The Determinants of Price Dispersion

in Cosmetic E-Commerce in China

By

Nan Li

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

Master of Arts

in

Economics

August, 2010

Nashville, Tennessee

Approved:

Ana R Vides De Andrade

Suhas L Ketkar

ABSTRACT

There are large numbers of potential customers of cosmetic e-commerce in China. From Jan to Aug in 2009, about 61% of Internet users have browsed shopping websites and about 18% of users have purchased a product online. In addition, more than 30 million which is about 8.3% of the total Netizen are cosmetic users. But there is a large price gap among the e-tailors. It might because of brand, product and e-tailor's characteristics. This thesis uses the HLM to test how the online cosmetic price is determined by brand, product and e-tailor's characteristics. It includes price data for 44 different product items from the 13 active Guangdong e-tailor web sites with different brands, product and e-tailor characteristics. The data consist of 269 observations collected during November, 2009. In order to capture the three-step process consumer use when selecting which e-tailor to buy from, it uses the HLM model.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 Background

The cosmetic e-commerce becomes more and more important in China. According to the data of State Statistics Bureau, during the first half of 2009, 17.2% growth has taken place in cosmetics while the total retail sales of consumer goods are 5871.1 billion yuan.

With 100% tariff on the imported cosmetic, the prices of foreign cosmetic in China are much higher than those in US. Chinese consumers, especially women, showed more astonishing purchasing power on cosmetics abroad than in China since it was cheaper to buy it from abroad. While with the development of e-commerce, more and more consumers are now preferred to buy cosmetics online.

From January to August, Chinese Netizen[1] increased 13.4% compared with the end of 2008. The rapid development of e-commerce has coincided with the rapid development of Internet hosts. According to the data from last year, about 61% of Internet users have browsed shopping websites and about 18% of users have purchased a product online. That is a large number of online consumers. In addition, more than 30 million which is about 8.3% of the total Netizen are cosmetic users.

The BAIDU search index shows that during the first half of 2009, the search index

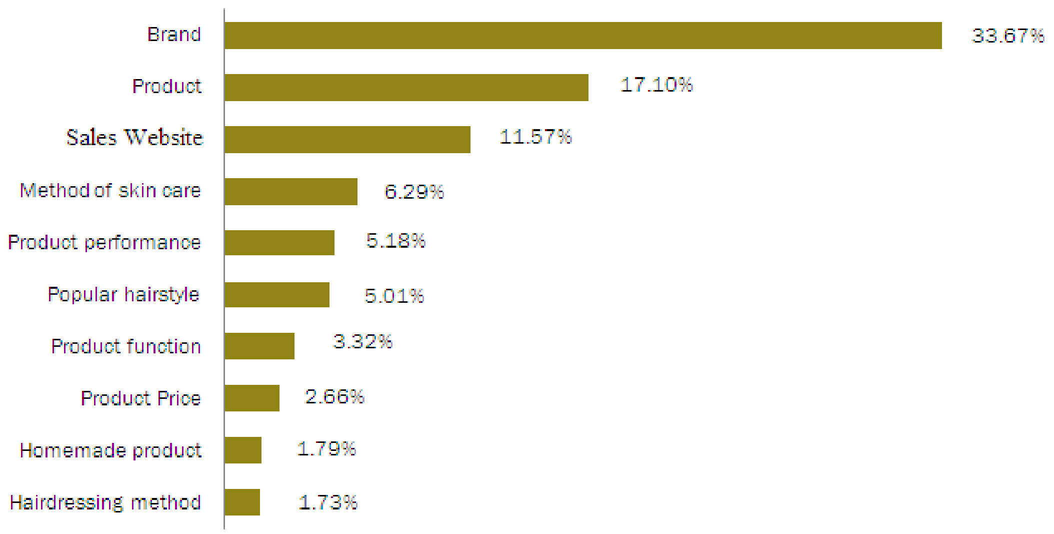
[1] A Netizen (a portmanteau of Internet and citizen) or cybercitizen is a person actively involved in online communities

of cosmetic industry is 10.7% higher than that in 2008.

Specific to China's cosmetic e-commerce market

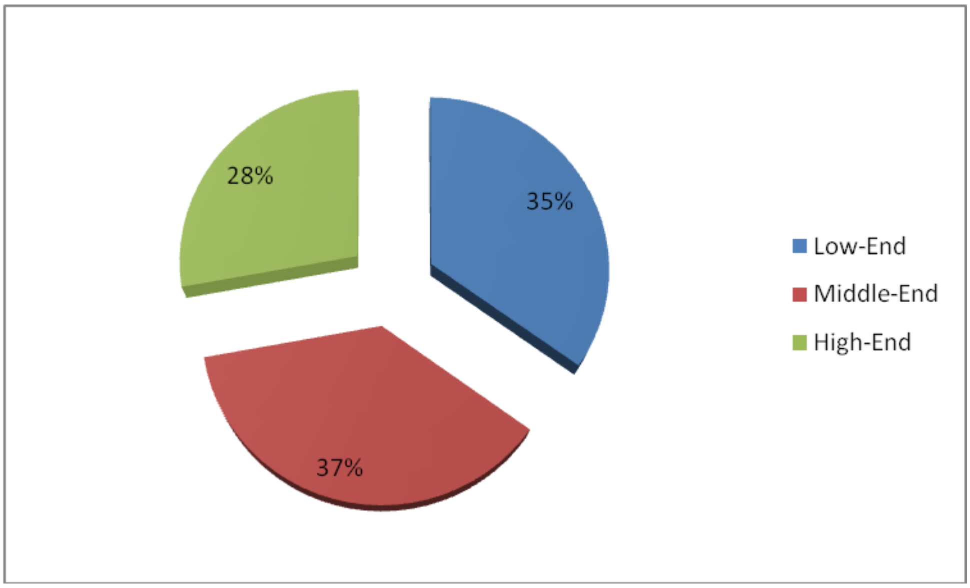
- The BAIDU Search keywords database reflects that when the consumers shop cosmetics online, the most important thing they care about is brand, and then product function, then sales website. Figure 1 shows that 33.67% of the total search of cosmetics is about the brand, 17.10% is about product function and 11.57% is about the sales website.

Figure 1 Ranking of cosmetica industry info in the first half of 2009



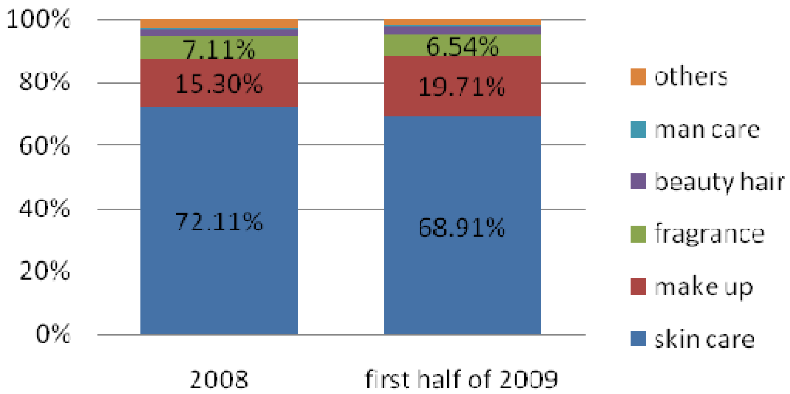
- Through Graph 1, during the first half of 2009, 34.92% of the transaction in cosmetics e-market belongs to Low-End brand, 37.21% is in Middle-End brand and 27.87% is High-End brand.

Graph 1 Consume of the Brand in Cosmetics E-Transaction



- In China, there are lots of different cosmetic products like skincare, make up, fragrance and so on. In graph 2, Skincare and makeup are the top two sellers in cosmetic products.

Graph 2 The Changes in Focusing on Category of Cosmetics



- When consumers shop online, they will evaluate three different characteristics of e-tailors: (1) whether the products are delivered on time; (2) did the product meet the expectation of the consumers; and (3), how are e-tailors after-sale services.
- In additions, the Figure 2 and Figure 3 show that no matter for the skincare or makeup, Guangdong, Shanghai and Beijing are the three largest region of cosmetic e-commerce or you can say they are the most developed cosmetic e-commerce markets in China.

