ESSAYS ON SUPPLY CHAIN AND HEALTHCARE ANALYTICS

BY YIJUN WANG

A Dissertation submitted to the

Graduate School – Newark

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Ph.D in Management

Supply Chain Management Major

written under the direction of

Dr. Yao Zhao and Dr. Lei Lei

and approved by

Newark, New Jersey

April 2017

ABSTRACTS OF DISSERTATION

Essays on Supply Chain and Healthcare Analytics

By Yijun Wang

Dissertation Directors: Dr. Yao Zhao and Dr. Lei Lei

Abundant and real-time data from manufacturing and healthcare industries opens up opportunities previously unimaginable to capture fast changing customer bases, identify new sources of profit, and devise personalized medicine. This thesis consists of three essays: the first essay develops a machine learning algorithm based on a novel pharmacokinetic pharmacodynamic model for inpatients under warfarin therapy to make personalized predictions on blood clotting times on a rolling horizon. The second essay develops a set of methodologies to detect patterns in the online sales data of a children-shoes manufacturer in China to predict price elasticity. The third essay solves a challenging production scheduling problem in a three-stage supply chain.

Specifically, the first essay designs a pharmacokinetic pharmacodynamic model to make individualized and adaptive international normalized ratio (INR) predictions for warfarin inpatients in changing clinical status. We tested a new model on 60 inpatients at Columbia Medical Center. The model personalizes four submodels and

minimizes the number of parameters to be estimated. Prediction accuracy was assessed by prediction error, absolute prediction error and percentage absolute prediction error. The INRs (International Normalized Ratio, to measure blood clotting times) were accurately predicted 5 days into the future. Median prediction error: 0.01–0.12; median absolute prediction error: 0.17–0.5 and median percentage absolute prediction error: 9.85–26.06%. Patients exhibit interindividual and intertemporal variability. The model captures the variability and provides accurate and personalized INR predictions.

In the second essay, we study the problem of pricing and promotion for an online seller of children shoes and try to estimate the price elasticity for different products at different seasons. By testing the various models to capture the seasonality, price elasticity and sales inertia, we may accurately forecast the sales lift for various price discounts, so as to provide a foundation for data-driven price / promotion optimization.

In the third essay, we study a three-stage supply chain, where the second stage refers to the operations of external contracted manufacturers. The external manufacturers each has multiple time windows in which the resources needed for the outsourced operation are available. Both internal and external operations are non-instantaneous, and the required processing times are subject to various resource constraints. The problem is to assign and sequence a given set of customer orders to both the internal and the external processes so the total tardiness in the order

fulfillment is minimized. We present mathematical models that define different variations of this problem, analyze the special cases that can be solved in polynomial times, and then develop heuristic algorithms that can be applied to quickly solve the problems with a reasonable quality. In particular, we show that a heuristic solution based on the 3D linear assignment algorithm has a potential to be an effective approach for this type of problems.

Acknowledgements

I owe the deepest of thanks to my advisor, Dr. Yao Zhao, whose dedication to supporting me has been unwavering. His brilliance is astounding, his course is a highlight of my experience, and his generosity of time and mind are second to none. To my dissertation committee, I express great thanks: to Dr. Lei Lei for her vast knowledge and insightful comments, to Dr. Lian Qi for mentoring me as both teacher and researcher, to Dr. ShengBin Wang for showing me how to make the most difficult of problems look easy.

I also extend special thanks to my family: to my husband, Can Chen, with whom I share a love that most think only exists in fairy tales and who unconditionally supported me throughout my education; to my little girl, Kelly Chen, who sacrificed hours of Mommy time so that I may type and think in peace and quiet; and to my parents, for being the most supportive in years of study and taking good care of Kelly.

I also thank all the professors who put their time into the Ph.D. program and make it such a rewarding experience. Special thanks to Dr. Lei Lei, who not only teach me how to be a strong woman and also to be a qualified and respectful mum and wife. She has a huge impact for me in the later working and personal life. Special thanks to Dr. Lian Qi, who brought me into SAP world. Without him, I would ever have a chance to know that I love teaching and how important to deliver messages to people effectively. Special thanks to Dr. Hui Xiong, whose encouragement and experience are much appreciated. And He taught me how to dance in the business world with an academic background. I was fortunate enough to cross their paths.

Finally, I would thank you to all my friends in US and China. Because of you, Dr. ShengBin Wang, JunMing Liu, Zhe Jian, Dr. Xin Xu, Dr. Hui Dong, Dr. Yang Song, Julie Wen, Xu Cai, Di Qiao, Yanting Li, XiaoMing Zhong, HuiJuan Zhao, Aubrew Lu, Dr. Qi Liu, Dr. Chang Tan, Dr. Francis Cai, Dr. LuZhuang Wang, XiaoWen, Fu, ChenChun Zhu, MinJun Yuan and SiCheng Wang, life in US and China were fantastic and I was so honored to have you as a friend, or a mentor to make me a better lady.

Dedications

In loving of my husband Can Chen, whose wisdom continues to guide me, whose accompanies makes me brave ever and after, and whose support has enabled me to pursue my dreams.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Abundant and real-time data from manufacturing and healthcare industries opens up opportunities previously unimaginable to capture fast changing customer bases, identify new sources of profit, and devise personalized medicine.

With the cooperation relationship to Columbia University Medical Center, we have the opportunity to do some interesting research work, by combining the mathematical methodology with the medical data analytics. Previously, professors from Columbia University Medical Center are more likely to use statistics to explain the mechanism, while we would like to try something new. Another personnel reason for this research was from my grandfather's death, due to the stroke. And warfarin turns out to be the primary medicine for the prevention of the recurrent stroke. Therefore, warfarin is the starting point of our research.

Warfarin has a long history of clinical use and is the most routinely prescribed oral anticoagulant. The INR is a standardized measure of blood clotting time for those taking warfarin and needs to be controlled within a narrow range typically between 2.0 and 3.0 to achieve anticoagulation. Numerous factors (e.g., comorbidities, diet, medications, metabolism, genetic differences), can affect individuals INR levels, which make warfarin dosing a challenging task for clinicians. The same dose of warfarin may result in very different INRs among different individuals (inter-individual variability). Patients with multiple co-existing comorbidities may be in a transient clinical status and exhibit strong inter-temporal variability; meaning that an individual’s INR response to warfarin could change considerably over time. No protocols for warfarin dosing have been uniformly accepted, resulting in a significant variation in clinical practice of dosing. Recent research has shown that computerized algorithms based on PK/PD models may outperform the manual approach and protocols used in practice in terms of efficacy and convenience. However, most studies on PK/PD models focus on patients in a relatively stable clinical state; much less information exists on how to predict INR and how to dose warfarin for inpatients with multiple co-existing comorbidities and in a transient clinical status. The purpose of this study was to develop a new PK/PD model to make individualized and adaptive predictions of INR measurements for cardiac clinical inpatients whose health status was in a state of flux.

The research results are promising and Essay 1 was successfully published in <Personal Medicine> in 2015.

Inspired with Dr. Lei Lei’s supply chain optimization and algorithms class, and also due to my family business background, Essays 2 and 3 were originated from the XiaoDingDang case, which is a kid-shoes brand ranked top 10 in China. The initial thought for Essay 2 was quite easy. XiaoDingDang would like to expand its online channels and did not want to damage the existing offline channels, therefore they should create some new products especially for online channel. The differences pricing strategies from offline channel confused them a lot. The online sales were so sensitive that even one Chinese Yuan changes would affect the revenue; besides, crucial competition from other online sellers and also non-stop marketing events will also play a role in customer's behavior.

In this research, we tried to establish dynamic pricing strategies considering the promotion events and price elasticity. Therefore, given certain products price, we can easily predict the sales of product so that the inventory could be better controlled. For the further impact, the production planning will be much accurate since the online sales information will give XiaoDingDang an instant overview on which product is more likely to sell a lot. By forecasting the sales, we were ready to do a deeper supply chain analytics in production planning.

In Essay 3, our focus was on the production planning, assuming that all the forecasted demand from Essay 2 are quite accurate. And the next step is to make sure that we can make enough shoes to sell under the time constraints. If we got a chance to visit a shoe-making manufacturer, you would find out that making a pair of shoes is much more complicated than you can image. The whole factory was labor-intensive; more than 500 people were working there in order to make more than 3 million pairs of shoes; the profit margin went down due to the rising materials, labor and legal costs. Our family business, on the behalf of most enterprises in China, had more than ever motivations to cut down the costs from any aspect to survive in this economy.

The scheduling problem seems to be appealing to any one in charge of production. However, after consulting with more than 20 big groups, evidences showed that almost no one was using automatic system to schedule the production planning. The main reasons for this are: 1) the planning was too customized that APS providers would not like to configure a suited system with a very reasonable price; 2) the people in charge of production planning were less likely to switch to an automatic system, in belief that he/she should have a better understanding/experiences on scheduling; 3) the machines might turn down and the workers would ask for leave, so that it's easier to recalculate by themselves, even though it's quite time consuming and probably the results were wrong; 4) some external contracts added in may cause the difficulties of the scheduling.

As one of the contributions to family's years of supports, we would like to make a simple version of production planning by models, for decision-maker's cross validation evidence. However, some assumptions were still applied and the model had room to improve in the future research.

Chapter 2. A rolling-horizon Pharmaco-Kinetic Pharmaco-Dynamic model for warfarin inpatients in transient clinical states

1. Introduction

Warfarin is the most routinely prescribed oral anticoagulant for the treatment and prevention of thromboembolic events [1-3]. The drug works by affecting the function of the coagulation cascade and decreasing the clotting ability of the blood [1-3]. Although warfarin is widely utilized in clinical practice, it can be challenging to manage. For example, it negatively interacts with many other common medications, such as antibiotics and over-the-counter medications (e.g. NSAIDS) [4-9]. Advancements in molecular genetics and technology have shown individual genetic variations, where particular polymorphisms in the genes of certain enzymes have been associated with altered sensitivity and metabolism of warfarin [10-16]. The International Normalized Ratio (INR) levels, a standardized measure of blood clotting time, of patients taking warfarin can also be significantly altered by changes in the dietary intake of vitamin K from foods, vitamins, or herbal products [17, 18]. Thus, warfarin has a variety of potentially dangerous side effects and is among the top drugs with the largest number of serious adverse event reports [19, 20].

To prevent such adverse side effects, patients taking warfarin are required to have regular INR testing. The INR needs to be controlled within a narrow range, typically between 2.0 and 3.0, to achieve anticoagulation [21]. A high INR may predispose some individuals to an increased risk of bleeding, while an INR below the therapeutic range may be insufficient to protect high-risk individuals against stroke and other thromboembolic events [22, 23].

Numerous factors can affect individuals' INR levels, which can make determining the need for increased or decreased warfarin challenging for the practitioner. These factors can include, mostly notably, variations in anticoagulation dosing and clinical protocols, individual genetic differences in metabolism of warfarin (which may not be known), significant chronic disease burdens (e.g., atrial fibrillation, heart failure, diabetes), and addition or deletion of other medications that might influence warfarin metabolism [24-30]. These factors coupled with a hospitalization can further alter an INR response to warfarin by exhibiting strong inter-temporal variability, meaning an individual's INR response to warfarin could change considerably over time, and inter-individual variability, meaning the same dose of warfarin may result in very different INRs among different individuals.

To ensure effective and safe warfarin dosing in clinical practice, an adaptive personalized medicine approach to care that will account for individual variability is needed. Warfarin dosing and INR prediction protocols and algorithms have been studied in the literature for initiation and maintenance of proper anticoagulation in patients under relatively stable clinical conditions [19, 22, 31-38], however limited literature exists on how best to optimize anticoagulation therapies for inpatients receiving intensive care and those who are in a transient clinical state. Regression models have been used to determine loading and maintenance doses, so as to improve time in therapeutic range [35-37]. Even after accounting for information on patients' specific genotypes (responders vs. non responders to warfarin), these regression-based models are not completely individualized but rather adopt a cross-sectional design, and make dose prediction based on association observed between the warfarin dose and INR changes in a particular cohort [12, 38].

Computerized algorithms based on PK/PD models have the potential of making personalized and adaptive INR predictions and dose suggestions for patients in a transient clinical state [39-43]. Several clinical trials demonstrate the effectiveness of PK/PD models for warfarin dosing in comparison to manual dosing practices [44, 45]. A few retrospective studies report on the accuracy of these algorithms in predicting INR and maintenance doses for warfarin patients [43, 46]. Most of these studies focus on patients in a relatively stable state (e.g., rehabilitation) where patients' initial drug responses can be used to predict maintenance dose or steady-state INR [43, 44, 46, 47]. Only a few studies examined patients with potentially unstable conditions [42, 44].

To examine inpatients whose clinical status may be quickly changing, we developed a PK/PD model that includes all sub-models of Holford [40] and personalized them by having parameters from each sub-model estimated by utilizing individual patients' data. Including all the sub-models and fully personalizing them (in comparison to partial personalization using population means [43, 46, 48]) provided more flexibility to capture the inter-individual and inter-temporal variability previously mentioned. Our refined model also has fewer parameters to be estimated (than previous work that include all the sub-models) reducing the amount of historical data needed for estimation.   Our model is also estimated on a rolling horizon (with dated history replaced by latest observations) using a Bayesian approach to capture the inter-temporal variation. The purpose of this study was to develop a new PK/PD model to make individualized and adaptive predictions of INR measurements for cardiac clinical inpatients whose health status was in a state of flux.

