter 3. We recall that the LCM is described as:

$$y_{ij}|\mu_i, \nu_j, \sigma_1^2 \overset{iid}{\sim} \text{Normal}\left(\mu_i + \nu_j, \sigma_1^2\right), \ i = 1, \ldots, \ell = 20, \ j = 1, \ldots, m = 11.$$

As before, we assume μ_i to be independent, and generate the data by drawing

$$\left(\mu_i \mid \underset{\sim}{\theta}^*,, \ \sigma_3^2\right) = \begin{cases} \text{Normal}\left(\log(\frac{\theta_1/2}{1-\theta_1/2}), \sigma_3^2\right), & \text{with probability} \quad \omega_1 \\ \text{Normal}\left(\log(\frac{(\theta_1+\theta_2)/2}{1-(\theta_1+\theta_2)/2}), \sigma_3^2\right), & \text{with probability} \quad \omega_2 \\ \text{Normal}\left(\log(\frac{(\theta_2+\theta_3)/2}{1-(\theta_2+\theta_3)/2}), \sigma_3^2\right), & \text{with probability} \quad \omega_3 \\ \text{Normal}\left(\log(\frac{(1+\theta_3)/2}{1-(1+\theta_3)/2}), \sigma_3^2\right), & \text{with probability} \quad \omega_4 \end{cases}$$

where, $0 \leq \theta_1 \leq \theta_2 \leq \theta_3 \leq 1$ is the order statistics, and

$$\nu_j \quad \sim \quad \text{Normal}\left(0, \sigma_2^2\right).$$

Our simulation experiment, in which we generate 1000 data sets, is as follows:

1. i Repeat the above process a thousand times to generate $y_{ij}^{(f)}$, $f = 1, \ldots, 1000$.

   ii Note the cluster in which each company belongs.

2. i Fit the Latent class model.

   ii Count how many times each company is categorized into a particular cluster.

Table 4.6: Simulation study showing the classification efficiency of the LCM

	Estimates (%)			
SIM	**1**	**2**	**3**	**4**
1	**85.34**	14.32	0.33	0.00
2	9.16	**56.75**	33.54	0.55
3	3.06	41.68	**52.70**	2.56
4	0.00	0.07	10.80	**89.13**

Table 4.7: Statistic for table of SIM by Estimates

Statistic	DF	Value	Prob
Chi-Square	9	29650.75	< 0.0001
Likelihood Ratio Chi-Square	9	27855.74	< 0.0001
Sample size = 20000			

From Table 4.7 the simulated - estimated values of 85.34% and 89.13% for the worst and best clusters, respectively, and 56.75% for cluster 2 and 52.70% for cluster 3 shows that the latent class model is very efficient in the classification of companies in the extreme clusters (worst and best).

From Table 4.8 we observe small p-values (0.0001), this shows a strong association between the simulated and estimated values.

From our simulation the modal class interval of the estimated correlations between the deleted residuals and the predicted values is (-0.07, -0.04). From section 4.2 we obtained a correlation of -0.06, which is consistent with the simulated result. We conclude that the Latent class model shows a good fit.