Figure 2 The Regional Distribution of Consumers Buying Skincares Online

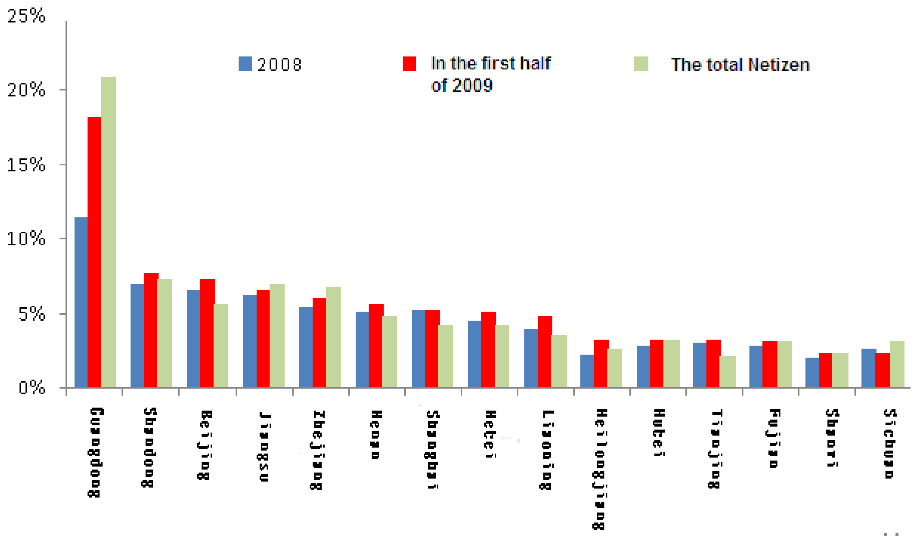
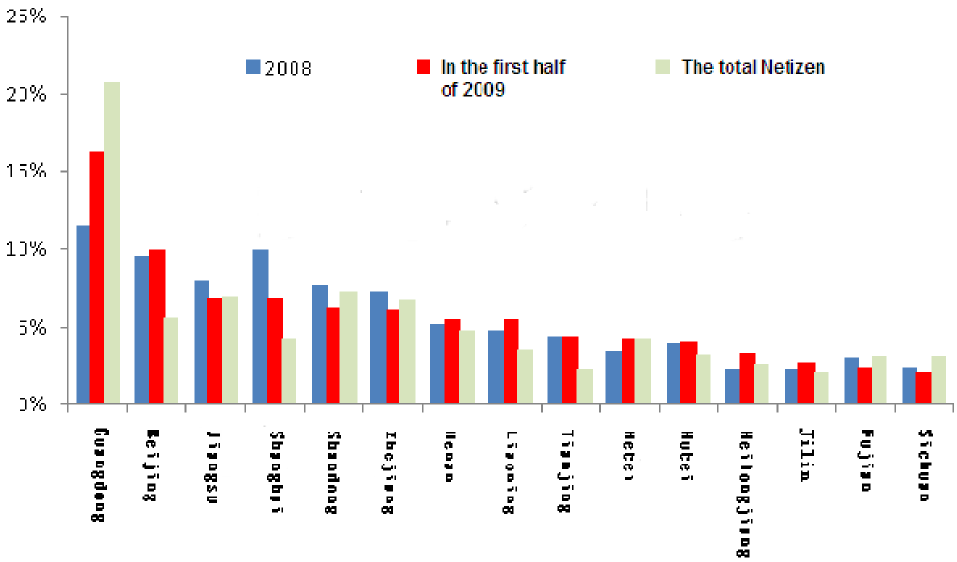


Figure 3 The Regional Distribution of Consumers Buying Makeups Online



Although there is a lot of information about cosmetic e-commerce in China, none of them show how competition and prices are evolving in China's cosmetic e-commerce market. In other words, the information doesn't show the e-tailor's pricing behavior in China's cosmetic e-commerce. This thesis will analyze the factors that determine the price dispersions of cosmetics in China's e-market. Some product functions, brands, and also the e-tailor's characteristics may result to the price dispersion.

1.2 Hypotheses

This thesis will test the following research questions:

1) There is price dispersion in China's cosmetic e-commerce.

2) The online cosmetic price is determined by product characteristics, brand and e-tailor's characteristics.

3) The most determinant factor of the price dispersion is the e-tailor characteristics.

4) HLM is better fit in the case of China's cosmetic e-commerce than OLS model.

1.3 Data Collection

This thesis collects price data on 44 different cosmetic items from the 13 active Guangdong e-tailor web sites. Prices are collected by two different product function (makeup and skin care) within different brands. The data which is collected during November 2009 consists of 269 observations. This dataset is comprised of price data, product function, brand and e-tailor characteristics for Cosmetics sold through Internet in China. Cosmetics were chosen because the physical products themselves are homogeneous and are sold through numerous Internet e-tailors. This enabled easier comparison of prices across levels and across a wide variety of firms. The thesis focuses on a sample of goods carefully matched across different e-tailors in China's e-commerce market to eliminate much of the unobserved heterogeneity in the sample, leaving a clearer measure of the difference between the characteristics of the e-tailors. It focuses on the pricing behavior in China's e-commerce market with data collected directly from each e-tailor's web site because there is no universally accepted shopping engine(s) in China at the time the data was collected2.

To test these hypotheses empirically, a Hierarchical linear model (HLM) is used in this thesis. After giving a general introduction of this thesis in Chapter 1, Chapter2 presents the former research on the field of China's E-commerce market and on the

[2] Dongmei Zhao, Scott J. Savage, and Yongmin Chen, 2008

applications of HLM Model. Chapter 3 describes how the data is used in this thesis. Chapter 4 shows the methodology used to test the hypotheses and Chapter 5 presents the estimations and results. Chapter 6 provides the conclusions and policy implications.

CHAPTER II

LITERATURE REVIEW

2.1 Ownership, location and prices in Chinese electronic commerce markets

In the article "Ownership, location and prices in Chinese electronic commerce markets" by Dongmei Zhao, Scott J. Savage, and Yongmin Chen (2008), it investigates (i) whether the widespread availability of information on the Internet has resulted in price convergence in Chinese e-commerce markets. (ii) the extent to which price divergence is explained by product differentiation.

It collects the price data for 535 product items from the 93 active Beijing e-tailor web sites. The data consist of 6316 observations for nine product categories(books, compact discs, laptop computers, gifts, cosmetics, digital video cameras, digital cameras, MP3 music players, and cellular telephones) during July, 2004.

The study first summarizes the statistics (number of items, observations, mean, SD, min, max of each product category) for prices and adjusted prices. It provides the evidences that the e-tailor's price has not converged in Chinese e-commerce market.

And then, to test the extent to which price divergence is explained by product differentiation, it uses the model based on the hedonic pricing approach.

The model is that firm i's price for product item j is described by

$$PRICE_{ij} = \alpha_j + \beta' s_i + \delta' x_i + \varphi' w_i + \gamma z_i + \varepsilon_{ij}\text{, i=1, … , 93, j=1, …. , 535,} \quad (1)$$

where PRICE is total adjusted price, s is a vector of e-tailor service characteristics which consist of DEL, CONV, REL, SPEED, INFO, and RETURN (DEL is factor score for the delivery and payment options, CONV is the factor score for shopping convenience, REL is the factor score for reliability of order fulfillment, SPEED is the factor score for the website's "ease of use", INFO is the factor score for the degree of info describing the product's attributes, and RETURN is the factor score for how easy customers return the product), x is a vector of e-tailor reputation characteristics which consist of CERT, BAIDU, and PURE PLAY (equals one when the e-tailor sells goods and services through an Internet web site only and zero when the e-tailor sells through an Internet web site and has a physical store), w is a vector of e-tailor ownership characteristics which consist of STATE (equals one when national, provisional and/or other levels of government have majority ownership interest in the e-tailor, and zero otherwise) and MNC(stands for multinational corporation), z is a strategic motive control variable which consist of FIRM AGE, α is an individual fixed effect specific to the product item, β , δ , φ and γ are parameters of variables and ε is an error.