2. Methods

2.1. Participants

This was a retrospective cohort study conducted at New York Presbyterian Hospital (NYPH)/Columbia University Medical Center (CUMC), New York, US. The study protocol was reviewed and approved by the Institutional Review Boards at Columbia University, New York City, New York, and Rutgers, the State University of New Jersey, New Jersey.

This study collected data from a convenience sample of 64 inpatients cases (admissions), which were seen as part of the cardiac services at CUMC during 2012-2013. These patients were also followed by the anticoagulation clinic at CUMC upon discharge for their outpatient warfarin management.

Patients were enrolled consecutively. Eligible patients were those who were initiated on warfarin therapy for any indication as part of their usual clinical care by their providers. We screened 346 cases total and excluded those patients under 18 years of age and those cases with fewer than 6 INR measurements. We also excluded 4 cases, which have missing data and/or errors. We did *not* exclude patients with other multiple chronic diseases, using potential medications known to interfere with warfarin metabolism, (as we aimed to create a “real world” approach that could be applied to clinical practice in the future).

Subjects’ clinical data was collected in a de-identified manner via chart review and through the electronic medical record (EMR), including age, gender, self-reported race/ethnicity, documented indication for anticoagulation therapy, concomitant medications (including those known to affect the metabolism of warfarin, such as amiodarone), cardiac history and cardiac testing preformed as part of their clinical care. All INR results measured as a part of routine clinical care were recorded from the first day of warfarin therapy until subjects were discharged from the hospital. A therapeutic INR target range for anticoagulation was defined for each subject at admission (between 2.0-3.0) and any changes in or out of the target range were recorded, from the electronic medical record review. The warfarin dosing history for each patient, including the daily dose taken, adjustments made and the time of warfarin administration was collected on all subjects with at least 6 INR measurements.

2.2. The PK/PD model

The PK/PD model developed in this paper includes all 4 sub-models of [40]. Specifically, we consider a single compartmental model and assume 100% bioavailability of warfarin and instantaneous absorption [43]. The pharmacokinetics model for warfarin metabolism is standard [40],

$$\frac{dW(t)}{dt} = -\beta W(t),$$

where **Error! Reference source not found.** is the amount of warfarin in plasma compartment at time **Error! Reference source not found.** and β is the warfarin elimination rate.

To describe the dose-effect in the pharmacodynamics model, we use a hyperbolic tangent function,

$$F\big(W(t)\big) = 1 - Tanh\big(bW(t)\big)$$

rather than the Emax function [40-42, 46, 48] because hyperbolic tangent has similar mathematical properties but only requires one parameter, **Error! Reference source not found.** [43]. In contrast, Emax requires two parameters.

The physiological model for the vitamin K dependent clotting factors is standard [40],

$$\frac{dC(t)}{dt} = -\alpha C(t) + \gamma F\big(W(t)\big)$$

where **Error! Reference source not found.** is the amount of clotting factors at time **Error! Reference source not found.** normalized to ensure **Error! Reference source not found.** (thus **Error! Reference source not found.** is not a parameter to be estimated) and α is first-order elimination rate. To keep the model simple, we assume that warfarin in the plasma is immediately available to inhibit clotting factor synthesis and **Error! Reference source not found.** is the same for all clotting factors [42].

To describe the relationship between clotting factors and INR, we use the inverse functions,

$$INR(t) = a + \frac{A\text{-}a}{C(t)/C_{\infty}}$$

where **Error! Reference source not found.** and **Error! Reference source not found.** (**Error! Reference source not found.**) are patient specific parameters [42], **Error! Reference source not found.** is **Error! Reference source not found.** representing the amount of clotting factors in the steady-state in absence of warfarin. We choose inverse function because theoretically, INR can be arbitrarily high as plasma concentration of clotting factors approaches zero. This trend can be properly modeled by the inverse function.

In summary, our PK/PD model has 5 parameters **Error! Reference source not found.**, which are patient specific and must be estimated based on individual patients' historical INR and doses. Once these parameters are estimated, the model can be used to predict future INRs for any dose regimen starting from any initial state, **Error! Reference source not found.** and **Error! Reference source not found.**.

The parameters are estimated by a Bayesian method using weighted least squares regression [43, 46]. Specifically, let **Error! Reference source not found.** be the set of parameters, **Error! Reference source not found.** (**Error! Reference source not found.**) be the **Error! Reference source not found.**observed (predicted) INR, **Error! Reference source not found.** be the corresponding time, **Error! Reference source not found.** be the mean of the **Error! Reference source not found.** parameter, and **Error! Reference source not found.** be the number of INR observations used to fit the model, the posterior objective function or Bayes risk is given by,

$$Min_{\theta} \sum_{j=1}^{n} \frac{\left(y_j\text{-}\hat{y}(\theta, x_j)\right)^2}{\sigma_y^2} + \sum_{p=1}^{P} \frac{\left(\theta_p\text{-}\mu_p\right)^2}{\sigma_{\theta_p}^2},$$

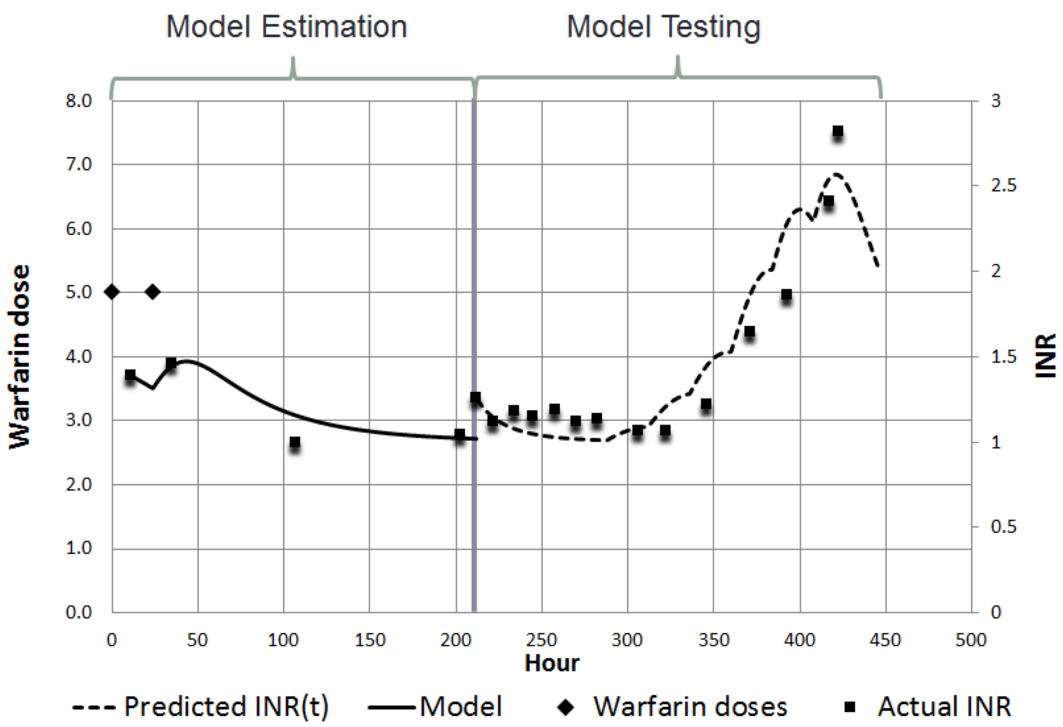
where **Error! Reference source not found.** is the variation of prediction error of INR, **Error! Reference source not found.** is the variation of the prediction error of the **Error! Reference source not found.** parameter. The first part in the sum represents prediction errors, and the second part measures the deviation of the parameters from the population means. The means of **Error! Reference source not found.** (0.02 **Error! Reference source not found.**, corresponding to a warfarin half-life of 36 hours) and **Error! Reference source not found.** (0.07) are obtained in the literature [43]. The mean of **Error! Reference source not found.** (0.05 **Error! Reference source not found.**, corresponding to a clotting factor half-life of 15 hours) is given by [40]. The mean of **Error! Reference source not found.** is set to 1 because a normal person without taking warfarin (**Error! Reference source not found.**) has an INR equal to 1 [44], and the mean of **Error! Reference source not found.** is set to 0.5. **Error! Reference source not found.** is chosen as a certain percentage of the mean for each parameter [43]. The model is implemented in Microsoft Visual Studio® and the Bayes estimators are found by a nested partitions method for global optimization [49]. This method has been validated and used in many research areas, such as medical treatment (e.g., radiation therapy), engineering and management [50]. The algorithm usually takes 1-2 minutes in computation to estimating one set of the parameters (Intel Core i5 CPU at 2.40 GHz, RAM 3 GB).

2.3. Parameter estimation

We estimate the parameters for our PK/PD model on a rolling-horizon basis, where we update the model upon each new INR response by estimating the parameters using the latest 5 INR responses. Figure 1 presents an example to demonstrate the procedure of model estimation and INR prediction. It shows the actual and predicted INRs over time for a case in our sample. The patient is a 48 year old female (in 2012) on warfarin therapy for atrial fibrillation.

Figure 2.1 A case study to demonstrate model estimation and to compare predicted vs. actual INRs.

The horizontal axis represents time (in hours), the right vertical axis is INR, and the left vertical axis is warfarin dose. The squares are actual INR responses, and the diamonds are warfarin doses. The solid line represents INR calculated by the model, which is estimated by the first 5 INR responses and all warfarin doses prior to the 5th INR response. The dash line represents the predicted INR ahead of the latest INR (i.e., the 5th INR).



For this patient, we first estimated parameters for our PK/PD model (the solid curve in Figure 2.1) using the initial five consecutive INR responses and all warfarin doses prior to the 5th INR response. Then we set the 5th INR as the starting INR (initial condition), and used the estimated PK/PD model to predict INR for the patient once every half-hour (the dash curve in Figure 2.1) in the next few days under the dose regimen recorded in the medical record. The dash curve does not connect to the solid curve because there is no guarantee that the model fits the 5th INR response perfectly. The dash curve shows that our predictions for INRs are sufficiently accurate for this patient up to about 200 hours ahead of the 5th INR. The prediction effectively captures INR changes, in particular the quick rising INR values, on this patient over the course of in hospital clinical care.

At each newly observed INR (the latest INR) beyond the first five, we estimated the model parameters again by using the latest five INR responses and all warfarin doses prior to the latest INR. INRs observed earlier than the latest five are dropped. For instance, if the latest INR is the 6th INR measure, then the 2nd through 6th INR values are used for model estimation but the 1st INR is dropped. To make INR predictions using the updated model, we reset the starting INR to be the latest INR.

2.4. Study design

For each patient case, we calculated three types of errors: the Prediction Errors (PEs – defined as the differences between the predicted INRs and the observed INRs), the Absolute Prediction Errors (APEs – defined as the absolute value of PEs), and the Percentage Absolute Prediction Errors (PAPEs – defined as the ratios between the APEs and the observed INRs) for at most five days ahead of the latest observed INR. We used PEs to measure the accuracy (bias) of the model, and APEs and PAPEs to evaluate the prediction precision [51]. We calculated the median and 95% confidence intervals (CIs) for the median of each type of errors, for each day up to five days into the future. These statistics allowed us to draw inference on the performance of the model in terms of accuracy and precision. In addition, we statistically compared PEs, APEs and PAPEs among different days into the future using the Kruskal-Wallis ANOVA test for medians.

To assess the importance of individualized treatments on these patients and the necessity of a fully personalized model, we studied the variation of the estimated values of all parameters for the entire cohort. For each parameter, a case may have multiple estimated values over time. If all patients are the same in their INR response to warfarin, we would expect that the mean estimated values for all cases will be close to the corresponding population mean of this parameter. To formally test this hypothesis, i.e., there is no inter-individual variability, for each parameter we calculated its mean estimated value for each case, and then conducted a t-test to see if these mean estimated values are significantly different from the corresponding population mean. To measure inter- temporal variability among patients, we calculated the coefficients of variation (CoV – standard deviation of the sample over sample mean) for each parameter estimated over time for each eligible case if the case has multiple estimated values. If each individual has the same INR response to warfarin over time, i.e., there is no inter-temporal variability, then we expect that these CoVs for each parameter have a mean zero. To formally test the existence of inter-temporal variability, we conducted a t-test for each parameter.

3. Results

3.1. Demographics

Table 2.1 Demographics of the cases (n=60) in the study. IQR stands for interquartile range.

Age (at admission) median (IQR)	60.5 (30.75)
18-65	33 (55%)
> 65	27 (45%)
Sex (n=60)	
Male	24 (40%)
Female	36 (60%)

Reason for Warfarin	
Atrial fibrillation	29 (48.33%)
Valve disease	10 (16.67%)
Vascular disease	17 (28.33%)
Stroke	1 (1.67%)
Others	3 (5%)
Ethnicity (n=60)	
White or White-Hispanic	7 (11.67%)
Black	1 (1.67%)
Asian	0 (0%)
Hispanic	21 (35%)
Others, not described*	31 (51.67%)
INR Target Range (at admission)	
1.5-2	2 (3.33%)
1.8-2.5	1 (1.67%)
2-2.2	1 (1.67%)
2-2.5	5 (8.33%)
2-3	46 (76.67%)
2.5-3.5	5 (8.33%)

The average number of INR readings in these cases is 12.7 (SD: 11; Range [6, 82]). The average number of warfarin doses received is 10.0 doses (SD: 7.6; Range [4, 60]). The median age (at admission) of these patients is 60.5 years, and 40% of subjects were males. The reason for warfarin administration includes atrial fibrillation (29 cases; 48%), valve disease (10 cases; 17%), vascular disease (17 cases; 28%), history of stroke and other reasons (4 cases; 7%). 46 cases (77%) have an

INR within the therapeutic target range of 2.0 to 3.0 (at admission); 9 cases (15%) have a sub-therapeutic INR below 2.0, and 5 cases (8%) had a high INR level above 3.0.