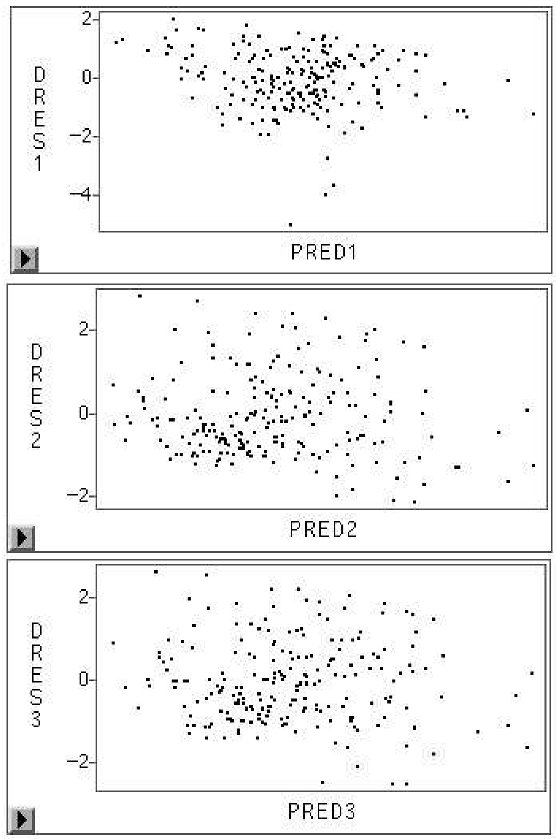


Figure 4.3: Scatter plot of the deleted residuals (DRES) versus predicted values (PRED) (top: logarithm, middle: square root, bottom: cube root).

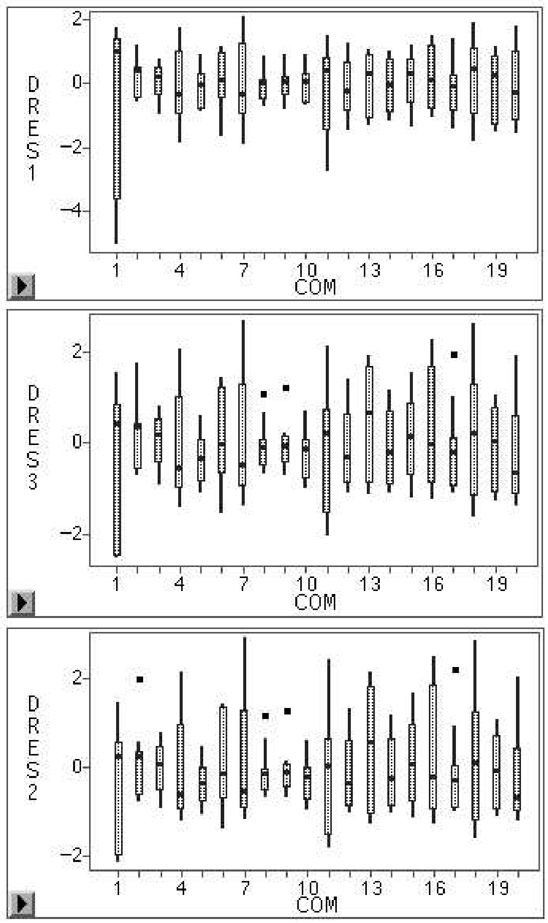


Figure 4.4: Box plot of the deleted residuals (DRES) by company (top: logarithm, middle: square root , bottom: cube root).