Because the effects of e-tailor characteristics on price may be different for different product categories, it estimates price Eq. (1) for each of the K = 9 product categories. To decide the service characteristics, the authors did a questionnaire survey and regressed the price dispersion, measured by standard deviation of price, on the standard deviation of the 27 service characteristics by ordinary least squares (OLS). Got 10 characteristics have significant and positive effects on price dispersion. Then factor analysis was used to classify the 10 characteristics into six independent variables and got the factor scores that we are going to use to run the regression. Here it actually uses Principal component analysis (PCA)

After controlling for different attribute levels within each product item, the results show that:

1. Prices of e-tailors have not converged in China's e-commerce.
2. Service characteristics are significant determinants of online prices, but no specific pattern between product categories can be discerned from the data.
3. PURE PLAY is an important determinant of prices for five of the nine product categories which suggest reputable e-tailors with physical stores charge a price premium.
4. Ownership structure appears to be an important source of observed price differences between e-tailors. The government owned company can set a higher price than those small private businesses.

2.2 Online Price Dispersion Within and Between Seven European Countries

The intention of the article "Online Price Dispersion Within and Between Seven European Countries" by J. Rupert J. Gatti and Paul Kattuman (2003) is to compare degrees of price dispersion between different product categories and between different national markets in Europe.

Firm specific pricing data was downloaded from the dominant European price comparison site "Kelkoo" in seven European countries (Denmark, France, Italy, Netherlands, Spain, Sweden and the United Kingdom), weekly, for 30 weeks from October 25, 2001 until June 6, 2002 giving 17644 individual price observations across 4699 country/product/date specific websites for a selection of 31 unique products across six distinct product categories (games, games consoles, PDAs, music, printers and scanners).

After the reduction of 900 records for which only one firm listed a price –leaving 3799 records with two or more price quotations, the summary statistics (# of items, # of observations, mean, SD, min, max) in different countries for different categories for three commonly used measures of price dispersion (coefficient of variation, range and gap, all in percentages) shows that there is considerable variation in the observed degrees of price dispersion between both countries and categories.

There are a number of possible explanations for these – including differences in the number of firms listing prices, differences in the value of the product as well as potentially important industry specific and country specific differences in market structure.

To obtain a clearer understanding of the relevance and relative importance of these alternative explanations, the authors run the Probit Regressions for positive price dispersions, OLS Regressions on Price Dispersion, and OLS regressions on coefficient of Variation within National Markets.

1. The dependent variables for probit regressions are Marg. Effect on Prob of Coeff. Of Variation, Marg. Effect on Prob of Range, and Marg. Effect on Prob of Gap,

separately. The variables are log of mean price, log of min price, Firm dummies, category dummies, country dummies, and time dummies.

2. For OLS regressions, the dependent variables are log of Coeff. Of Variation, log of Range, and log of Gap, separately. The variables are log of mean price, log of min price, Firm dummies, category dummies, country dummies, and time dummies.
3. For the third kind of regressions, the dependent variables are log of coeff. Of variance in 7 countries separately. The variables are log of mean price, firms dummies, category dummies, country dummies, and time dummies.
4. A robust finding is that relative price dispersion falls as the level of prices rise. It finds that the impact of changes in the number of firms listing prices depends on which measure of price dispersion and which country is considered.

2.3 Can Price Dispersion in Online Markets be Explained by Differences in E-tailor Service Quality?

The article "Can Price Dispersion in Online Markets be Explained by Differences in E-tailor Service Quality?" tested whether price dispersion in online markets can be explained by differences in service quality of e-tailors.

The data which consist of 6,739 price quotes for 581 identical products distributed across eight product categories (Book, CD, DVD, Desktop, Laptop, Personal digital assistant, Software, and Electronics) from 105 e-tailors are drawn from BizRate.com in November 2000. The authors also collect the average consumer ratings of various measures of that e-tailor's service from BizRate.com, information on the type of e-tailor (pure play vs. bricks-and-clicks), whether the item was popular, and the stage in product life cycle of each item to be used as additional factors in the hedonic price regression.

It first calculates the Mean, SD, Minimum, and Maximum for each product category. Then it uses Principal component analysis (PCA) to transform 10 correlated variables (Ease of ordering, Product selection, Product information, Price, Web site navigation and look, On-time delivery, Product representation, Level and quality of customer support, Tracking, Shipping and handling) into a 5 independent variables which consist of Reliability in fulfillment, shopping convenience, Product information, shipping and handling, and price policy). This is a factor analysis of E-tailor Services.

After that, it uses the price index variable for each item which eliminates the cross-item differences and the factor scores from the factor analysis of the services to run the hedonic price regression (Price dispersion due to product/service differentiation) within each product category and compares the results.

$$P_j / \overline{P} = \alpha_0 + \sum_i \alpha_i S_{ij} + V_j^{'}$$

where P_j refers to the price of the e-tailor or retailer j. α is an individual fixed effect specific to the product item. $V_j^{'}$ is idiosyncratic differences in price charged by j due to differences in cost or pricing policy.

Then they perform a stepwise hedonic regression for each category in which each of the three factors is added to the model with e-tailor service attributes, one factor at a time.

1. the type of e-tailor (pure play vs. bricks-and-clicks)

   After controlling for e-tailor service quality, the e-tailor type factor is significant in six of the eight product categories

2. popularity of the product
3. stage in the product life cycle of the item studied (introduction/early growth vs. late growth/maturity)

   The effects of product popularity and stage in product life cycle or market knowledge are insignificant in all the categories except the desktop computer category.

The empirical analysis shows that the proportion of price dispersion explained by heterogeneity in e-tailors is small and a portion of the remaining price dispersion is explained by factors such as e-tailor's type (pure-play vs. bricks-and-clicks). Evidence also indicates that electronic markets are not necessarily information efficient. There are apparently gains associated with searches conducted by those who do not know what the best deal is.

2.4 Analyzing the Determinants of Housing Price: Regional Comparison within Seoul Korea

Rho, Youngwoo(2009) concentrates on analyzing the apartment price difference between Kangnam and Kangbuk. It tests three main hypotheses. 1) The apartment price in Seoul is determined by individual and neighborhood characteristics. 2) The route of neighborhood effect is different from apartment characteristics. 3) Among the neighborhood characteristics, educational environment is the most important factor explaining regional price differences.

To answer this research question, this thesis selects 250 apartments' data-randomly selecting 10 transactions (real estate sales) from each of the 25 Gus-which includes individual characteristics (location, square area, floor type, age of houses, and distance to the subway, region) and neighborhood characteristics (education rate, population density and crime rate) corresponding to each transaction from the housing survey gathered by Kookmin Bank.[3] in 2007. The Kookmin bank surveys the characteristic of houses being financed and publishes the result..

In this paper, it first uses the simple Pearson correlation coefficient between price and some apartment characteristics (area, age, floor) and then analyses the correlation between average price of apartment in each Gu and neighborhood characteristics. It uses a two level Hierarchical Linear Model (HLM) to analyze how much neighborhood-specific characteristics effect on housing price in Kangnam and Kangbuk.

The dependent variable is the real trading prices of apartments of Seoul during the first quarter of 2007. The independent variables are individual characteristics of apartment and the neighborhood characteristics.

The empirical results of this paper suggest that the most important individual apartment feature determining the apartment's price is size. Second, a large portion (more than 38%) of Seoul apartment price variation is due to differences in neighborhood characteristics. Third, the most important neighborhood characteristic

[3] Kookmin Bank "The survey of the state of demand for housing finance" Dec. 2006.

in determining apartment prices is education services. These conclusions answer some of the initial questions and hypotheses of this thesis.

2.5 Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models

The paper focus on the school effects models, designed for data on individuals nested within naturally occurring hierarchies (e.g., students within classes, children within families, teachers within schools);.

It shows general strategies for working with multilevel data in SAS and for creating data sets at several levels.

The data set consists of information for 7,185 students in 160 schools (with anywhere from 14 to 67 students per school). The student-level which is level-1 outcome is MATHACH. The level-1 covariate is SES. There are two school-level (level-2) covariates. One is an aggregate of student level characteristics (MEANSES); the other is a school-level variable (SECTOR). MEANSES and SES are centered at the grand mean (they have means of 0). SECTOR, a dummy variable, is coded 0 and 1.