Our cohort also had multiple co-existing cardiovascular comorbidities, which further necessitated the need for warfarin. Of the 60 cases, 47 (78%) had hypertension, 20 (33%) had diabetes, 16 (27%) had coronary artery disease, and 15 (25%) had heart failure. This is not surprising given we acquired our data from our cardiovascular units.

3.2. Assessment of prediction accuracy

Table 2.2 show the median and C.I. for PE, the APE and the PAPE on future INR predictions for the PK/PD model on a fixed-horizon. Figure 2.2 shows the box-plots of PE. We made predictions for a maximum of five days into the future. The first column of the table and the caption of x-axis for Figure 2.2, "days into the future", indicate the day ahead of the latest observed INR response; that is, Day 1 means 0 to 24^{th} hour (including 24^{th} hour) ahead of the latest INR response; Day 2 means 24^{th} hour (excluding 24^{th} hour) to 48^{th} hour (including 48^{th} hour) ahead of the latest response and so on.

Table 2.2 Assessment of the prediction accuracy using median prediction error (PE = Predicted INR – Actual INR), median absolute prediction error (APE = |PE|), percentage absolute prediction error (PAPE = |PE|/observed INR).*

Days into the future	# of obs.	Prediction Error (PE)		Absolute PE		Percentage Absolute PE	
		Median	95% CI	Median	95% CI	Median	95% CI
1	381	0.01	(-0.01, 0.05)	0.17	(0.14, 0.19)	9.85%	(8.47%, 10.90%)
2	447	0.01	(-0.03, 0.05)	0.24	(0.21, 0.27)	13.45%	(12.09%, 14.82%)
3	387	0.04	(-0.01, 0.1)	0.34	(0.29, 0.42)	18.00%	(16.10%, 21.19%)
4	333	0.11	(0.02 , 0.19)	0.42	(0.34 , 0.51)	22.08%	(19.54%, 25.81%)
5	301	0.12	(0.05, 0.25)	0.50	(0.40, 0.60)	26.06%	(20.21%, 29.62%)

* In the column "Days into the future", 1 means 0 to 24th hour (including 24th hour) ahead of the latest actual INR, 2 means 24th hour (excluding 24th hour) to 48th hour (including 48th hour) ahead of the latest actual INR, and so on. The number of observations is greater than the number of inpatient cases because most cases have more than 6 INRs. The 95% CIs are for the median.

Figure 2.2 Boxplots of PE (Predicted INR – Actual INR) for the PK/PD model on a rolling horizon.

Each box represents the median, 75th and 25th percentile of the prediction error and the whiskers extend to cover 99 % of the data. The horizontal axis represents the number of days into the future.

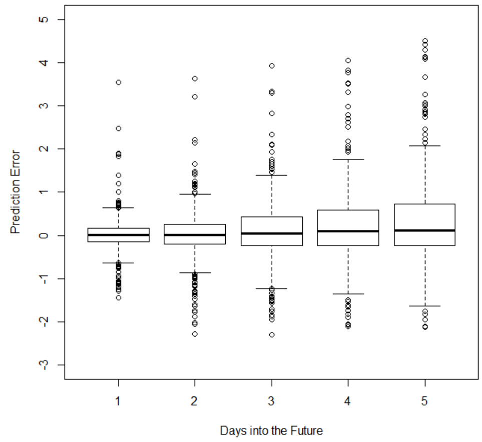


Table 2.2 shows that the predictions made by the PK/PD model on a rolling-horizon is unbiased within the first three days into the future because the median PE is not statistically different from zero. In the 4th and 5th days, the model slightly over-predicts the INR values. Table 2.2 presents the precision of the model using median APE and median PAPE, which shows that the INR predictions of the model are reasonably accurate for five days into the future. The model is particularly precise for the first two days with a median APE of no more than 0.25 (95% CI is [0.21, 0.27]) and a median PAPE of no more than 13.45% (95% CI is [12.09%, 14.82%]). The Kruskal-Wallis ANOVA tests on PEs (APEs and PAPEs) show that the model is more accurate for a shorter term prediction, i.e., the prediction of the model becomes more accurate the closer they are made from the last value into the future (p-value<0.001).

3.3. Inter-individual and inter-temporal variability

Table 2.3 shows the summary statistics of all estimated model parameters for all cases over time and the t-tests results for inter-individual and inter-temporal variability.

Except for two parameters, we found strong statistical evidence on inter-individual variability (p<0.05), supporting that individual patients do present very different INR responses to warfarin. For all of the five model parameters, individual CoVs over the course of our study are significantly different from zero, indicating strong inter-temporal variability, (i.e., patient INR response to warfarin does change over time).

Table 2.3 Summary statistics for the estimated values of all model parameters.

Parameter	**Warfarin half-life (hours)**	$\boldsymbol{b}$	**Clotting factor half-life (hours)**	$\boldsymbol{A}$	$\boldsymbol{a}$
Population Mean	36	0.1	15	1	0.5
Sample mean	37.74	0.10	15.25	1.01	0.46
p-value[1]	0.02	0.54	0.18	0.01	0.03
Mean sample CoV	6.65%	11.36%	5.55%	1.49%	17.25%
p-value[2]	<0.01	<0.01	<0.01	<0.01	<0.01

Note:
p-value[1] <0.05 suggests strong statistical evidence of inter-individual variability.
p-value[2] <0.05 suggests strong statistical evidence of inter-temporal variability.

4. Discussions

4.1. Our PK/PD model versus PK/PD models in the literature

A PK/PD model for INR response to warfarin typically has 4 sub-models [40]: (1) A pharmacokinetic model for the absorption, distribution, and elimination of warfarin. (2) A pharmacodynamic model to describe the impact of warfarin on clotting factors.

(3) A physiological model for the synthesis and degradation of clotting factors. (4) A model to link the activity of clotting factors to the prothrombin time (or equivalently, INR). Earlier studies attempted various mathematical equations for each of these four sub-models [41-43, 46, 48].

Our PK/PD model is novel in that it uses a new combination of the mathematical equations, for similar mathematical properties but fewer parameters. Specifically, we use hyperbolic tangent function for sub-model (2) and inverse function for sub-model (4), see Section 2 for further details. Our model includes all 4 sub-models but only has 5 parameters, with 1 for warfarin elimination, 1 for clotting factor elimination, 1 for dose sensitivity, and 2 for INR; all to be estimated for individual patients. Thus, the model is able to capture inter-individual and inter-temporal variability in all sub-models. In comparison, Vadher and Patterson proposed a PK/PD model with 11 parameters, 9 of which come from the literature and thus are identical for all patients, the other 2 parameters (warfarin elimination, dose sensitivity) are estimated for individual patients [43]. Wright and Duffull considered a PK/PD model with 8 parameters and 6 of them are individualized [46]. Wright and Duffull studied a model with 8 parameters [41], and the model of Hamberg et al. has 11 parameters [48].

The results in Section 3.3 imply that the inpatients in our sample not only have inter-individual variability but also exhibit a strong inter-temporal variability in all 4 sub-models of Holford NHG [40]. These results confirm the importance of an individualized PK/PD model particularly for those who were not in their stable state of health (as in our cohort). The rolling-horizon estimation method updates the model at each new INR observation by adding the latest INR but dropping the oldest data point; as such, it can capture the changing patient sensitivity to warfarin over time. Our PK/PD model uses fewer numbers of parameters as compared to previous models reported in the literature. Thus, our PK/PD model not only reduces the amount of data required for a reliable estimation but also helps to keep the model current.

Pharmacogenomics information can be included in PK/PD models [48]. However, this may not be necessary for the model proposed in this paper because it is already individualized to each patient [41]. That is, for each patient, we estimate a model based on the patient's historical doses and responses, and then use the model to make prediction for that patient. The model is also updated as the patient's condition progresses to capture the inter-temporal variability.

4.2. Predicting INR for patients in stable vs. transient clinical states

Previous studies of computerized algorithms of PK/PD models focus on patients in a stable health status. For instance, Vadher and Patterson predicted the long-term maintenance dose by the initial 4 INR responses [43], and Wright and Duffull predicted the long-term steady-state INR by the initial 0-6 INR responses [46]. Pasterkamp et al. considered a reduced model where one parameter can be updated after each visit for outpatients where INRs are monitored on a weekly or monthly basis [42]. In contrast, a majority of the inpatients in our study had fluctuating INRs because of their changing clinical status.

Wright and Duffull tested a variant of the PK/PD model of Hamberg et al. in a study of 55 warfarin patients on its predictive performance of the steady-state INR, defined to be the second of two subsequent INR measurements within 80–120 % of the target (usually 2.5) separated by at least 7 days [46]. Wright and Duffull also found that the model performs the best when the initial 6 consecutive INR responses are included, where 99% of the prediction errors are within (−1, 1) and 50% of them are within (−0.5, 0.5) [46]. Rather than the steady-state INR, we made predictions on INRs at any time in the next 5 days because the patients in our study exhibit strong inter-temporal variability and rarely reached steady-state during their stay in the hospital. This likely explains why our prediction errors are larger (especially on day 5); 95% of the prediction errors (for INRs five days ahead) are within (−1.7, 4.6), 50% of them are within (−0.2, 0.8), and some extreme outliers have PE>4 (see also Figure 2).

Our model may outperform previous ones on the prediction accuracy of INRs in the near future (in contrast to steady-state). For instance, Vadher and Patterson made an accurate prediction for the 4th INR based on the initial 3 with the median PAPE of 15.3% and a 95% CI of (10.9%, 17.7%) in a sample of 74 inpatients [43]. Based on the latest 5 INR responses, our model makes a more accurate prediction for the next INR (see Table 2) where the PAPE has a median of 9.85% and a CI of (8.55%, 10.93%). A few factors contribute to the improved predictive performance of our model and algorithm. Apart from the enhanced flexibility provided by individualizing all 4 sub-models and more historical data used, we use a global optimization algorithm to find the model parameters. In contrast to the commonly used Excel solver, our algorithm ensures that our fitted model provides the *best* possible fit to data [49].

INR levels are difficult to predict for warfarin patients, and it is particularly challenging to dose and obtain an INR within a target range of 2.0-3.0 for inpatients given their transient clinical state where their sensitivity to warfarin may change over time. Our results provide the evidence that the fully personalized PK/PD model developed in this paper, for which the parameters are estimated on a rolling-horizon, can provide reasonably accurate INR prediction for these patients up to five days into the future. This method is much better than manual protocols used by clinicians that are based on generic response tables, the alternative currently available for such unstable patients, because the latter typically can only predict INR one day ahead.

4.3. Limitations

Compliance with warfarin therapy and diet were not always available on study participants. While a majority of patients have only one case in this study, a few patients have multiple cases (up to six INR measures) as they were admitted multiple times to the hospital during the course of the study.

5. Conclusions

We have provided evidence that inpatients exhibit strong inter-individual and inter-temporal variability in their response to warfarin doses. Our novel PK/PD model combined with the rolling-horizon method can adapt to patients' changing clinical conditions over time, and provide an accurate and individualized prediction of INR levels 5 days into the future.

Given the requirement of at least 6 INR measurements over the course of treatment, the patients included in this study only represent those who had significant underlying cardiovascular risk factors and were admitted to the hospital for acute care. We plan to expand this research from INR prediction to warfarin prediction dosing in the future. We also plan to study inter-temporal variations and fully personalized models for outpatients receiving warfarin therapy, in whom INR levels are typically monitored less frequently than inpatients. This individualized approach holds the promise to improve anticoagulation management in the future.

Future Perspectives: Our computerized model provides a personalized rolling horizon approach that is able to adapt in real time to individual warfarin PK and PD variability. The application of this model will allow for better prediction of anticoagulation response outcomes and better dosing without the need for intensive sampling of drug concentrations (INR levels) and the use of static "one size fits all" dosing algorithms. Future research needs to focus on developing computerized dosing tools based on our proposed methods, and testing such tools via clinical trials on patients receiving warfarin treatment.

Chapter 3. XiaoDingDang Case——Dynamic Pricing

1. Background

XiaoDingDang is a Chinese children-shoe brand, which was well-known among the offline kid-shoe distributors in most of eastern cities. With the development of the E-Commerce in China in the latest 5 years, more and more brands are eager to increase their customer recognition not only through the traditional offline distributors, but also through the online markets via T-Mall (a B2C E-Commerce platform in Alibaba, NYSE: BABA, $103.1, as of 03/03/2017), JD.com (rival Alibaba, NASDAQ: JD, $30.93, as of 03/03/2017), and even Vipshop (a leading online discount retailer for brands in China, NYSE: VIPS, $12.92, as of 03/03/2017). In this paper, we only analyze the data from T-Mall.

Back to the year of 2011, the traditional way to sell shoes is straight forward. All the regional distributors gathered in one hotel and XiaoDingDang would like to show around 300 pairs of sample shoes. The regional distributors wrote down the numbers of the shoes they would like to order in the half year, marked with the specified colors. For some colors that not included in the sample shoes, some large distributors could ask to add more colors, if the new order exceeds the minimum production order quantity. There will be four shows per year to meet the demand in four seasons.

The disadvantages of the traditional selling ways are obvious: **1) disconnecting the relationship between the manufacturer and the end-users.** The information from the end-users will never be collected by the manufacturer promptly and accurately, such as the rationality of the pricing, the quality of shoes, the design defectives and the desire of repeating shopping, etc. **2) Losing control of the final retailing price.** The manufacturer did not know the selling pricing in various cities, so that it was making a meager profits through production, which is around 5%-7% of the revenue. If there were product returns with defectives, the profits margin would be much lower than 5%. **3) The long-term accounts receivables.** The regional distributors would delay the payment until they got paid from the end-users. Therefore, the longest payment term could be up to 240 days.