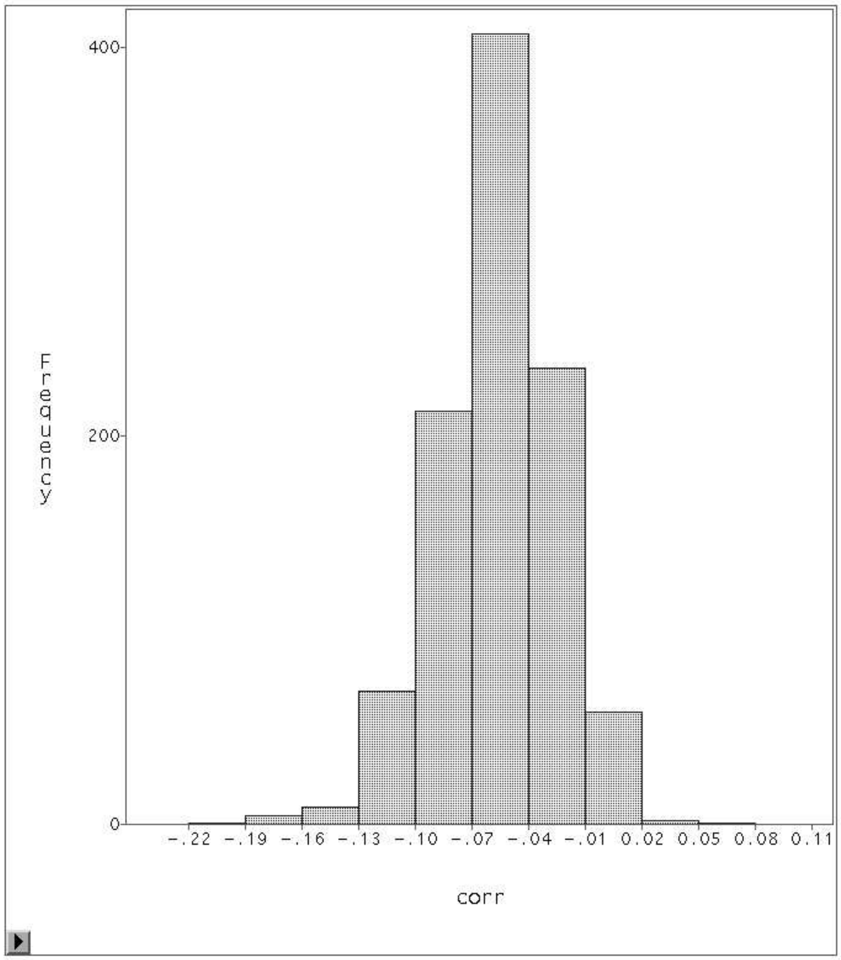


Figure 4.5: Histogram of the correlation between deleted residuals (DRES) versus predicted values (PRED) for the 1000 simulated experiments.

Chapter 5

Conclusion

Ranks are often used as a summary measure of a company's relative performance. We observe that the ranks of the twenty eCommerce companies vary across different perspectives. There is also no evident trend between the overall performance and an individual perspective. We have shown that measuring the performance of an organization by the old measures of success like, return-on-assets (ROA), return-on-equity (ROE), and return-on-investment (ROI) provides a partial information about an organization. We have also shown that the qualitative BSC approach can be quantified to evaluate performance using the latent class model. Therefore, the BSC approach of analyzing an organization's overall performance by critical indicators is an excellent methodology.

5.1 Discussion of the classification and ranking process

An eCommerce site's greatest assets are its customers, which is why some of the most valued and sophisticated metrics used today evolve around customer behavior [1]. For example, Amazon.com, one of the highest rated eCommerce companies with a large customer base ranked second in the customer perspective. Its weakest performance, in the financial perspective was a result of a comparatively huge loss at the time of the publication of the eCommerce Almanac. Overall, Amazon.com ranked sixth.

The resulting classification and ranking procedure shows that a company's overall performance does not depend on how well it performed in a particular perspective, but rather how it ranked relative to the others in all four perspectives. There is no evident trend between the overall performance and an individual perspective. For example, the overall best company was FTD.com, its performance across the critical perspectives was consistent except for the customer perspective where it ranked fifteenth. A similar conclusion can be drawn about NetB@nk and Peapod Inc., sharing the second and third positions respectively. It is observed that a weak performance in the financial and customer perspectives for NetB@nk, and a very weak performance in the learning and growth perspectives for Peapod Inc. did less harm to their overall performance relative to FTD.com. From this observation, we have at least shown how the BSC approach analyzes an organization's overall performance. And can conclude that the BSC methodology, based on the idea that assessing performance through for example, financial returns only provides partial information about the success of an organization.

We also observed that the effectiveness of the LCM will very much depend on the number of measures across each perspective. In other words, for a better classification process the analyst will have to generate more measures to obtain an efficient assessment of the set of companies to be evaluated.

It is from this observation that we propose a further study, known as *"small area estimation"* to study limited data or measures in any of the four perspectives.

Our methodology, although applicable to particular sectors of an industry, applies generally to all companies. We have at least shown how to work with a multivariate population, how to estimate the parameters of the selected population, and the innovative introduction of the latent variables, which resulted in the simplified classification of the twenty B2C eCommerce companies into relative "winners" and "losers".

The resulting work offers a practical tool with the ability to identify profitable investment opportunities (buy and sell decisions) for financial managers andanalyst.

Note: The ranks are represented in ascending order, where lowest number (1) denotes the strongest rank, and largest number (20) denotes the weakest rank.