The author begins by fitting an unconditional means model which is viewed as a one-way random effects ANOVA model, examining variation in MATHACH across schools.

Combined $Y_{\mathrm{ij}} = \beta_{0j} + r_{ij}$ where $r_{ij} \sim$ iid N(0, σ^2)

and $\beta_{0j} = \gamma_{00} + u_{0j}$ where $u_{0\mathrm{j}}$~iid N(0, τ_{00}),

it gets $Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$ where $r_{ij} \sim$ iid N(0, σ^2) and $u_{0\mathrm{j}}$~iid N(0, τ_{00}).

Here β_{0j} is an intercept for the student's school, γ_{00} is a average MATHACH score in the population at level-2, u_{0j} is a series of random deviations from the mean, and r_{ij} is a random err associated with the ith student in the jth school. τ_{00} tells the variability in school means and σ^2 tells us about the variability in MATHACH within schools.

$$\hat{\rho} = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}^2}$$

ρ here tells us what portion of the total variance occurs between schools.

And then he sequentially examine the effects of a school-level (level-2) predictor (MEANSES) and a level-1 predictor (student SES).

$$Y_{ij} = \beta_{0j} + r_{ij} \quad \text{and} \quad \beta_{0j} = \gamma_{00} + \gamma_{i1} MEANSES_j + u_{0j} + r_{ij}$$

Substituting the level-2 into level-1 equation yields:

$$Y_{ij} = \lfloor \gamma_{00} + \gamma_{i1} MEANSES_j \rfloor + \lfloor u_{0j} + r_{ij} \rfloor$$

MEANSES indicates the average SES of the children within the school.

Then it illustrate the effect of including student level predictors by initially examining a model with only one student –level predictor (SES).

$$\begin{aligned} Y_{ij} &= \gamma_{00} + u_{0j} + (\gamma_{10} + u_{i\backslash 1j})(SES_{ij} - \overline{SES_J}) + r_{ij} \\ &= [\gamma_{00} + \gamma_{10}(SES_{ij} - \overline{SES_j})] + [u_{0j} + u_{1j}(SES_{ij} - \overline{SES_j}) + r_{ij}] \end{aligned}$$

Finally, he combined both types of predictors into a single model.

$$\begin{aligned} Y_{ij} = {} & \gamma_{00} + \gamma_{01} MEANSES_j + \gamma_{02} SECTOR_j + \gamma_{10}(SES_{ij} - \overline{SES_j}) + \gamma_{11} MEANSES_j (SES_{ij} - \overline{SES_j}) \\ & + \gamma_{12} SECTOR_j (SES_{ij} - \overline{SES_j}) + u_{0j} + u_{1j}(SES_{ij} - \overline{SES_j}) + r_{ij} \end{aligned}$$

The empirical analysis shows that:

1. Schools do differ in their average MATHACH scores and that there is even more

variation among students within schools.

2. ρ =.18 tells us that there is a fair bit of clustering of MATHACH within school which suggests that an OLS analysis of these data would likely yield misleading results.
3. 69% of the explainable variation in school mean math achievement scores is explained by MEANSES.
4. The main effect of SECTOR tells us the intercepts in these two models are significantly different.

2.6 A Hierarchical Linear Model Approach for Assessing the Effects of House and Neighborhood Characteristics on Housing Prices

The objective in this paper is to demonstrate how HLM can be used to separate the variation in housing prices into the portion that depends on house-specific characteristics and that portion that depends on neighborhood-specific characteristics. It uses HLM model because this model accounts for the inherent hierarchy in the housing purchase decision and also housing price.

Geographic information systems (GIS) data which constructed with 725 dwellings located in 60 different Census Block Groups (referred to as a “neighborhood” here) sold in 1984 and 1985 in a single mid-size city (with a population of 43,765 in 1980 and 45,090 in 1986) in the upper Midwest are being used. This data comes from the city assessor’s office, the 1980 Census Tapes, the State Tax Department and the office of the Assistant Superintendent of Elementary Education.

Since the focus here is a simple HLM model, only two important variables are employed. “Area” represents the total square feet for a particular house. The data comes directly from the city assessor’s office. “Median Travel Time” represents the median travel time to be related negatively to the house price holding all else constant. This data comes from the 1980 Census.

Then three models are used to test the hypothesis:

1. The standard one-way ANOVA model $\mathrm{Y}_{ij} = \gamma_{00} + u_{0j} + r_{ij}$ which is combined from $\mathrm{Y}_{ij} = \beta_{0j} + r_{ij}$ and $\beta_{0j} = \gamma_{00} + u_{0j}$
2. Y_{ij} is house i’s price in neighborhood j. In the level-one model, $\beta_{1\mathrm{j}}$ is set equal to zero for all j, β_{0j} will be the mean outcome for the jth neighborhood and r_{ij}, the level-one error or a house effect, is normally distributed with mean zero and variance σ^2. In the level-two model, γ_{01} is set to zero, γ_{00} will be the grand-mean house price for all dwellings in the sample, u_{0j} is a neighborhood effect and is assumed to have mean zero and variance τ_{00}. This model provides a simple test to determine whether there is statistically significant variance at the neighborhood level. The results of this model can be used to calculate the “intraclass correlation coefficient” (Raudenbush and Bryk, 2002) given by $\rho = \tau_{00} / \left(\tau_{00} + \sigma^2\right)$
3. The next model is the random- coefficients model.

   $$Y_{ij} = \gamma_{00} + u_{0j} + \gamma_{10}(X_{ij} - \overline{X_j}) + u_{1j}(X_{ij} - \overline{X_j}) + r_{ij}$$

   This allows both β_{0j} and β_{1j} to vary across neighborhoods while controlling for the area of the house at level one.

   To find the relationship between neighborhood mean house prices and the neighborhood effects of square footage, it calculates the correlation between means and slopes across neighborhoods using $\frac{\tau_{01}}{\sqrt{\tau_{00}\tau_{11}}}$

4. The results of the random-coefficients model suggest there is significant unexplained variance in housing prices that remains at level two. To explain some of this variation by including an explanatory variable at level two, it uses the complete model

   $$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{00}(X_{ij} - \overline{X_j}) + \gamma_{11}W_j(X_{ij} - \overline{X_j}) + u_{0j} + u_{1j}(X_{ij} - \overline{X_j}) + r_{ij}$$

   Where Y_{ij} is house i's price in neighborhood j; X_{ij} is the area of house i in neighborhood j measured in square feet; $\overline{X}_{\mathrm{j}}$ is the average area for all houses in neighborhood j; and r_{ij} is a stochastic error term assumed to be distributed N(0, σ^2). W_j represents the median travel time to work in neighborhood j.

   For comparison purposes, it also presents the OLS estimates using robust standard errors for this model.

The empirical analysis shows that:

1. The standard one-way ANOVA model shows that the variance at the neighborhood level is statistically significant at better than the 1% level of significance. By calculating the "intraclass correlation coefficient" ρ, it implies that more than 51% of the variance in house prices is due to differences at the neighborhood level which strongly suggests that housing prices vary across neighborhoods.
2. The results of the random-coefficients model indicate there is strong evidence that house prices differ significantly across neighborhoods and also the little

correlation which is 0.10 suggests that the rate of increase in house prices for additional square footage is likely to be similar across neighborhoods regardless of the neighborhood's mean house price.

3. The neighborhoods with higher median travel times tends to have lower average house prices but has positive relationship with the increase in house prices as square footage increases.
4. HLM provides "better" estimates from a statistical standpoint with little added computational difficulty

2.7 Summary

Through reviewing these literatures, both OLS Model and Hierarchical Linear Model can be applied to analyze the hypothesis. But the three-level HLM is more appropriate than the simple OLS[4] because the dependent variables in the case (products' prices made by each e-tailor) share some similar characteristics and brands that makes error term in the OLS model neither unbiased nor consistent while the error term is not assumed to be independent and doesn't need to have a constant variance in the HLM.