By trial of the online sells, XiaoDingDang hopes to combine the advantages from both online and offline sales. Offline sales will continue to be a big portion of the total sales, while online sales can help derive the market and customer insights much more effectively and efficiently. From the year 2012, XiaoDingDang jumped into the online sells market and up till now, it generated some valuable data sets, such as the sales order, customer information, the product information, etc. XiaoDingDang wondered that by all the evidence provided by the online data, is it possible to take the dynamic pricing policy to some products and how this strategy will take effect?

2. Data Processing

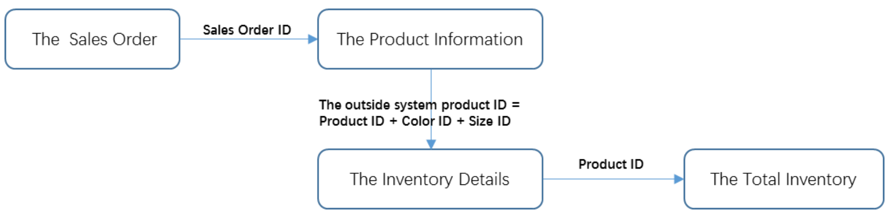
The data sets we got are from four categories in T-Mall: 1) the sales orders; 2) the product information, 3) Inventory details and 4) Total inventories. For more details, please check out the following table (Table 3.1).

Table 3.1 Illustration of the Data Sets

ID	Data Set	Contents	Notes
1	Sales Orders	Order Status, the buyer's/receiver's information, the logistics information, the payment information, etc.	1) 2013/1/1-2016/8/17 2) 387556 records
2	Product Information	The sales order ID, product name, price, quantity, the product property, order status, the product ID, etc.	1) 13327 product IDs 2) 2561 SKU
3	Inventory Details	Date, suppliers, warehouse, the product ID, product name, color ID, size ID, price, quantity, etc.	1) 2015/1/1-2016/8/17 2) 17257 records
4	Total Inventories	Product ID, suppliers, year, season, costs, the ending inventory quantity and revenue, etc.	1) 1005 records 2) On the date of 2016/7/31

By linking all the data together (Figure 3.1), we get a table with all the products, customers, suppliers, and sales orders connected.

Figure 3.1 The linkage between four data sets



When processing the data sets, we found out that the order status varied. The effective orders are 296640 out of 387556, about 76.5%. For all the following analysis, we mainly focused on the effective orders, ignoring the others.

Table 3.2 The Sales Order Status and Quantities

The Order Status	Order Quantities
Effective Orders	296640
Closing Orders	89244

Orders been shipped and waited for the buyer's confirmation	1433
Orders been paid and waited for the shipment	159

3. Exploratory Analysis

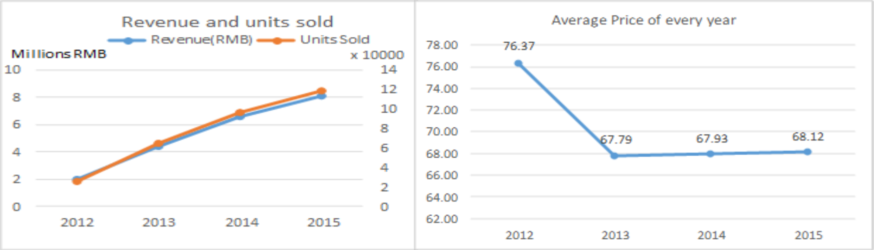
To have better understanding of the data, we explored the data by analyzing the trend, region and the discounts.

3.1. Trend Analysis

3.1.1 Revenue, Units Sold and Success Ratio from 2012 to 2015

From the year 2012 to 2015 (Figure 3.2), the total sales were increasing greatly, from around 2 million RMB to more than 8 million RMB. However, the average price of product were quite steadily, around 68 RMB, except for 76.37 RMB (2012). The reason for price deduction may be due to the severe competition in the shoe industry.

Figure 3.2 The Revenue, Units Sold and Average Price from 2012 to 2015



From Figure 3.3 and Figure 3.4, we saw significantly growth in both revenue, units sold and order success ratio from 2012 to 2015, which indicates the XiaoDingDang brand recognition is increasing among the online mass market.

Figure 3.3 The Success Ratio1 from 2012 to 2015

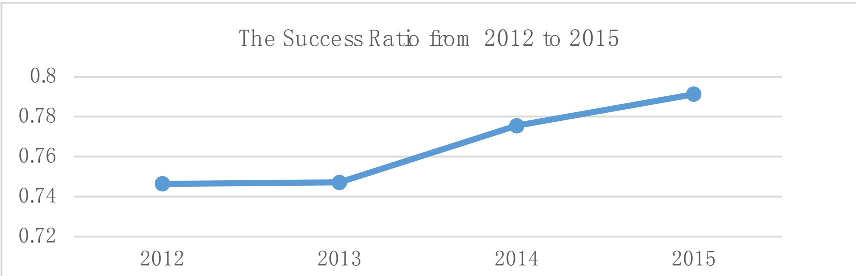
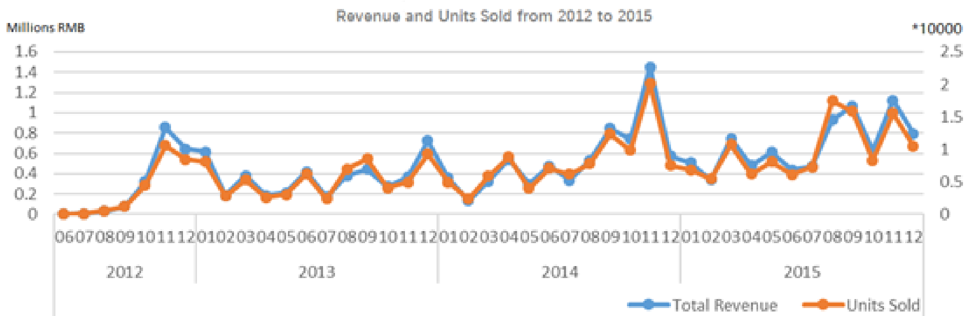


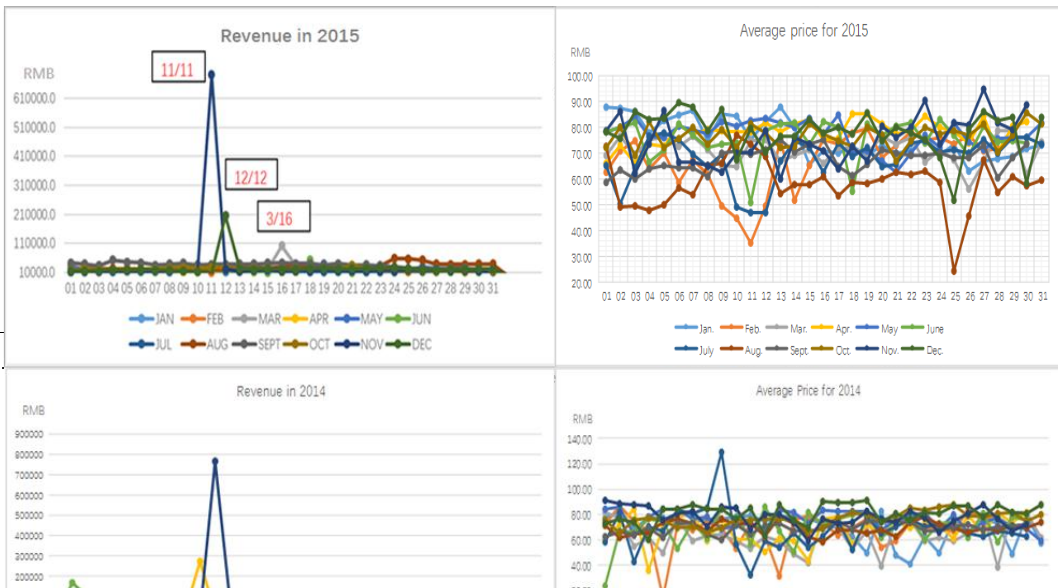
Figure 3.4 Revenue and Units Sold each month from 2012-2015



3.1.2 Average Price and Key Days' Sales in 2013, 2014, and 2015

Even though the yearly revenue increased steadily, it not means the average price for the product increase accordingly (Figure 3.5), which also indicates the profit margins for the company probably decreases in these three years.

Figure 3.5 Revenue and Average Price for each month from 2013-2015



1

From Table 3.3, we can see that the yearly activities are mainly in January, March, July or August, November and December. Some others insights include: 1) The "Double 11" on 11/11 ranks the top for all years around, taking up at least 5% (2013) revenue. For the year 2014 and 2015, "Double 11" is up to 11.6% and 8.5% respectively. 2) Except "Double 11" and "Double 12", the peak days may be different across years. 3) A few days' sales account for a significant percentage of annual sales, however no consistent pricing behavior across months and years.

Table 3.3 Key Days' Sales from 2013 to 2015

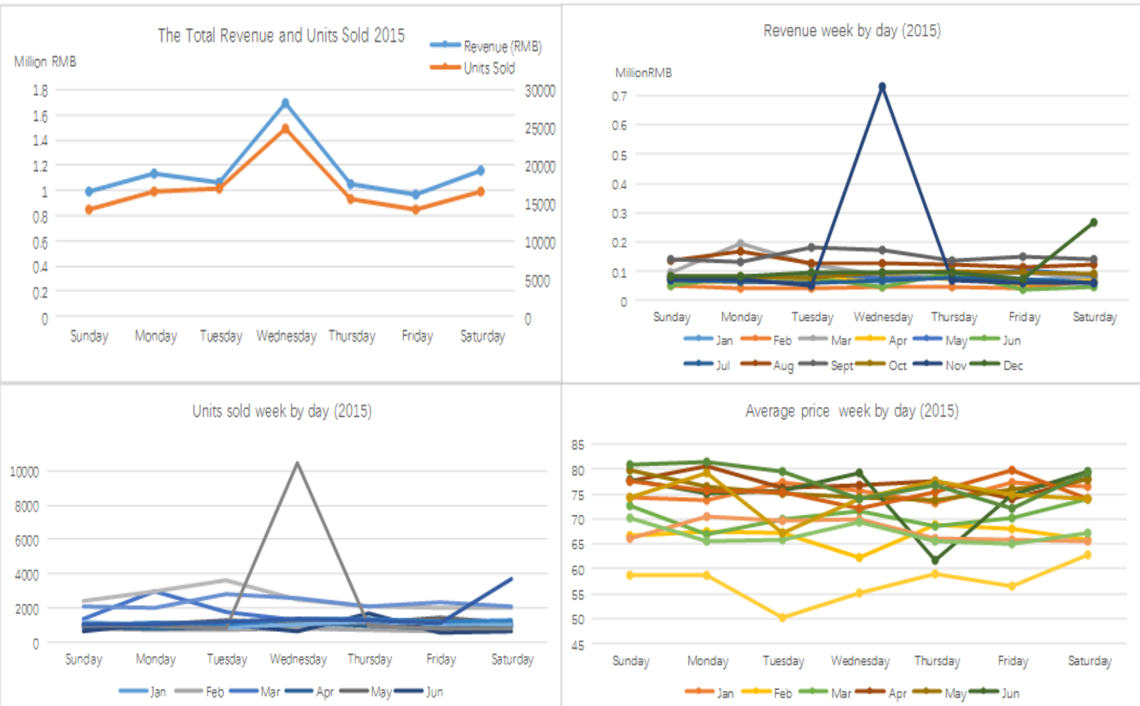
Year	Date	Revenue (RMB)	Units Sold	Revenue Ratio	Unit Sold Ratio
	03/16/15	101589.6	1537	1.25%	1.30%
	8/24/15	58126.7	991	0.71%	0.84%
	8/25/15	56957.8	2343	0.70%	1.98%
2015	**8/26/15**	52251.5	1151	0.64%	0.97%

	11/11/15	691918.8	9847	8.50%	8.31%
	12/12/15	205171.1	2908	2.52%	2.45%
	Total	**1166015**	**18777**	**14.33%**	**15.85%**
2014	**01/06/14**	106933.6	1463	1.6%	1.5%
	01/07/14	78020.5	1057	1.2%	1.1%
	04/10/14	272974.8	4582	4.1%	4.7%
	07/02/14	164733.6	2461	2.5%	2.5%
	07/03/14	95055.2	1443	1.4%	1.5%
	08/15/14	73855.6	1185	1.1%	1.2%
	11/11/14	765559.5	11167	11.6%	11.5%
	12/12/14	103410.4	1629	1.6%	1.7%
	Total	**1660543**	**24987**	**25.1%**	**25.8%**
2013	**01/08/13**	65823.49	823	1.5%	1.3%
	01/14/13	139091.9	1774	3.2%	2.8%
	06/03/13	101202.5	1446	2.3%	2.3%
	07/19/13	126351.5	2124	2.9%	3.3%
	08/09/13	132113.9	2656	3.0%	4.1%
	11/11/13	217725.4	2876	5.0%	4.5%
	12/02/13	294061	3676	6.7%	5.7%
	12/03/13	141284	1775	3.2%	2.8%
	12/12/13	41467	529	0.9%	0.8%
	Total	**1259120**	**17679 17679**	**28.70% 28.70%**	**27.60% 27.60%**

3.1.3 Revenue by Month, by Week, by Day and by Hour in 2013, 2014, 2015

In 2015, if we consider the sales by day, Wednesday is ranking the top of sales, because “Double 11” in 2015 was Wednesday. The average price was quite steady, except August with 10% lower than the normal price.