		Overall	Financial	Customer	Internal	L & Growth
	Company	*Rank*	*Rank*	*Rank*	*Rank*	*Rank*
13	FTD.com	1	1	15	1	2
16	NetB@nk	2	11	18	2	1
18	Peapod Inc	3	7	3	3	19
11	E*TRADE Group Inc	4	12	1	7	5
6	CDnow Inc	5	15	5	4	3
1	Amazon.com	6	18	2	8	7
2	Autobytel.com Inc	7	9	8	5	4
15	iOwn	8	4	7	15	13
3	Autoweb.com	9	6	12	6	8
4	Bolt Inc	10	2	20	11	9
20	Webvan Group Inc	11	20	4	12	20
19	PlanetRx.com Inc	12	10	10	13	11
17	Nextcard Inc	13	14	13	14	6
14	Furniture.com Inc	14	5	11	16	12
12	Fogdog Inc	15	8	16	9	17
10	eToys Inc	16	19	17	17	18
9	E-Loan Inc	17	17	6	18	14
8	drugstore.com Inc	18	16	9	19	16
7	800.com Inc	19	3	14	10	15
5	CarsDirect.com Inc	20	13	19	20	10

Table 5.1: Final ranking of the twenty B2C eCommerce companies by the square root transformation

Table 5.2: Cluster 3 - "Best" companies from the classification process

	Company	Rank
13	FTD.com	1
16	NetB@nk	2
18	Peapod Inc	3
11	E*TRADE Group Inc	4
6	CDnow Inc	5
1	Amazon.com	6
2	Autobytel.com Inc	7
15	iOwn	8
3	Autoweb.com	9

Table 5.3: Cluster 2 - "Worst" companies from the classification process

	Company	Rank
4	Bolt Inc	1
20	Webvan Group Inc	2
19	PlanetRx.com Inc	3
17	Nextcard Inc	4
14	Furniture.com Inc	5
12	Fogdog Inc	6
10	eToys Inc	7
9	E-Loan Inc	8
8	drugstore.com Inc	9
7	800.com Inc	10
5	CarsDirect.com Inc	11

Chapter 6

Appendix - Classification and ranking tables under the four perspectives

The tables below represent the classification and ranking of the twenty eCommerce companies under the four perspectives.

<u>Note:</u>

- Clusters to the right are better, i.e., $C4$ is considered the best cluster.

- The ranks are represented in ascending order, where lowest number (1) denotes the strongest rank, and largest number (20) denotes the weakest rank.

Table 6.1: Financial Perspective

	Company	C1	C2	C3	C4
1	Amazon.com	0.000	0.776	0.224	0.000
		0.000	0.005	0.005	0.000
2	Autobyel.com Inc	0.000	0.490	0.510	0.000
		0.000	0.006	0.006	0.000
3	Autoweb.com	0.000	0.479	0.521	0.000
		0.000	0.006	0.006	0.000
4	Bolt Inc.	0.000	0.147	0.850	0.000
		0.000	0.004	0.004	0.000
5	CarsDirect.com Inc.	0.000	0.883	0.117	0.000
		0.000	0.004	0.004	0.000
6	Cdnow.com Inc	0.000	0.680	0.320	0.000
		0.000	0.006	0.006	0.000
7	800.com Inc	0.000	0.206	0.793	0.002
		0.000	0.005	0.005	0.000
8	Drugstore.com Inc	0.000	0.796	0.204	0.000
		0.000	0.005	0.005	0.000
9	E-Loan Inc	0.000	0.764	0.236	0.000
		0.000	0.005	0.005	0.000
10	eToys Inc	0.000	0.898	0.102	0.000
		0.000	0.004	0.004	0.000