After collecting the data, first the thesis tests whether there is price dispersion of cosmetics between China with summarizing the statistics (number of items, number of observations, mean, SD, min, max) for different items of cosmetics. Next, the thesis uses the simple Pearson to test correlation coefficient between price and some product functions and also analyzes the correlation between average price of

[4] Stephen W. Raudenbush and Anthony S. Bryk, 2002

products in each brand and brand characteristics to see whether they are statistically significant. Then it analyzes the correlation between average price of products in each e-tailor and e-tailor characteristics to see whether they are statistically significant. If they are true, it means that these variables are related to the price different. Finally, the thesis uses the three-level HLM model to investigate the e-tailor's pricing behavior in China's cosmetic e-commerce market.

CHAPTER III

EMPIRICAL MODEL AND DATA

3.1 Data Description

his thesis collects price data on 44 different cosmetic items from the 13 active Guangdong e-tailor web sites. Prices are collected by two different product function (makeup and skin care) within different brands. The data which is collected during November 2009 consists of 269 observations. This dataset is comprised of price data, product characteristics, brand and e-tailor characteristics for Cosmetics sold through Internet in China. Cosmetics were chosen because the physical products themselves are homogeneous and are sold through numerous Internet e-tailors. This enabled easier comparison of prices across levels and across a wide variety of firms. The thesis focuses on a sample of goods carefully matched across different e-tailors in China's e-commerce market to eliminate much of the unobserved heterogeneity in the sample, leaving a clearer measure of the difference between the characteristics of the e-tailors. It focuses on the pricing behavior in China's e-commerce market with data collected directly from each e-tailor's web site because there is no universally accepted shopping engine(s) in China at the time the data was collected.

This thesis gathers the price of 44 different products of 26 brands sold in 13 different e-tailors. The price is the dependent variable and product, brand and e-tailor characteristics are the independent variables in the empirical model. Skincare, size (in milliliter) and imported are three characteristics of products. The variable of Skincare equals to 1 if the product belongs to skincare. If the product is makeup, the variable of Skincare equals to 0. The variable of imported is 1 if the product is imported. Otherwise, it is 0. The brands are sorted from 1 to 26. 1 to 9 belongs to high street brand, 10 to 17 belongs to middle-end brand and 18 to 26 belongs to high-grade brand. E-tailor's characteristics are: customer satisfaction index, product delivered on time, product met customer's expectation, customer support, would like to shop here again, and Reliability. These characteristics are all collected from each e-tailor's website. Customer satisfaction index is ranging from 0 to 1. The higher the score is, the more satisfied the customer is with the e-tailor. Later in this thesis, it will be referred as CustSat. Product delivered on time which will be referred as Ontime in the thesis is ranging from 1 to 5. The longer the time it takes to deliver the product, the larger score is. Both product met customer's expectation (which will be referred as MetExp) and customer support (which will be referred as CustSup) are ranging from 1 to 5. The higher the score is, the better the product met the expectation and the customer's support. Would like to shop here again (shophere) refers to how many customers are willing to shop here again and Reliability (Relia) refers to how many customers has shopped from the e-tailor.

In sum, this thesis uses information on price per item, brand, product characteristics such as imported, size, and skincare, etc. In addition, the scores of e-tailors, such as CustSat, Ontime, MetExp, CustSup, Shophere and Relia.

Table 1 shows the example of cosmetics data used in this thesis.

Table 1 - Example of Cosmetic Data

Product	1	1	43
Skincare	1	1	0
Size (ml)	360	360	30
Imported	1	1	0
Brand	18	18	1
Etailor	1	2	8
CustSat	0.2561	0.6203	0.2134
Ontime	4.8	4.9	4.5
MetExp	4.7	4.9	4.7
CustSup	4.8	4.8	4.8
Shophere	36278	2896	9770
Relia	141649	4669	45774
Price (RMB)	328	320	0.8

3.2 Descriptive statistics

To analyze the correlation between price and product's characteristics the price of each item needs to be translated to average prices of each product.

Table 2 – Average Product Price and Product Characteristics

Product	Price	Imported	size	skin care
1	332	1	360	1
2	321	1	30	1
3	55	0	50	1
4	214	0	5	1
5	213	1	125	1
6	71	1	50	1
7	196	1	30	1
8	160	1	250	1
9	323	1	15	1
10	38	1	60	1
11	169	1	15	1
12	50	0	15	1
13	129	1	60	1
14	393	1	30	1
15	10	0	60	1
16	34	0	400	1
17	531	1	50	1
18	533	1	50	1
19	298	1	60	1
20	79	0	30	1
21	10	1	20	1
22	195	1	500	1
23	14	0	400	1
24	19	0	160	1
25	78	1	30	0
26	31	1	10	0
27	19	0	150	0
28	56	0	30	0
29	48	1	30	0
30	45	1	30	0
31	16	0	30	0
32	112	1	30	0
33	141	1	30	0
34	95	1	30	0
35	15	0	50	0
36	6	0	50	0
37	140	1	200	0
38	57	1	250	0
39	51	0	200	0
40	11	0	80	0

41	36	1	10	0
42	150	1	5	0
43	1	0	5	0
44	10	0	10	0
42	150	1	5	0
43	1	0	5	0
44	10	0	10	0

Table 3: The correlation between Average Product Price and Product Characteristics

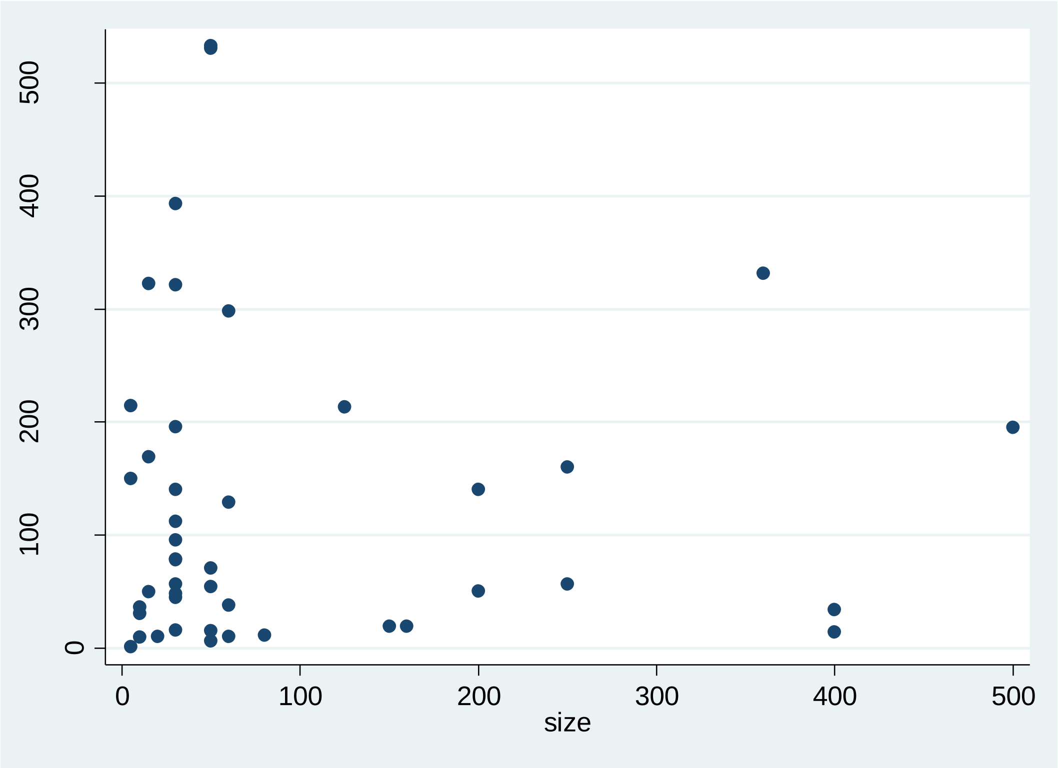
Correlation	Imported	Size	Skincare
Price	0.5069	-0.0021	0.4680
P-value	0.001	0.551	0.001

Table3 shows that the price and Imported have a positive, statistically significant (p=0.001) and high correlation (r=0.5069), suggesting that price is relatively close related to whether the product is imported or not. It's an economic common sense that the price of imported cosmetics is higher than that of the domestic cosmetics. So the result is coincided with common sense.