Figure 3.6 Revenue, Units Sold and Average Price by day in 2015



Clearly, because of the "Double 11", the 11th date is the most productive throughout the years (Figure 3.6 and 3.7). Besides, 2014 and 2015 are the years that XiaoDingDang spent much money on advertising brand, so that the sales are sky high. The shopping time preference is quite steady in Figure 8, showing that one big peak is at 10:00AM in the morning and the other small peak in the evening around 21:00. However, the order success ratio is not consistent with the shopping peaks. We figure out that the noon time has a relatively high success ratio, but the variance of the day is small, no more than 10% difference (Figure 3.8 and 3.9).

Figure 3.7 Revenue for each day of month from 2012-2015

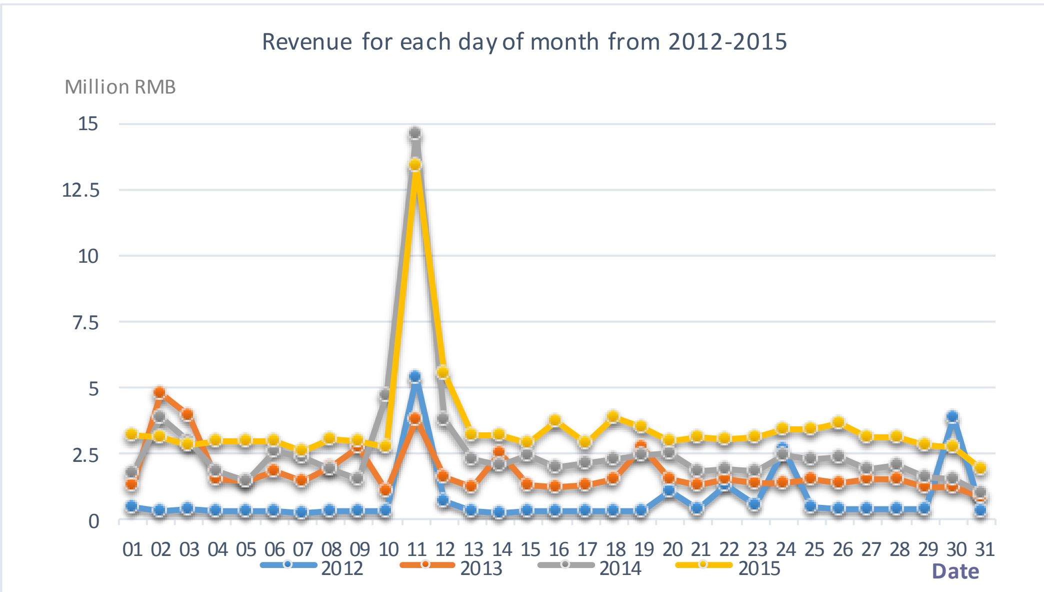


Figure 3.8 Units Sold by hour from 2012-2015

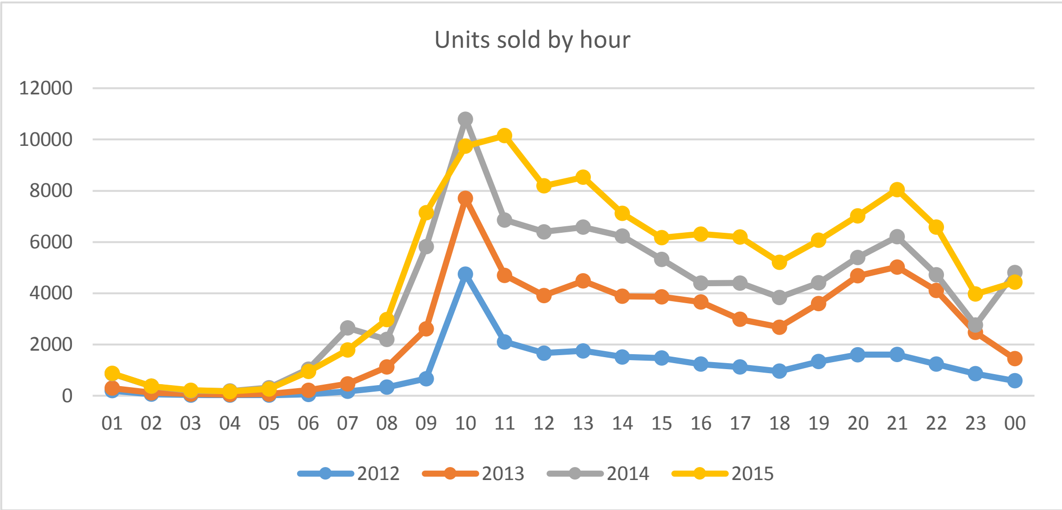
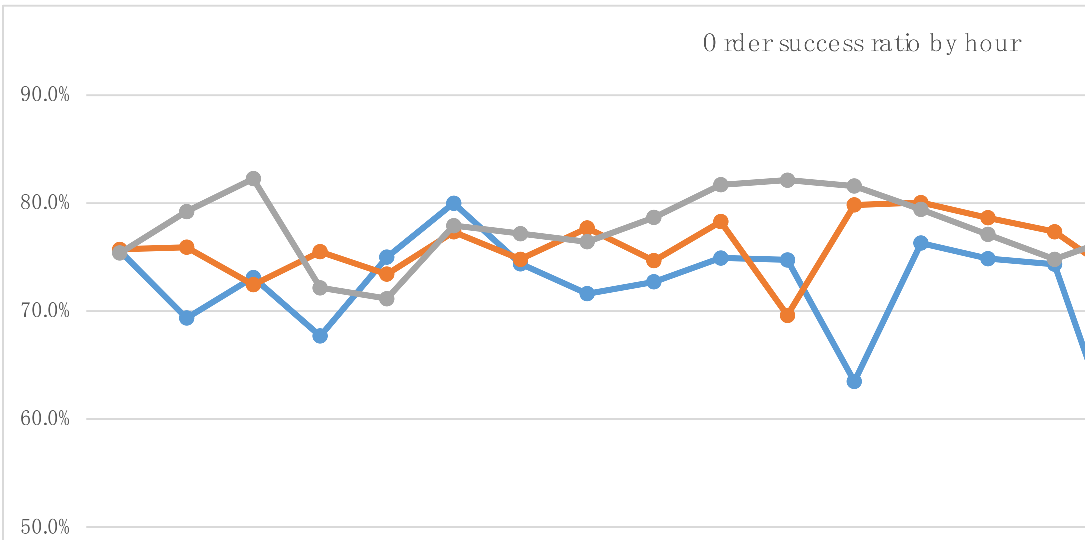


Figure 3.9 Order Success Ratio by hour from 2013-2015

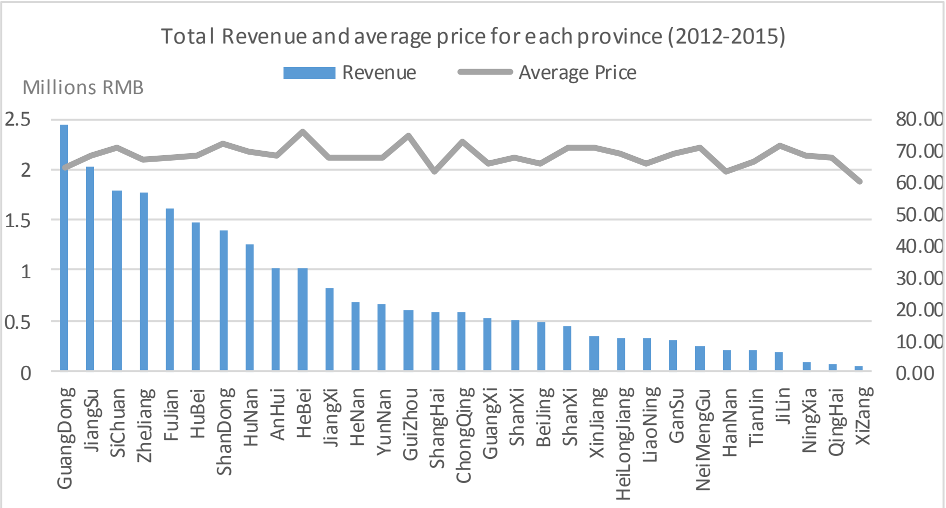


3.2. Regional Analysis

3.2.1 Total Revenue and average price for each province (2012-2015)

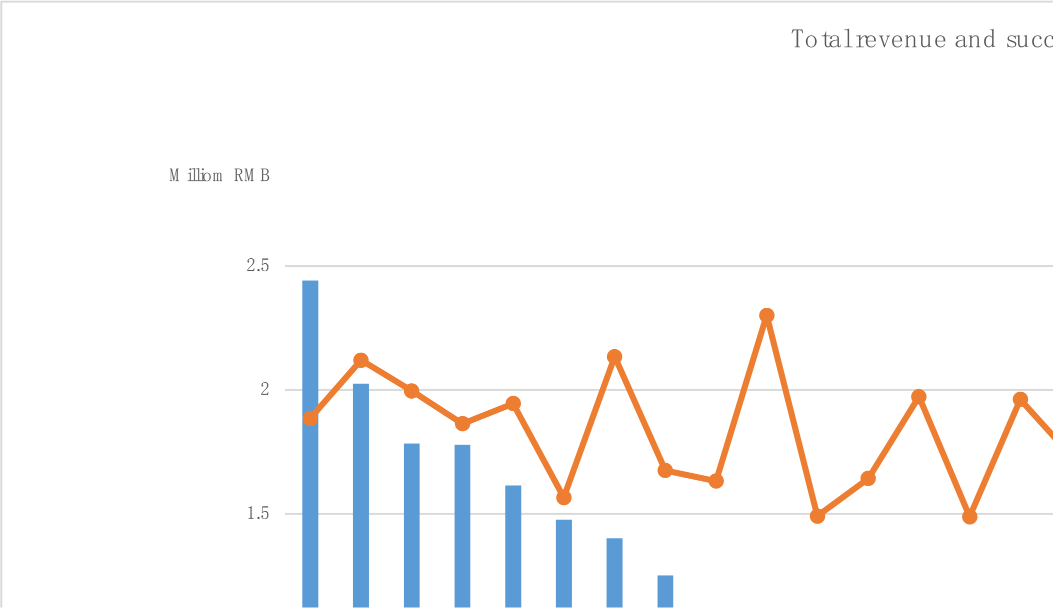
Geographically, XiaoDingDang sells shoes more than 30 provinces in the Greater China both online and offline. For some provinces, it's hard to cover by the physical distributors, but for the online shoppers, there is no geographic constraints. It, to some extents, increases the coverage of the sales. From Figure 3.10, GuangDong, JiangSu, SiChuan, ZheJiang and FuJiang are the top sales provinces. It is interesting to note that lower prices will not definitely lead to the higher sales. Comparing HeBei with ShangHai, we can tell that the higher average price in HeBei (￥78) is with relatively higher sales, around 150% higher sales than that in ShangHai.

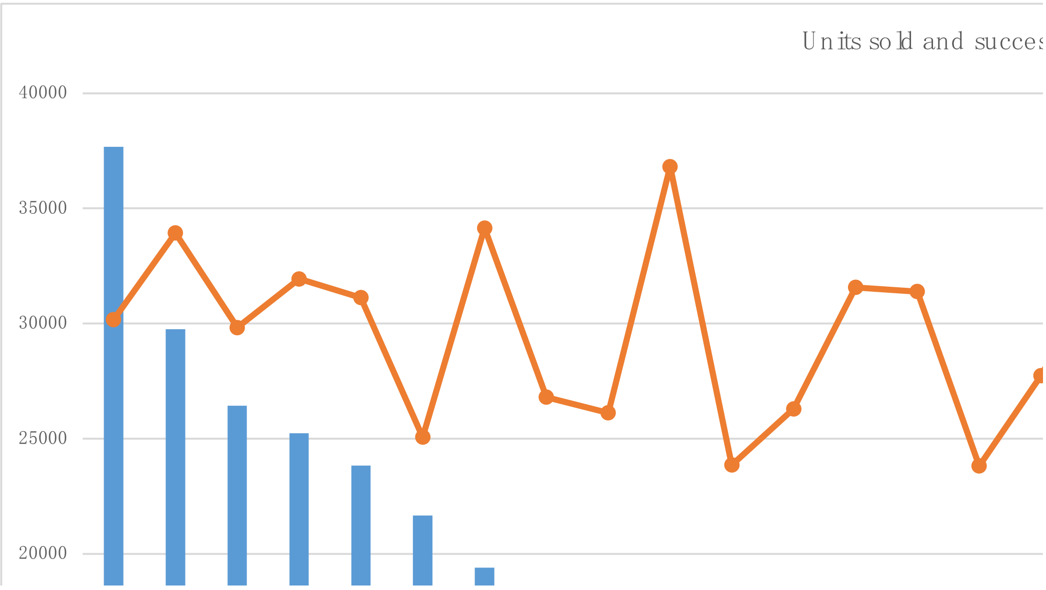
Figure 3.10 Total Revenue and average price for each province (2012-2015)



The success ratio varies across different regions, and the lowest ratios is at Ningxia, with less than 0.7 (Figure 3.11). The top five provinces for the total revenue and units sold are almost consistent, except the swift for SiChuan and ZheJiang. HeiBei and HaiNan and ShanDong are with the highest success ratio, almost around 0.79, while ShanXi, GanSu and NingXia, these three Northwest provinces are with quite low success ratio, around 0.73.