Table 6.2: Financial Perspective

	Company	C1	C2	C3	C4
11	E*Trade Group Inc	0.000	0.609	0.391	0.000
		0.000	0.006	0.006	0.002
12	Fogdog Inc	0.000	0.652	0.348	0.000
		0.000	0.006	0.006	0.000
13	FTD.com	0.000	0.054	0.886	0.059
		0.000	0.003	0.004	0.003
14	Furniture.com Inc	0.000	0.408	0.592	0.000
		0.000	0.006	0.006	0.000
15	iOwn	0.000	0.139	0.858	0.000
		0.000	0.004	0.004	0.000
16	NetB@nk	0.000	0.628	0.372	0.000
		0.000	0.006	0.006	0.000
17	NextCard Inc	0.000	0.887	0.113	0.000
		0.000	0.004	0.004	0.000
18	Peapod Inc	0.000	0.058	0.905	0.037
		0.000	0.003	0.004	0.002
19	PlanetRx.com Inc	0.000	0.712	0.288	0.000
		0.000	0.005	0.005	0.000
20	Webvan Group Inc	0.000	0.828	0.172	0.000
		0.000	0.005	0.005	0.000

Table 6.3: Financial perspective-classification into clusters and Ranks

	Company	Cluster	Rank
1	Amazon.com	2	1
2	Autobytel.com Inc	2	3
3	Autoweb.com	3	7
4	Bolt Inc	3	4
5	CarsDirect.com Inc	2	13
6	CDnow Inc	2	11
7	800.com Inc	3	5
8	drugstore.com Inc	2	6
9	E-Loan Inc	2	7
10	eToys Inc	2	8
11	E*TRADE Group Inc	2	5
12	Fogdog Inc	2	12
13	FTD.com	3	1
14	Furniture.com Inc	3	6
15	iOwn	3	3
16	NetB@nk	2	10
17	Nextcard Inc	2	4
18	Peapod Inc	3	2
19	PlanetRx.com Inc	2	9
20	Webvan Group Inc	2	2

Table 6.4: Customer Perspective

	Company	C1	C2	C3	C4
1	Amazon.com	0.000	0.009	0.028	0.963
		0.000	0.003	0.052	0.006
2	Autobyel.com Inc	0.005	0.417	0.577	0.001
		0.002	0.016	0.016	0.001
3	Autoweb.com	0.008	0.419	0.572	0.001
		0.003	0.016	0.016	0.001
4	Bolt Inc.	0.037	0.499	0.133	0.000
		0.015	0.016	0.011	0.000
5	CarsDirect.com Inc.	0.027	0.597	0.375	0.001
		0.005	0.016	0.015	0.001
6	Cdnow.com Inc	0.000	0.219	0.701	0.080
		0.000	0.013	0.015	0.009
7	800.com Inc	0.556	0.372	0.072	0.000
		0.016	0.015	0.008	0.000
8	Drugstore.com Inc	0.009	0.413	0.575	0.003
		0.003	0.016	0.016	0.002
9	E-Loan Inc	0.020	0.523	0.456	0.001
		0.004	0.016	0.016	0.001
10	eToys Inc	0.014	0.496	0.488	0.002
		0.004	0.016	0.012	0.001

Table 6.5: Customer Perspective

	Company	C1	C2	C3	C4
11	E*Trade Group Inc	0.000	0.028	0.110	0.862
		0.000	0.005	0.010	0.011
12	Fogdog Inc	0.166	0.622	0.212	0.000
		0.012	0.015	0.013	0.000
13	FTD.com	0.003	0.385	0.609	0.003
		0.002	0.015	0.015	0.002
14	Furniture.com Inc	0.076	0.589	0.335	0.000
		0.008	0.016	0.015	0.000
15	iOwn	0.030	0.578	0.390	0.002
		0.005	0.016	0.015	0.014
16	NetB@nk	0.002	0.331	0.646	0.021
		0.001	0.015	0.015	0.005
17	NextCard Inc	0.017	0.465	0.517	0.001
		0.004	0.016	0.016	0.001
18	Peapod Inc	0.002	0.328	0.637	0.033
		0.001	0.015	0.015	0.006
19	PlanetRx.com Inc	0.120	0.617	0.263	0.000
		0.010	0.015	0.014	0.000
20	Webvan Group Inc	0.041	0.614	0.345	0.000
		0.006	0.015	0.015	0.000