The coefficient of correlation between price and skincare is 0.4680 and statistically significant (p=0.001). The coefficients of the size are negative and are not statistically significant (p=0.551). The results suggest that the price and the size are not directly related to each other. The negative relation between price and size means that the bigger the product is, the lower the price is. It is not coincided with common sense.

Graph 3 shows the relation between price and size as predicted by the correlation coefficients.

Graph 3: The correlation between price and size



To analyze the correlation between price and brand, the price of each item needs to be translated to average prices in each brand

Table 4: Average brand price and brand

Brand	Price	Brand	Price
1	1.071429	14	160.3333
2	6.365	15	214.1667
3	14.08	16	30.5
4	10.24	17	195.4286
5	10.12429	18	331.5
6	36.16667	19	54.60769
7	11	20	72.87027
8	53.09375	21	298.1429
9	38.165	22	162.3933
10	14.375	23	128.8571
11	19	24	121.65
12	56.7375	25	278.6471
13	15.28333	26	384

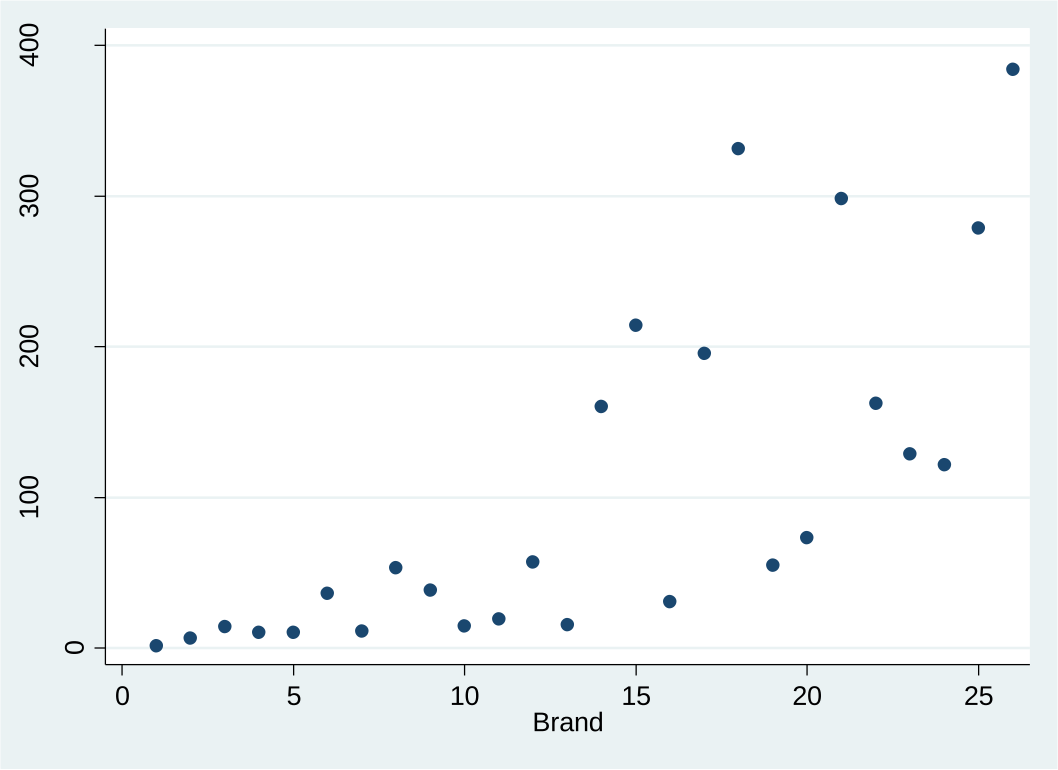
The Pearson correlation coefficients between price and brand are shown in Table 5.

Table 5-The coefficient of correlation between average brand price and brand

Correlation	Brand
Price	0.7384
P-value	0.000

The high correlation coefficient is between price and brand (r=0.7384). This suggests that the cosmetics price is very closely related to the brand. The brand can be thought as a representative of cosmetics products. So table 5 shows that brand is a very important factor in explaining the cosmetics prices.

Graph 4: The correlation between average brand price and brand



To analyze the correlation between price and e-tailor's characteristics the price of each item needs to be translated to average prices in each e-tailor.

Table 6 Average E-tailor price and E-tailor's characteristics

Etailor	Price	CustSat	Ontime	MetExp	CustSup	Shophere	Relia
1	163.38	0.2561	4.8	4.7	4.8	36278	141649
2	135.6	0.6203	4.9	4.9	4.8	2896	4669
3	159.28	0.2805	4.6	4.6	4.7	3699	13185
4	119.7	0.1327	4.8	4.7	4.6	6415	48324
5	141.38	0.3577	4.7	4.7	4.7	26458	73974
6	133.3	0.2713	4.7	4.7	4.7	9751	35948
7	82.68	0.229	4.8	4.7	4.7	9460	41304
8	171.42	0.2134	4.5	4.7	4.8	9770	45774
9	191.29	0.8261	4.8	4.8	4.7	15203	18403
10	171.41	0.333	4.7	4.6	4.7	34251	102842
11	84.01	0.1782	4.7	4.7	4.8	18606	104417
12	85.95	0.2543	4.8	4.8	4.7	2034	7997
13	113.06	0.3741	4.7	4.6	4.8	5300	14167

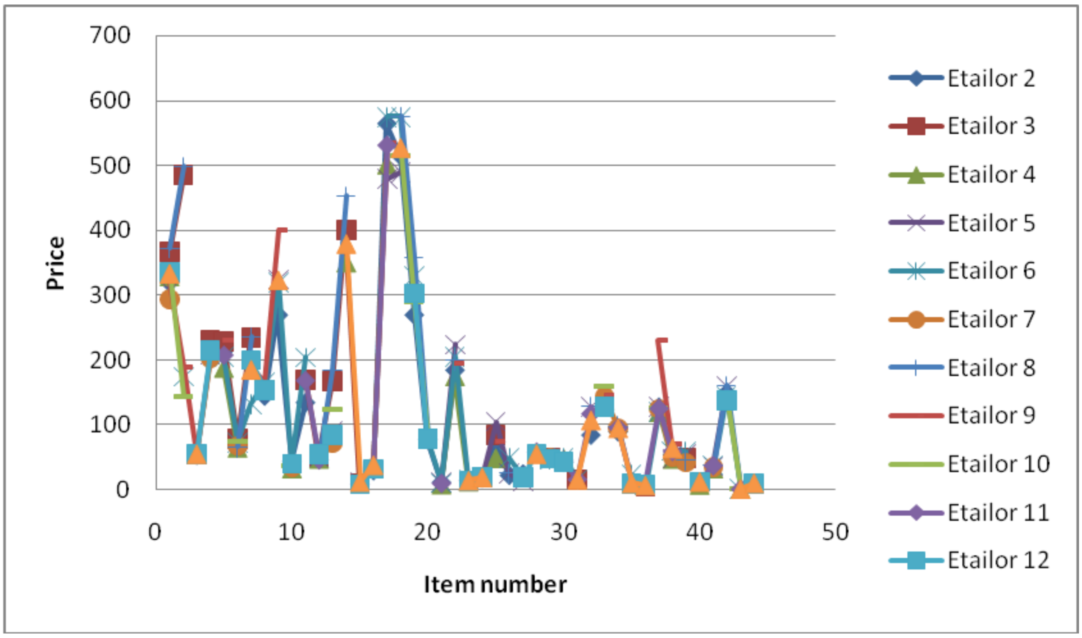
The Pearson correlation coefficients between average e-tailor's price and e-tailor's characteristics are shown in Table 7.

Table 7 Correlation coefficients between average e-tailor's price and e-tailor's characteristics

Correlation	CustSat	Ontime	MetExp	CustSup	Shophere	Relia
Price	0.4801	-0.2838	-0.0804	0.0312	0.3950	0.1277
P-value	0.164	0.079	0.776	0.368	0.851	0.779

The correlation coefficients between price and e-tailor's characteristics are not high enough to say any of the e-tailor's characteristics can be thought as a representative of e-tailor's quality in specific e-tailor. But the ontime is statistically significant (α=0.10). It shows that the product price per e-tailor is slightly negatively related to the score of product has been delivered on time. Usually it is argue that customer prefer to pay a premium to the product which could be delivered faster. The empirical evidence shows that this common sense can be apply to China's cosmetics e-market. In other words, the longer the delivery takes, the cheaper the price of the product.