Figure 3.11 The Total Revenue, Units Sold and Success Ratio for Each Province





3.2.2 Total Revenue and average price for each province (2012-2015)

When we compare the average price and success ratio for city (Figure 3.12), we see that the correlation is 0.04 for all cities (across province). However, for ShanDong and HeBei provinces, they are positively correlated (Figure 3.13 and Figure 3.14), while it's negatively correlated in LiaoNing Province (Figure 3.15).

Figure 3.12 The Average Price and Success Ratio for City

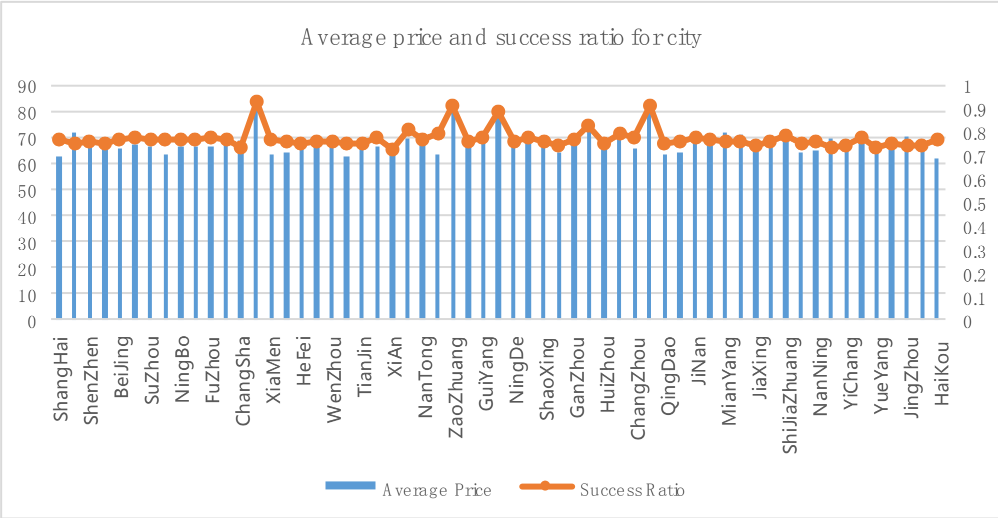


Figure 3.13 The Average Price and Success Ratio for City for ShanDong Province

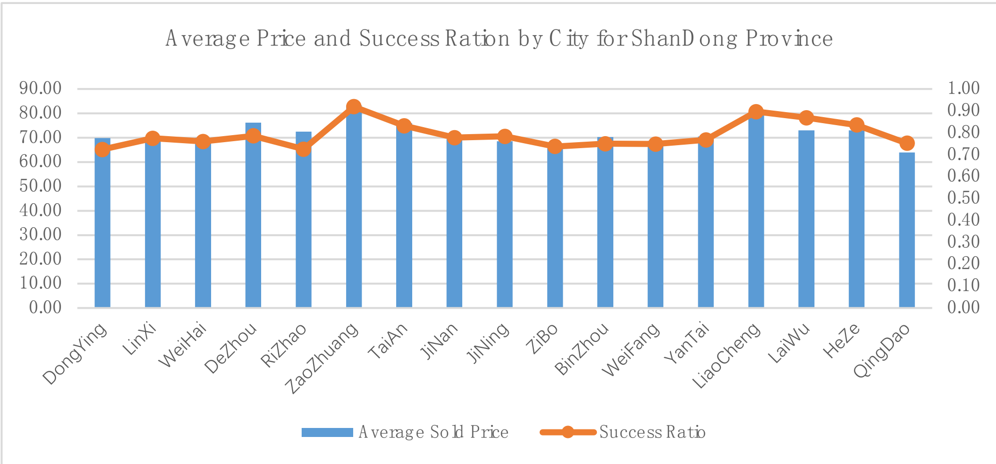


Table 3.4 The Regression Model for ShanDong Province

	Estimate	Standard Variance	T-Value	Pr > |t|
Intercept	18.6374883	10.3707434	1.80	0.0912
ratio	67.2677220	12.9673449	5.19	<.0001

Table 3.4, 3.5 and 3.6 are showing results of the linear regression models running in ShanDong, HeBei and LiaoNing, respectively. Both ShanDong and LiaoNing are with the highest prediction accuracy.

Figure 3.14 The Average Price and Success Ratio for City for HeBei Province

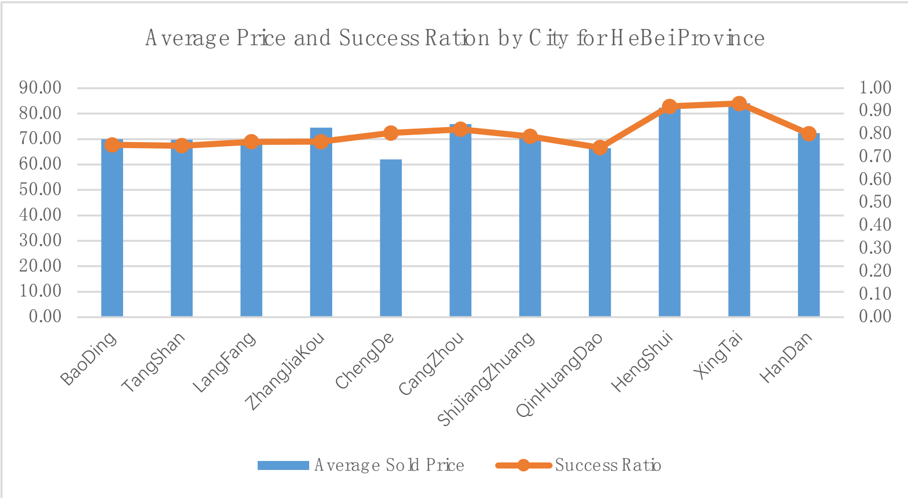


Table 3.5 The Regression Model for HeBei Province

	Estimate	Standard Variance	T-Value	Pr > |t|
Intercept	8.6283906	6.48187900	1.33	0.2127
ratio	79.4514747	7.56132237	10.51	<.0001

Figure 3.15 The Average Price and Success Ratio for City for LiaoNing Province

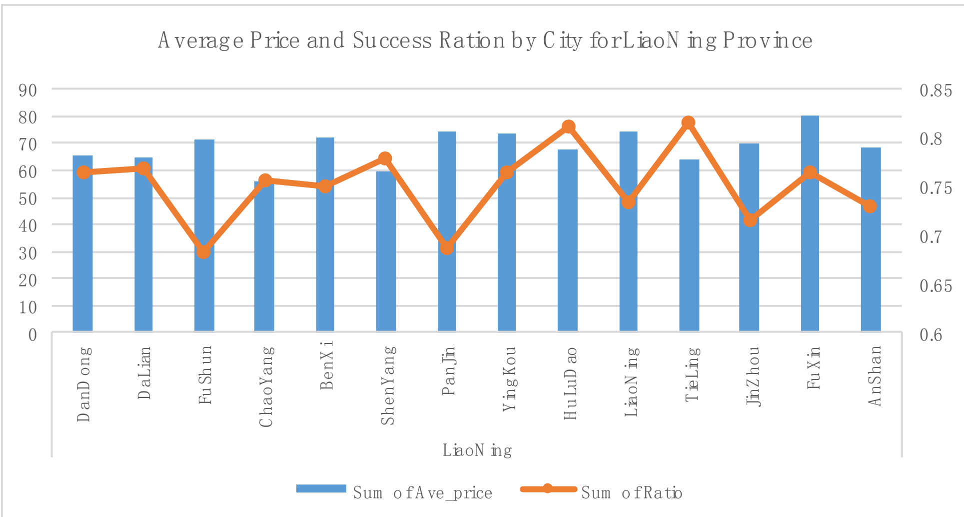
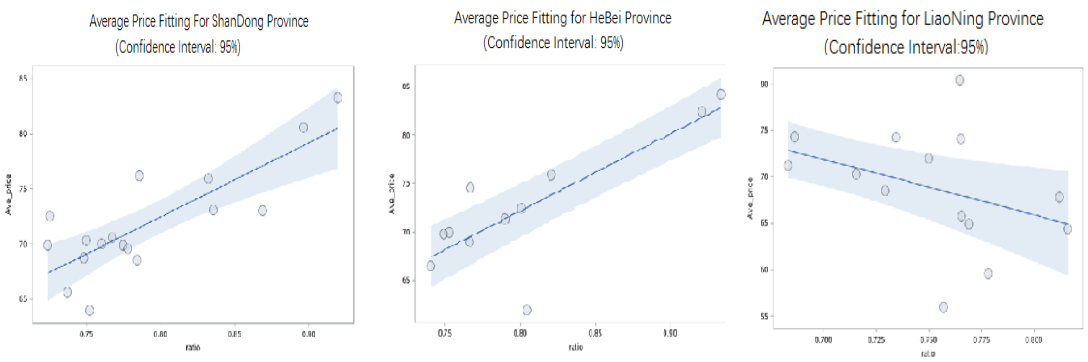


Table 3.6 The Regression Model for LiaoNing Province

	Estimate	Standard Variance	T-Value	Pr > |t|
Intercept	113.497117	13.7707159	8.24	<.0001
ratio	-59.508888	19.5100950	-3.05	0.0093

From Figure 3.16, there is a strong positive relationship in ShanDong and HeBei provinces, while LiaoNing is strongly negative under confidence level of 95%.

Figure 3.16 The Average Price Fitting for ShanDong, HeBei and LiaoNing Province (CI:95%)

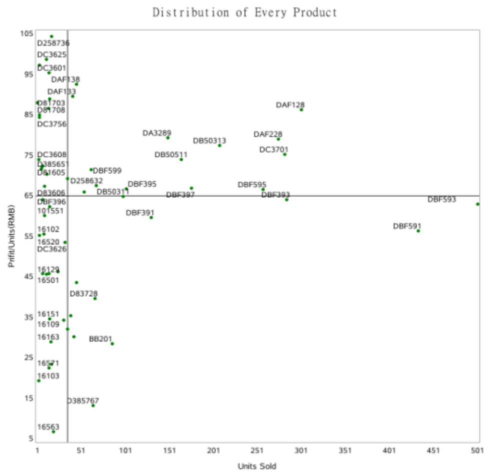


3.3. Products Analysis

3.3.1 SKU Rationalization

The fashion of kid’s shoes is not only fast, but also changeable, so that tons of inventories and long-tailed SKUs are inevitable. In Figure 3.17 and Table 3.7, by dividing Profits/Units (65) and Units Sold (35) into four categories, we will see 1) **Junk**, 2) **Diversity**, 3) **Commodity**, and 4) **Cash Cow** (Most desired).

Figure 3.17 The Distribution of Every Product



Data resource: Xdd

For the Junk products, both Profits/Units is less 65 and the Units Sold is less than 35, which means the more we sell, the more we lose. This kind of product, we should try to eliminate in the next season to reduce the loss. For the Categories in Diversity and Commodity, if we do not control the Profits/ Units or Units Sold well, it may lead to lose, but also may result in a small profits margin, depending on how we run the opration. For the Cash Cow category, those products are the main focus of XiaoDingDang, which not only bring us more profits, but also have high sales volumns.

Dynamically classifying the products by season will help the management to adjust the marketing strategies and pricing tactics. Prompt response and opinions from the market and end-users will increase the success of R&D and dramatically decrease the management and operation costs. By spending less ads fees, we can target more precise end-users with right products.

Table 3.7 The Four Categories of all products

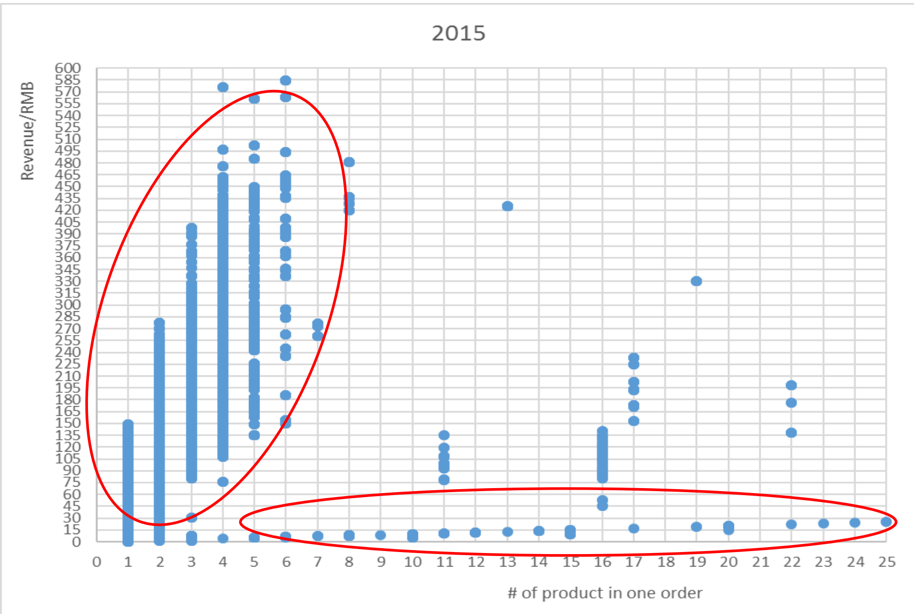
Category	**Profit/Units**	**Units Sold**	**Product ID**
Junk	<65	<35	DC3626, DA60109, DB3362, 16563, 16582 ,16163, 16151, DBF396, DBF597,16501,16571, DC3601, D81703,

			DB3361, 16129, DC3625, 101551, D83606,16102, D258638, D81605, DC3726, DC3653, D385651, 16520, 16103
Diversity	>=65	<35	D258736, 16163, 16151, DBF396, DBF597, 16501, 16571, DC3601, D81703, DB3361, 16129, DC3625, 101551, D83606, 16102, D258638, D81605, DC3726, DC3653, D385651, D81708, DC3756, DC3788, 16103, DC3608, DB60076
Commodity	<65	>35	DC28603, DB5001, DBF591, DBF391, BB201, D83728, D385767, DB3367, DA60209, 16509, 16109
Cash Cow	>=65	>35	DAF275, DAF175, D83725, D81716, DBF592, DBF593, DAF128, DBF393, DC3701, DAF228, DBF595, DB50313, DBF397, DB50511, DA3289, DBF395, DB50311, D258632, DBF599, DB50320, DAF138, DAF133, DC3708

3.3.2 The Volume Discount

From Figure 3.18, there are two different categories of products / markets. There are economies of scale (volume discount) but not very significant. For the right-handed red circle, the more volumes you get, the more differences of the shipping fees, which means the payment equals to the units sold.