Table 6.6: Customer perspective-classification into clusters and Ranks

	Company	Cluster	Rank
1	Amazon.com	4	1
2	Autobytel.com Inc	2	1
3	Autoweb.com	2	7
4	Bolt Inc	2	12
5	CarsDirect.com Inc	2	11
6	CDnow Inc	3	3
7	800.com Inc	2	10
8	drugstore.com Inc	2	9
9	E-Loan Inc	2	8
10	eToys Inc	2	13
11	E*TRADE Group Inc	3	1
12	Fogdog Inc	2	6
13	FTD.com	3	6
14	Furniture.com Inc	2	5
15	iOwn	2	4
16	NetB@nk	3	4
17	Nextcard Inc	2	3
18	Peapod Inc	3	2
19	PlanetRx.com Inc	2	2
20	Webvan Group Inc	3	5

Table 6.7: Internal Business Processes

	Company	C1	C2	C3	C4
1	Amazon.com	0.979	0.021	0.000	0.000
		0.005	0.005	0.000	0.000
2	Autobyel.com Inc	0.001	0.119	0.788	0.0920
		0.001	0.010	0.013	0.009
3	Autoweb.com	0.009	0.137	0.786	0.068
		0.003	0.011	0.013	0.008
4	Bolt Inc.	0.041	0.618	0.341	0.000
		0.006	0.015	0.015	0.000
5	CarsDirect.com Inc.	0.671	0.298	0.031	0.000
		0.015	0.015	0.006	0.000
6	Cdnow.com Inc	0.010	0.167	0.777	0.046
		0.003	0.012	0.013	0.007
7	800.com Inc	0.021	0.526	0.448	0.005
		0.005	0.016	0.016	0.002
8	Drugstore.com Inc	0.025	0.546	0.423	0.006
		0.005	0.016	0.016	0.002
9	E-Loan Inc	0.026	0.571	0.402	0.001
		0.005	0.016	0.016	0.001
10	eToys Inc	0.688	0.286	0.026	0.000
		0.015	0.014	0.005	0.000

Table 6.8: Internal Business Processes

	Company	C1	C2	C3	C4
11	E*Trade Group Inc	0.062	0. 717	0.218	0.003
		0.008	0.014	0.013	0.002
12	Fogdog Inc	0.014	0.268	0.703	0.015
		0.004	0.014	0.014	0.004
13	FTD.com	0.001	0.005	0.011	0.983
		0.001	0.002	0.003	0.004
14	Furniture.com Inc	0.015	0.433	0.546	0.006
		0.004	0.016	0.016	0.002
15	iOwn	0.026	0.571	0.398	0.005
		0.005	0.016	0.016	0.002
16	NetB@nk	0.000	0.004	0.017	0.979
		0.000	0.002	0.004	0.005
17	NextCard Inc	0.025	0.516	0.455	0.004
		0.005	0.016	0.016	0.002
18	Peapod Inc	0.188	0.684	0.126	0.002
		0.0124	0.015	0.011	0.001
19	PlanetRx.com Inc	0.281	0.634	0.085	0.000
		0.014	0.015	0.009	0.000
20	Webvan Group Inc	0.919	0.070	0.011	0.000
		0.009	0.008	0.003	0.000

Table 6.9: Internal bus. processes-classification into clusters and Ranks

	Company	Cluster	Rank
1	Amazon.com	2	1
2	Autobytel.com Inc	3	1
3	Autoweb.com	3	2
4	Bolt Inc	3	6
5	CarsDirect.com Inc	2	6
6	CDnow Inc	3	3
7	800.com Inc	3	5
8	drugstore.com Inc	3	11
9	E-Loan Inc	3	12
10	eToys Inc	2	5
11	E*TRADE Group Inc	3	8
12	Fogdog Inc	3	4
13	FTD.com	4	1
14	Furniture.com Inc	3	7
15	iOwn	3	9
16	NetB@nk	4	2
17	Nextcard Inc	3	10
18	Peapod Inc	2	4
19	PlanetRx.com Inc	2	3
20	Webvan Group Inc	3	2