Figure 4 Price dispersion within different e-tailors



As it is shown in Figure 4, for different 44 items, the prices provided by 13 e-tailors are not all the same as each others. There is a huge price dispersion in Cosmetic E-Commerce in China.

CHAPTER IV

METHODOLOGY

Given the evidence above, this thesis will apply a three-level HLM model to the Chinese cosmetics e-market and analyze how much product, brand and e-tailor characteristics effect on cosmetics price in Guangdong.

Level-1: $$Y_{ijk} = \pi_{0jk} + \sum_{p=1}^{P} \pi_{pjk} a_{pjk} + e_{ijk}$$

Level-2: $$\pi_{pjk} = \beta_{p0k} + \sum_{q=1}^{Q} \beta_{pqk} X_{pjk} + r_{pjk}$$

Level-3: $$\beta_{pjk} = \gamma_{pq0} + \sum_{s=1}^{S} \gamma_{pqs} W_{pqk} + u_{pqk}$$

Level 1 explained product characteristics. Level 2 explained brand characteristics and level-3 explained e-tailor's characteristics.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z	proportion
Intercept	Productid	8055.70	2408.11	3.35	0.0004	0.4391
Intercept	BrandNum	9259.01	4116.74	2.25	0.0123	0.5047
Intercept	EN	88.7868	59.9266	1.48	0.0692	0.0048
Residual		943.02	91.4857	10.31	<.0001	

The product, brand and e-tailor are all significant (α=0.1) which means they are important determinants of the price variance. The total variance is 18346.52. 44% of price variance is explained by product level. Brand explains 50% of the price variance. While the e-tailor level only explains 0.48% of the price dispersion, it will be added as fixed effects in this thesis.

Then the three-level HLM model becomes two-level HLM model. It is divided into two parts, product nested in brand. The first level of the HLM is expressed by following equations.

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}, \qquad (1)$$

where Y_{ij} is price of product i of brand j; X_{ij} is the item's brand id; r_{ij} stochastic error term assumed to be distributed N(0, σ^2).

HLM assumes that the coefficients of equation (1) are expressed as equation (2) ~ (4) which belongs to the second level of HLM.

(2) $\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + U_{0j}$ $\qquad U_{0j} \sim N(0, \tau_{00})$

(3) $\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + U_{1j}$ $\qquad U_{1j} \sim N(0, \tau_{11})$

Wj: characteristics of brand j

substituting equation (2) and (3) for β_{0j} and β_{1j} in equation (1) we have the

Equation (4) combined model:

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{00} X_{ij} + \gamma_{11} W_{ij} X_{ij} + U_{0j} + U_{1j} X_{ij} + r_{ij} \qquad (4)$$

This thesis tests how the vector of X characteristics impacts the price of cosmetics by using HLM model. Given the nested nature of the data where items are clustered within each product and within brand, the HLM model is a superior than the OLS model. OLS assumes that error term in the model is independent and has constant variance, but in this case, item prices of each product share the same brand and therefore the independent of errors is violated. In HLM models, the error term is not assumed to be independent and doesn't need to have a constant variance. With this data, the OLS estimators are neither unbiased nor consistent, and therefore a Hierarchical Linear Models is more appropriate than the simple OLS. (Stephen W. Raudenbush and Anthony S. Bryk, 2002)

Hierarchical linear models let capture the two-step process Chinese use when selecting what to buy. At first, they decide the brand in which they will buy the cosmetics and then, decide the specific product characteristics of the cosmetics they want to use.

The first model is the one-way ANOVA model which estimates how much product and brand affect the cosmetics price. The second one is the random Coefficient model shows how product characteristics affect the cosmetics price. The third one is the "Combined Model" shows how both product characteristics and brand affect the cosmetics price. This model provides useful preliminary information regarding the outcome variability at each of the two levels.

- In the level-one model, β_{ij} is set equal to zero for all j, giving:

$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}$

Let β_{ij} equal to zero for all j

equation (1) leads to $$Y_{\text{ij}} = \beta_{0j} + r_{ij} \qquad (5)$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad (6)$$

β_{0j} is the mean outcome for the jth brand and r_{ij} is the level 1 error follows N(0, σ^2).

The standard one-way ANOVA model is:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \qquad (7)$$

- Random Coefficient model:

  In the level-two, where β_{1j} is specified as $\beta_{1j} = \gamma_{10} + U_{1j}$ (8)

  Now the equation becomes:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + U_{1j} X_{ij} + U_{0j} + r_{ij} \qquad (9)$$

  In equation (9), β_{0j}, β_{1j} vary across brands.

- Combined Model:

  In equation (4), β_{0j}, β_{1j} varied across brand and they are a function of brand.

CHAPTER V

RESULTS OF REGRESSION MODELS

1. One-way Anova model

The empirical results of using the one-way ANOVA are shown in table 8.

Table 8 Result from the One-Way ANOVA Model

Fixed Effect	**Coefficient**	**Standard Error**	**t-value**	**P-value**
Intercept	109.85	24.7902	4.43	.
Random Effect	**Variance**	**Standard Error**	**Z-Value**	**Pr Z**
Intercept (Product)	8137.05	2434.80	3.34	0.0004
Intercept (BrandNum)	9280.24	4140.22	2.24	0.0125
Residual	1023.844	96.52513	10.31	<0.0001

According to the above results, the intercept coefficient of the fixed effect shows the overall average price across all cosmetics in the data sold in Guangdong's e-market across products and brand without controlling for any characteristic is ￥109.85 with an standard error of ￥24.79. The range of cosmetics prices at 95% confidence interval is ￥109.85± 1.96* ￥24.7902 = (￥61.26 to￥158.44).

The random effect shows the variance at the product and brand level, both p-values are less than 0.05% suggesting that the variance at both levels are statistically significant which means both of them are important affects of price dispersion.

In addition, the intra-class correlation coefficient 44.12% of cosmetic price variation (8137.05/ (8137.05+9280.24+1023.844)) is due to difference at the product level and 50.32% is explained by brand difference. This empirical result shows that the product characteristics and brand are important factors determining the prices of cosmetics in Guangdong’s e-market.

2. Random Coefficient Model

Table 9 Result from the Random Coefficient Model with Product Characteristics

Fixed Effect	**Coefficient**	**Standard Error**	**t-value**	**P-value**
Imported (0/1)	111.78	39.9099	2.80	0.0055
Skincare (0/1)	111.26	27.1530	4.10	<.0001
Random Effect	**Variance**	**Standard Error**	**Z-Value**	**Pr Z**
Intercept (Product)	5276.28	1667.76	3.16	0.0008
Intercept (Brand)	5675.88	2760.66	2.06	0.0199
Residual	1023.85	96.5263	10.61	<0.0001

Table 9 shows whether the cosmetics are imported or domestics and whether it is skincare or makeup are important determinants of the price variations in Guangdong's cosmetics e-market in China.

The total random effect drops from 18441.1 to 11976 which is about 35% (11976/18441.1) after adding the Imported and Skincare which means these two factors are important determinants of the price variance.

In addition, the Imported and Skincare are not highly correlated with product and brand because the product variance and brand variance decreased by about 35% and 38% respectively which are almost the same proportion as the total coefficient drops (35%).

The slope coefficient of Imported are 111.78 which means imported cosmetics are more expensive, consumers are willing to pay on average ￥111.78 more than for domestic cosmetics.

The coefficient of Skincare is 111.26 which shows that skincare cosmetics are also more expensive than makeup. On average, the consumers are willing to pay ￥111.26 more than that for makeup.

3. Adding product & E-tailor's factors

Table 10 The result of controlling the product and e-tailor's characteristics.