Figure 3.18 Total Revenue with the Number of Product in One Order



3.3.3 The Pareto Analysis (80/20 rules)

From the sales analysis, we are so surprised to find out that it's quite consistent with the Pareto Analysis. Figure 20 shows the sales of Top 15 products in 2015 accounted for 58% of total revenue, others accounted for 42%.

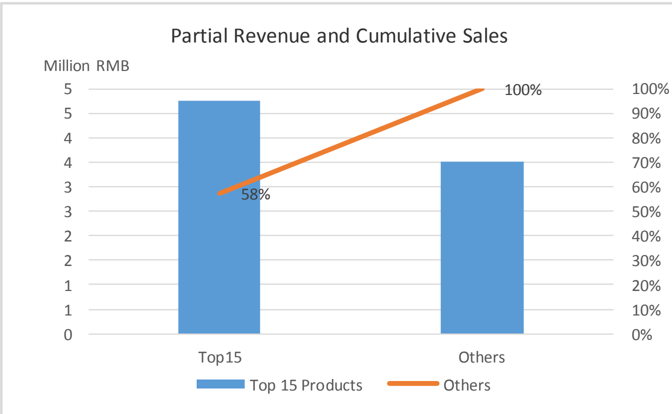
Another interesting finding (Table 3.8) is that girl's parents are spending more money on shoes than boy's parents. Therefore, the unit price for the girl's shoes is at least 10% higher than the boy's shoes. As one Chinese old saying goes: wealthily raised the girl but restricted with boys.

Table 3.8 The Top 17 Products Sales Revenue

Product ID	*Revenue*	*Units Sold*	*Ave. Price*	*Gender*
DC28603	1078670.5	14746	73.2	B
DW51701	1025888.2	16186	63.4	G
DB50303	353001.4	4288	82.3	G
DB50503	316919.2	3886	81.6	G
D585771	269141.8	3117	86.3	B
DA5190	234972.5	2525	93.1	G
D581725	210068.7	4416	47.6	B
D585671	185312.4	2125	87.2	B
DA5290	181061.3	1968	92.0	G
DB5001	175809.0	5286	33.3	M
DC50603	173304.0	1795	96.5	G
D581626	150787.9	3798	39.7	B
DC50607	144407.9	1566	92.2	G
DC50703	138182.0	1468	94.1	G
DA5051	123952.0	1296	95.6	G
DB50313	95327.3	1197	79.6	G
DC50707	94230.0	1049	89.8	B
Total	**4761479**	**68466**		

Gender	Revenue (RMB)	Units Sold	Rev. Rate	Sold Rate
G	2691688.41	34978	56.5%	51.1%
B	1893981.3	28202	39.8%	41.2%
M	175809	5286	3.7%	7.7%
Total	4761478.71	68466		

Figure 3.19 Partial Revenue and Cumulative Sales



From Figures 3.20, 3.21and 3.22, we could see that 1) the top two SKUs account for 25.45% of the sales; 2) a very long-tailed bullwhip effect; 3) the control of the Top 15 products may lead to a huge success with less marketing efforts, less advertising fees, less operation cost, less labors work, etc. 4) getting rid of the dead products is one of the most urgent work to do, by either markdowns or sell to the 3rd channel party.

Figure 3.20 Revenue and Cumulative Percent (2015)

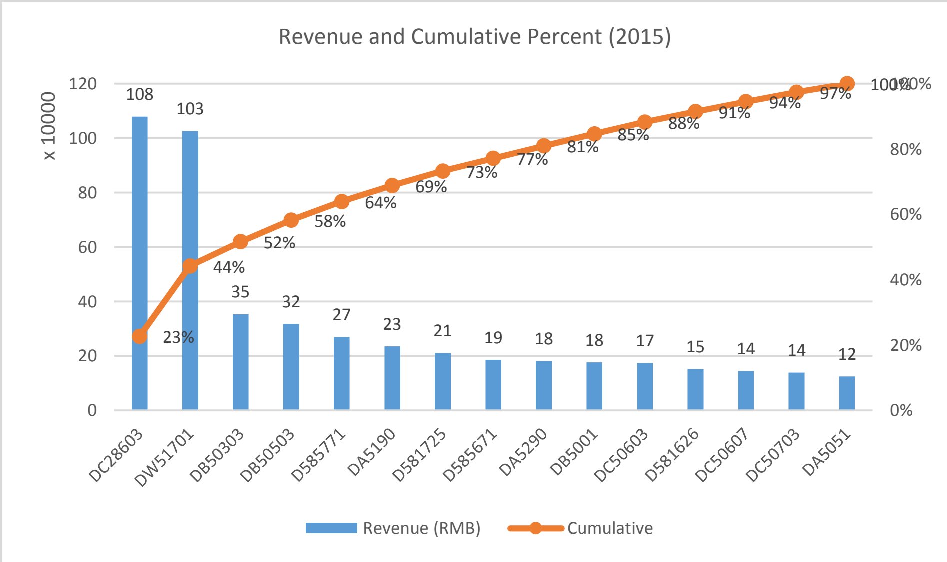


Figure 3.21 The distribution of revenue on all products (2015)

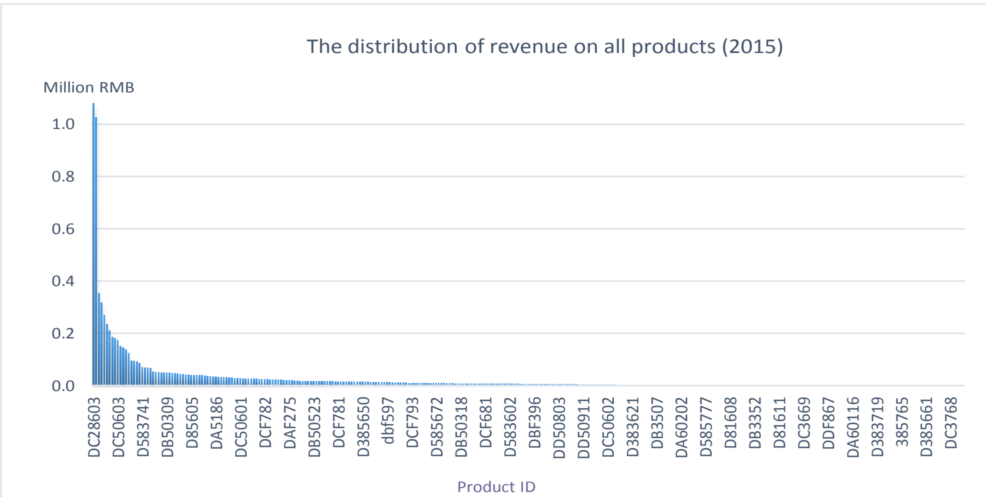
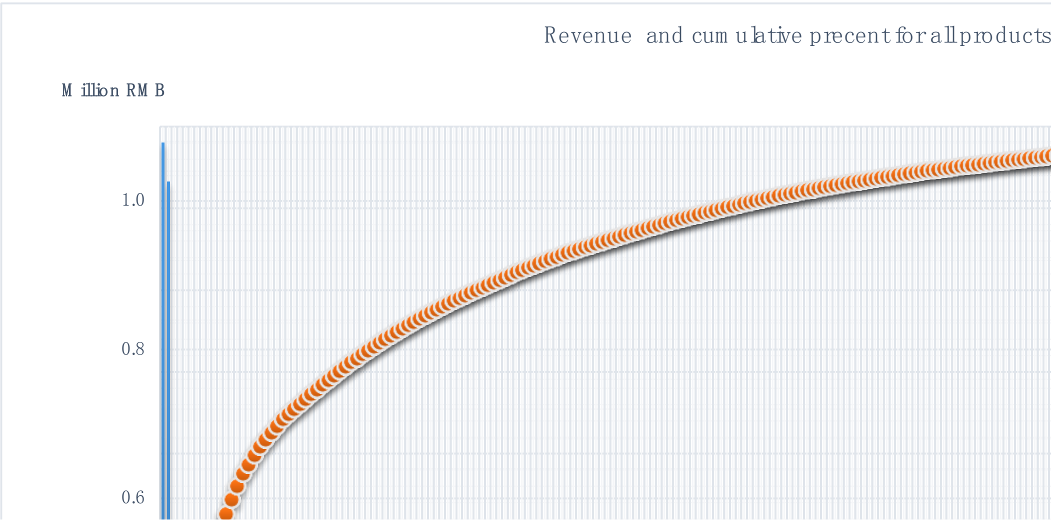
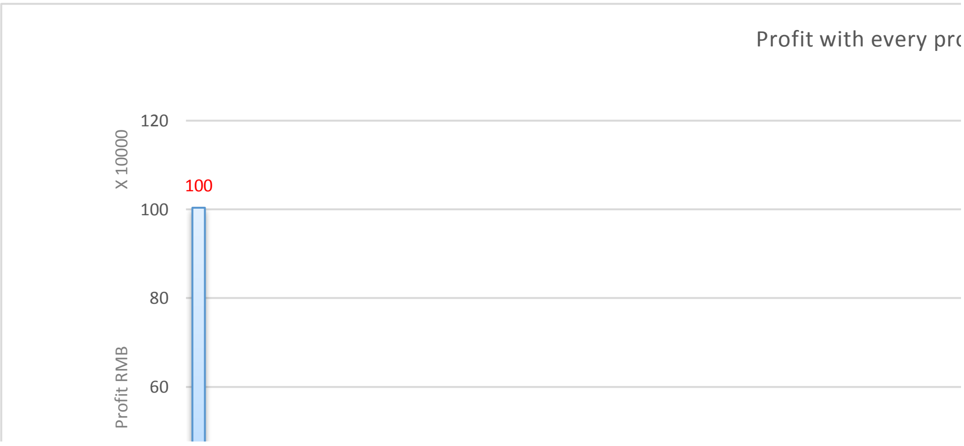


Figure 3.22 Revenue and Cumulative percent for all products (2015)



After the combination of products with purchase, we totally get 68 product ID from the all the data. The Product DC28603 generates more than one million of profits, which is far away from the second place.

Figure 3.23 Profit with every product (2012-2016)