Table 6.10: Learning and Growth Perspective

	Company	C1	C2	C3	C4
1	Amazon.com	0.006	0.193	0.440	0.361
		0.002	0.013	0.016	0.015
2	Autobyel.com Inc	0.006	0.247	0.475	0.272
		0.002	0.014	0.016	0.014
3	Autoweb.com	0.014	0.274	0.508	0.204
		0.004	0.014	0.016	0.013
4	Bolt Inc.	0.236	0.501	0.262	0.001
		0.013	0.016	0.014	0.001
5	CarsDirect.com Inc.	0.284	0.513	0.203	0.000
		0.014	0.016	0.013	0.000
6	Cdnow.com Inc	0.002	0.155	0.338	0.505
		0.001	0.011	0.015	0.016
7	800.com Inc	0.234	0.497	0.266	0.003
		0.013	0.016	0.014	0.002
8	Drugstore.com Inc	0.050	0.378	0.538	0.034
		0.007	0.015	0.016	0.006
9	E-Loan Inc	0.074	0.426	0.486	0.014
		0.008	0.016	0.016	0.004
10	eToys Inc	0.173	0.501	0.321	0.005
		0.012	0.016	0.015	0.002

Table 6.11: Learning and Growth Perspective

	Company	C1	C2	C3	C4
11	E*Trade Group Inc	0.002	0.203	0.381	0.414
		0.001	0.013	0.015	0.016
12	Fogdog Inc	0.095	0.477	0.422	0.006
		0.009	0.016	0.016	0.002
13	FTD.com	0.000	0.063	0.167	0.770
		0.000	0.008	0.012	0.013
14	Furniture.com Inc	0.083	0.496	0.417	0.004
		0.009	0.016	0.016	0.002
15	iOwn	0.101	0.447	0.440	0.012
		0.010	0.016	0.016	0.003
16	NetB@nk	0.001	0.047	0.161	0.791
		0.001	0.007	0.012	0.013
17	NextCard Inc	0.055	0.377	0.538	0.030
		0.007	0.015	0.016	0.005
18	Peapod Inc	0.066	0.402	0.509	0.023
		0.008	0.016	0.016	0.005
19	PlanetRx.com Inc	0.245	0.517	0.234	0.004
		0.014	0.016	0.013	0.002
20	Webvan Group Inc	0.427	0.466	0.105	0.002
		0.016	0.016	0.010	0.001

Table 6.12: Learning and Growth-classification into clusters and Ranks

	Company	Cluster	Rank
1	Amazon.com	3	7
2	Autobytel.com Inc	3	4
3	Autoweb.com	3	8
4	Bolt Inc	2	1
5	CarsDirect.com Inc	2	2
6	CDnow Inc	3	3
7	800.com Inc	2	7
8	drugstore.com Inc	2	8
9	E-Loan Inc	2	6
10	eToys Inc	2	10
11	E*TRADE Group Inc	3	5
12	Fogdog Inc	2	9
13	FTD.com	3	2
14	Furniture.com Inc	2	4
15	iOwn	2	5
16	NetB@nk	3	1
17	Nextcard Inc	3	6
18	Peapod Inc	2	11
19	PlanetRx.com Inc	2	3
20	Webvan Group Inc	2	12

Bibliography

[1] John Berry. Internetweek; "the world according to e-biz metrics, 1996.

[2] Homepage: CEOReview.com. Balanced scorecard: Translating strategy into action, 2002.

[3] L. Devroye. In *Non-uniform Random Variate Generation*. New York - Springer-Verlag.

[4] H. Goldstein and Spielgelhalter. League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society*, A, 1996.

[5] S. S. Gupta and S. Panchapakesan. *Multiple Decision Procedures*. John Wiley & Sons, Inc., New York, 1979.

[6] J. Low, T. Siesfeld, and D. Larcker. Capital thinking. *Forbes*, 199(9), 1999.

[7] C. N. Morris and C. L. Christiansen. Hierarchical models for ranking and for identifying extremes, with applications. In J.M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics V*. Oxford University Press, 227-296, 1996.

[8] J. Roos. Intellectual capital: what you can measure, you can manage.

[9] S. Schartz. Transcending financial measures. *Insurance and Technology*, 22(4), 1997.

[10] Homepage: SearchCIO.com. Definitions: Balanced scorecard methodology, April 2003.