Fixed Effect	**Coefficient**	**Standard Error**	**t-value**	**P-value**
Imported (0/1)	114.43	39.8469	2.87	0.0044
Skincare (0/1)	111.64	27.0254	4.13	<.0001
On-time Delivery (1-5)	-21.1057	11.1960	-1.89	0.0605
Customer Support (1-5)	108.27	37.3595	2.90	0.0041
Shop Here Again	-0.00042	0.000207	-2.05	0.0416
Random Effect	**Variance**	**Standard Error**	**Z-Value**	**Pr Z**
Intercept (Product)	5217.92	1648.45	3.17	0.0008
Intercept (Brand)	5684.48	2748.18	2.07	0.0193
Residual	995.65	94.9257	10.49	<0.0001

The random effect doesn't change too much (from 11976 to 11898.05 which is less than 1%) which means adding the e-tailor characteristics doesn't affect too much of the price variation. Imported (0.0044) and Skincare (<.0001) are still significant even after adding e-tailor's characteristics.

To the e-tailor's characteristics:

The score of Customer Support (0.0041) and Consumers would like to Shop in that e-tailor again (0.0416) are all significant factors explaining price variation. On time delivery was not significant ($p>0.05$).

The coefficient of Customer Support (108.27) shows e-tailors with one point higher in Customer Support score are able to charge ￥108.27 more than e-tailors with low customer support score. This goes along with the philosophy that pays for what you get. People are willing to pay higher price for the better service they get.

The slope coefficient of Shop Here again is -0.00042.It tells e-tailor with lower prices are the one that more customer agreed to shop at that e-tailor again which let that e-tailor could set ￥0.00042 lower than other e-tailors if one more customer would like to buy from this e-tailor again . E-tailors prefer higher sale volume instead of higher price to gain high profit margins.

4. Adding product factors and E-tailor

Instead of using e-tailor's characteristics, use the e-tailor itself as the fixed effects. And at the same time, adding the product characteristics. Table 11 shows the result of controlling the product characteristics and e-tailor.

Table 11 shows the result of controlling the product characteristics and e-tailor

Fixed Effect	**Coefficient**	**Standard Error**	**t-value**	**P-value**
Imported	111.80	39.3124	2.84	0.0048
Skincare	108.64	26.0475	4.17	<.0001
E-tailor 8	34.0633	9.3576	3.64	0.0003
Random Effect	**Variance**	**Standard Error**	**Z-Value**	**Pr Z**
Intercept (Product)	5114.70	1585.05	3.23	0.0006
Intercept (Brand)	5505.30	2620.00	2.10	0.0178
Residual	946.27	91.9035	10.30	<0.0001

The random effect doesn't change too much (from 11898.05 to 11566.27 which is less than 1%) which means adding the e-tailor doesn't explain too much of the price variation. Imported (.0048) and Skincare (<.0001) are still significant with the e-tailor was added.

For the E-tailors, only e-tailor 8 is statistically significant. The slope coefficient is 34.0633 which means the e-tailor 8 can set ￥34.06 higher than other e-tailors.

Table 12 E-tailor Characteristics Scores

E-tailor	On-time	Met Expectation	Custom Support	Shop here again
1	4.8	4.7	4.8	36278
2	4.9	4.9	4.8	2896
3	4.6	4.6	4.7	3699
4	4.8	4.7	4.7	6415
5	4.7	4.7	4.7	26458
6	4.7	4.7	4.7	9751
7	4.8	4.7	4.6	9460
8	4.5	4.7	4.8	9770
9	4.8	4.8	4.7	15203
10	4.7	4.6	4.7	34251
11	4.7	4.7	4.8	18606
12	4.8	4.8	4.7	2034
13	4.7	4.6	4.8	5300

In table 12, the e-tailor 8 has higher custom support score and less customers would like to shop there again. So it charges higher prices. This is consistent with the main findings from HLM estimation.

5. Comparing the goodness of fit

Table 13 shows the goodness of fit of each model. Combined model with e-tailors and product characteristics have the smallest index of goodness of fit. It means that this model can have the larger explanatory statistical power than other estimated models. In other words, by including the product factor and e-tailor as fixed effects in the model is able to explain the cosmetics prices in China better.

Table 13 Comparing the goodness of fit

Model Fit Statistics	Model 1	Model 2	Model 3	Model 4
Deviance-a	2814.0	2776.6	2749.5	2636.9
AIC-b	2820.0	2782.6	2755.5	2642.9
AICC	2820.1	2782.7	2755.6	2643.0
BIC-c	2814.0	2776.6	2749.5	2636.9

a: Log Likelihood Statistic

b: Akaike Information Criterion

c: Bayesian Information Criterion

**a~c: As the value is smaller, the model is statistically better.

In short, 94% of the price dispersion in cosmetic E-commerce in China can be explained by product characteristics (44.12%) and brand differences (50.32%).

In general of cosmetics sold online in Guangdong Province, the prices of imported cosmetics are ￥111.78 higher than domestic cosmetics. The prices of skincare are ￥111.26 higher than makeup.

Some of the e-tailor’s characteristics are also important

Custom support: the better the custom support, the higher price the e-tailor can set (108.27)

Shop here again: e-tailor can set a lower price if more customers would like to shop in that e-tailor again. Although the price decreased, the profit can still be increased since the sale volume is increased. (-0.00042).

CHAPTER VI

CONCLUSION

Don’t pay too much on makeup. Better skin condition can make you save a lot on doing the makeup. Since the skincare is more expensive than the makeup, it also has more promotions than makeup. So further research may need to include how many promotions has that product did during that time period so that the price drop according to the promotion can be tested.

More study need to be done on the on-time delivery. On-time delivery behaved a little bit weird in this dataset. In common sense, the higher on-time delivery score could let the e-tailor set higher price while in this paper it seems the opposite way. More products or more price data need to be included to test whether this on-time delivery is positive affect for the price variation.

E-tailors experience such as how long have they opened their business needs to be investigated too. The number of customers would like to shop here again is highly correlated with the time they opened their business. The longer the e-tailor has settled down its business, the more customers could have shopped in that e-tailor.

This thesis doesn’t concern the consumer characteristics which may include such as the age, gender, income, etc. Young girls may pay more on the makeup than the skincare while old women or men who barely do the makeup only purchase skincare which makes them spend all their budgets on the skincare. In addition, the student without any income may not be able to afford the high-end product such as Dior or LANCOME. But the millionaires may disdain to purchase the low-end product such as Gongdeng. So more research needs to be done on customers characteristics.

REFERENCE

1. Dongmei Zhao, Scott J. Savage, and Yongmin Chen, Ownership, location and prices in Chinese electronic commerce markets. *Information Economics and Policy 20,* 2008, 192–207

2. J. Rupert J. Gatti and Paul Kattuman, Online Price Dispersion Within and Between Seven European Countries *Faculty of Economics, University of Cambridge,* 2003, Sep

3. Xing Pan, Brian T. Ratchford and Venkatesh Shankar, Can Price Dispersion in Online Markets be Explained by Differences in E-tailor Service Quality? *Journal of the Academy of Marketing Science,* 2002, 30, 433

4. Rho, Youngwoo, Analyzing the Determinants of Housing Price: Regional Comparison within Seoul Korea. Unpublished manuscript, Vanderbilt University, Department of Economics, 2009

5. Judith D. Singer, Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models. *Journal of Educational and Behavioral Statistics.* 1998, 24 (4): 323-355

6. Kenneth H Brown and Bulent Uyar, A Hierarchical Linear Model Approach for Assessing the Effects of House and Neighborhood Characteristics on Housing Prices. *Journal of Real Estate Practice and Education,* 2004, 7 (1): 15

7. Bulent Uyar and Kenneth H. Brown, Neighborhood Affluence, School-Achievement Scores, and Housing Prices: Cross-Classified Hierarchies and HLM. *Journal of Housing Research*, 2007, 16, 2

8. Stephen W. Raudenbush and Anthony S. Bryk, Hierarchical Linear Models-Applications and Data Analysis Methods, Second edition, Sage Publications, Inc., 2002

9. Bradbury, Katharine L., Christopher J. Mayer, and Karl E. Case, Property Tax Limits and Local Fiscal Behavior: Did Massachusetts Towns Spend Too Little on Town Services under Proposition 2 ½, *Mimeograph, Federal Reserve Bank of Boston*. 1997

10. Downes, T.A. and J.E.Zabel, The Impact of School Characteristics on House Prices: Chicago 1987-1991, *Journal of Urban Economics,* 2002, 52, 1-25