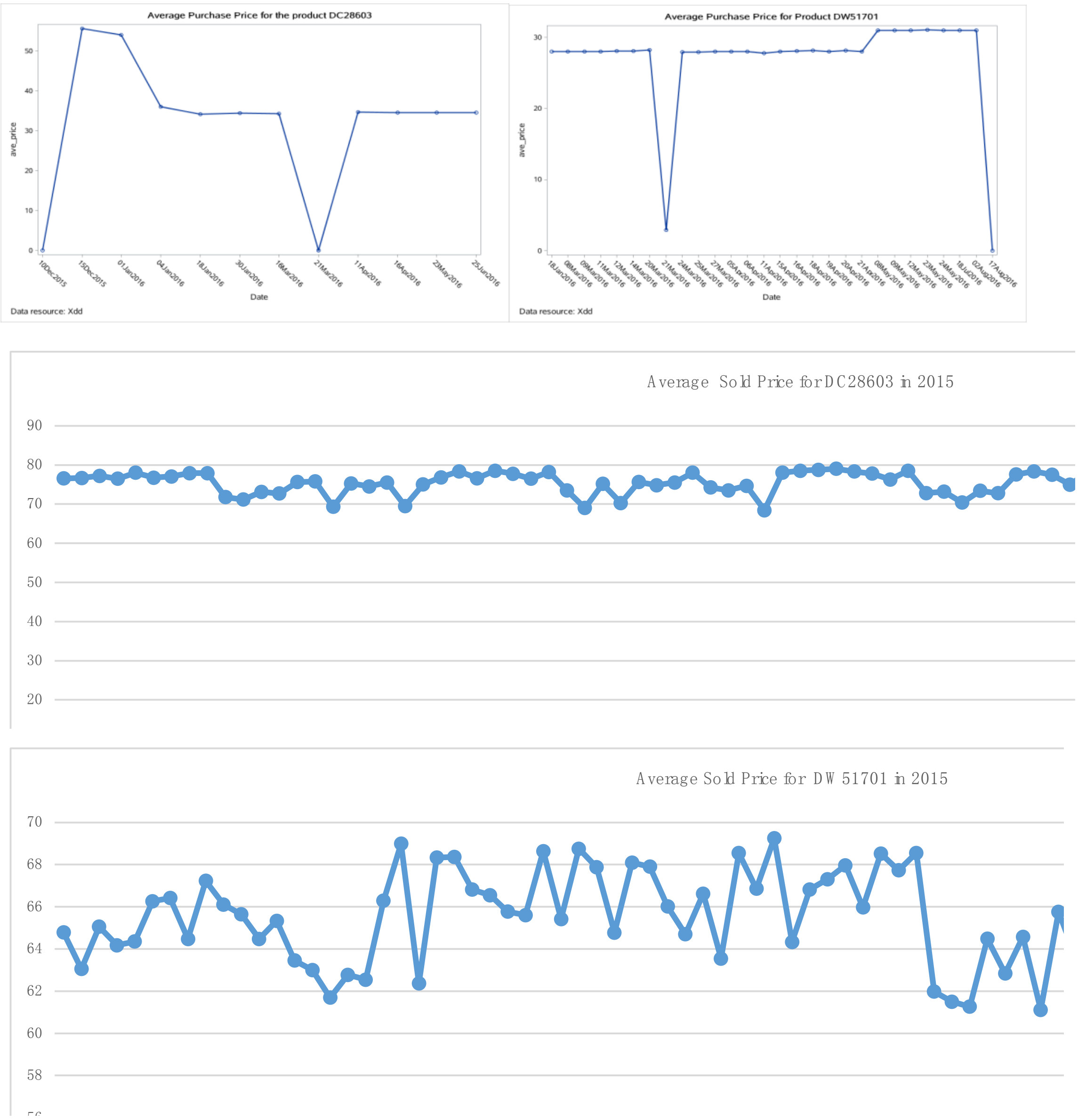
4. The Mathematical Model

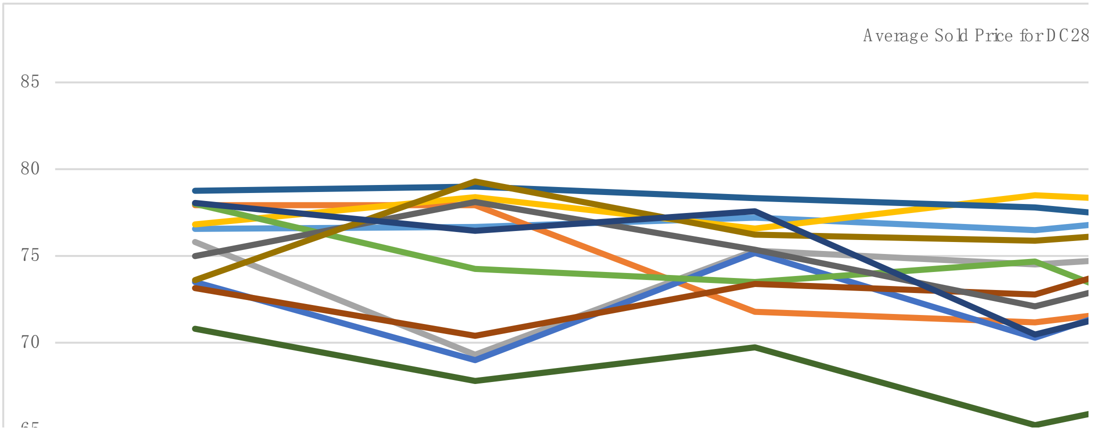
4.1. Two Special Products Overview

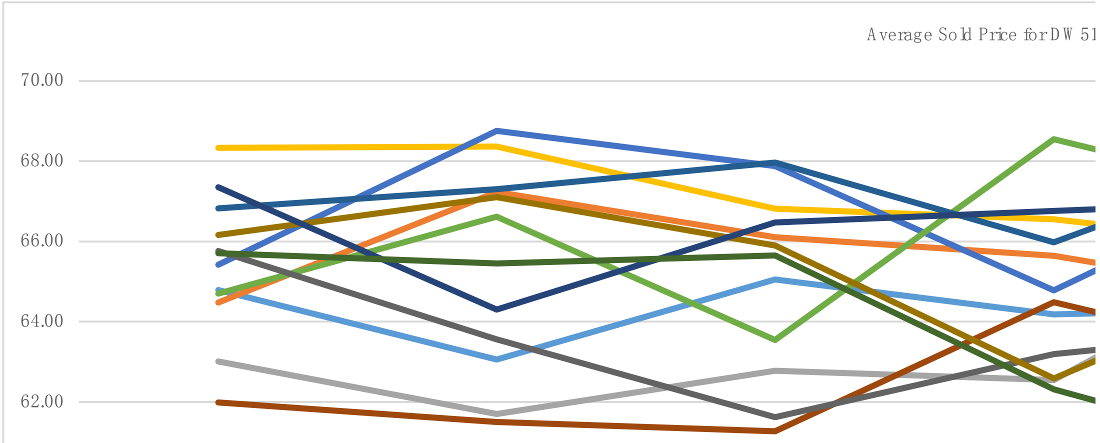
From the exploratory analysis in Section 3, we could see that the average unit price varies depends on the different holidays and marketing events. Therefore, we wonder which is the optimal price for some products under holidays or non-holidays? How much exactly we can sell based on the certain price policies?

To explain these two questions, we first choose two products: one boy's shoe (DC28603) and one girl's shoes (DW51701). A quick overview for these two products are followed in Figure 3.24.

Figure 3.24 Average Price Comparison for products DC28603 and DW51701







From Figure 3.24, we see that some unit prices of original data are zero, such as March 21st, and August 17th 2016. From Figure 25 and 26, the demand is quite independent of price. We guess that there is something else that is driving the demand except price. Besides, high price in Jan 2015 is with a higher demand, however the low price in Dec 2015 is with a lower demand.

Clearly seasonality on DC28603, but unclear impact of price on units sold. Clearly seasonality on DW51701, also clear impact of price on demand.

Figure 3.25 Distribution of Average Sold Price of All Products

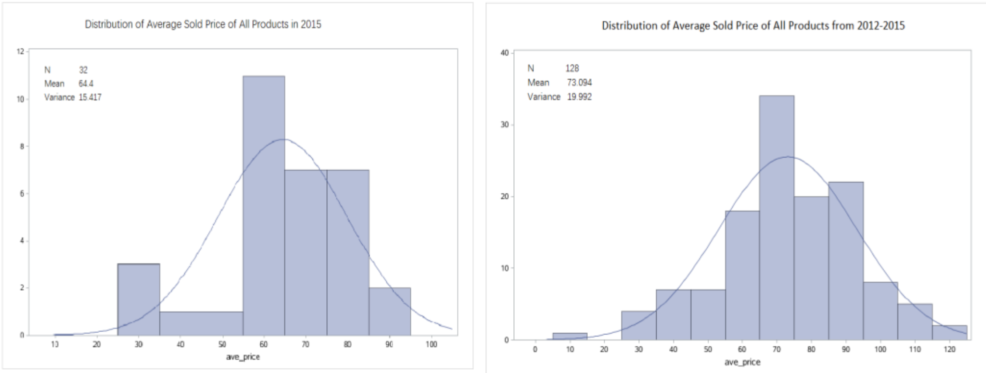
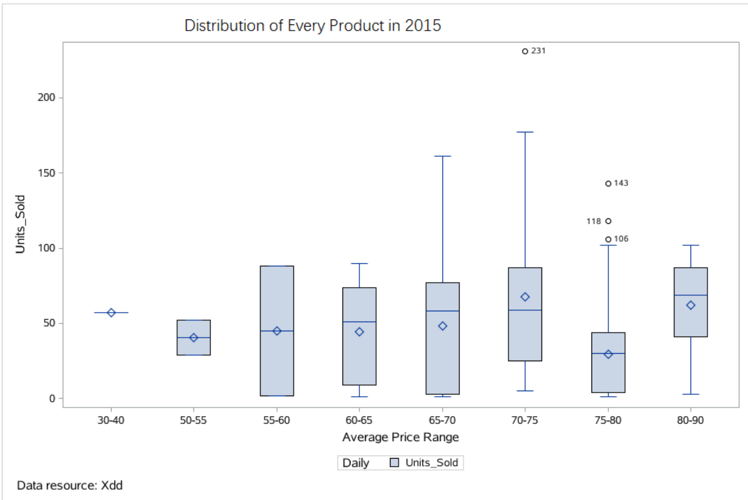


Figure 3.26 Distribution of Every Product in 2015



4.2. Linear and Log Regression Models for Two Products

In the fast fashion industry, we prefer to redefine the seasons not by spring, summer, fall and winter. In contrast, we define two weeks as one season. Therefore, there is 24 new seasons in one year. Price elasticity may be season dependent. For instance, the first 15 days of November may have a different price elasticity than the first 15 days of January. We need to group data from different years to build models to estimate season dependent price elasticity.

Next, we will test two products by the linear and log regression models. At the same time, we will also consider some special circumstances, such as seasonal factors, holidays and big marketing events “Double 11”, etc.

Controlling the seasonal factors, we plot a graph for price elasticity (y-axis is units sold vs. x-axis which is price). Take off the seasonal effect from the original data and plot it against the price. Here is how:

Suppose $\ln(y) = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n$, **Error! Reference source not found.**which is **Error! Reference source not found.** $y = e^{a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n}$. Then **Error! Reference source not found.** $y \times e^{-a_2 x_2 - \ldots - a_n x_n} = e^{a_0 + a_1 x_1}$. Let **Error! Reference source not found.** $\{y(i), x_1(i), x_2(i), \ldots, x_n(i)\}$ be the i^{th} observation (sample), we plot a graph with y-axis being **Error! Reference source not found.** $y(i) \times e^{-a_2 x_2(i) - \ldots - a_n x_n(i)}$ (seasonally adjusted) and x-axis being $x_1(i)$. Here x_1 is the average daily price, x_2 etc. are seasonal variables. It looks like an exponential distribution with a negative a_1 .

4.2.1 Optimal pricing - NLR

Let $\tilde{y}$ be the seasonally adjusted units sold. We know **Error! Reference source not found.** $\tilde{y} = e^{a_0 + a_1 x_1}$. Suppose we maximize the total revenue, thus max $\pi = x_1 e^{a_0 + a_1 x_1}$. If lower limit < x_1 < upper limit, then $\frac{d\pi}{dx_1} = e^{a_0 + a_1 x_1}(1 + a_1 x_1) = 0$ **Error! Reference source not found.** and therefore**Error! Reference source not found.** $x_1 = -\frac{1}{a_1}$.

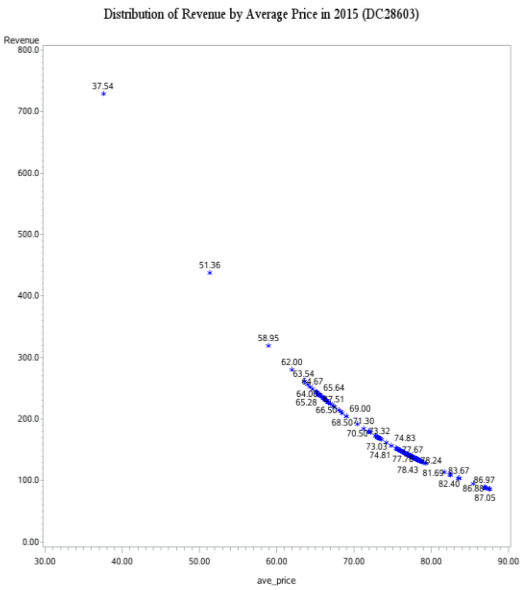
Because **Error! Reference source not found.** $a_1 = -0.10434$, the optimal price is 9.58, which is out of data range and below the lower limit, so the optimal price should be the lower limit. If **Error! Reference source not found.** $0 < x_1 < -\frac{1}{a_1}$, the revenue function $\pi(x_1)$ should be increase and when $x_1 > -\frac{1}{a_1}$ **Error! Reference source not found.**, the revenue function should be decrease. In this project the average price range at [40, 75], so in this price range, the revenue function is the decrease function that the value of revenue will be decreased with the increase of average price.

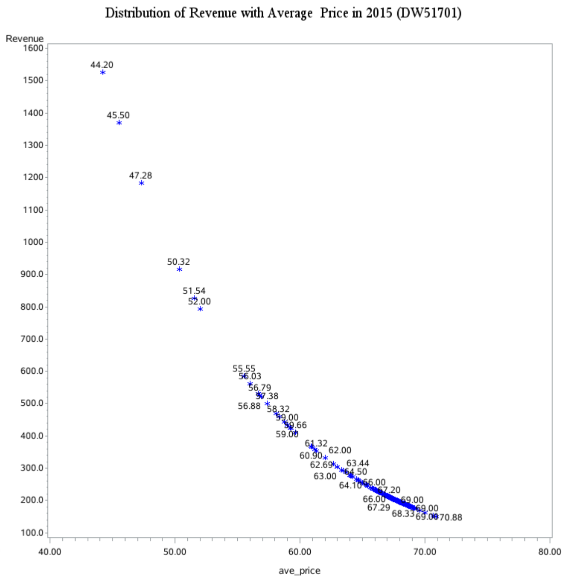
Therefore, we get optimal pricing for the following products:

$$y_{DW51701} = x_1 e^{8.15305 - 0.1043\, x_1} \text{ and } y_{DC28603} = x_1 e^{5.20506 - 0.05962\, x_1}.$$

If we are maximizing profit, then the problem should be max $\pi = (x_1 - c) \times e^{a_0 + a_1 x_1}$, given $c < x_1 <$ upper limit. Therefore, $\frac{d\pi}{dx_1} = e^{a_0 + a_1 x_1}(1 + a_1(x_1 - c)) = 0$ and $x_1 = -\frac{1}{a_1} + c$. Assuming the cost of the product is ￥30, then the optimal price is about ￥40. The point is, the gross margin should be $-\frac{1}{a_1} = 9.58$ for DW51701. Applying the same procedure, the optimal gross margin should be $-\frac{1}{a_1} = 16.77$ for DC28603**Error! Reference source not found.**.

Figure 3.27 Distribution of Revenue by Average Price for two Products in 2015





E

rror! Reference source not found.

The average observed price is 63. The model can improve profit by about ￥100 / day. It seems that customers have a higher price sensitivity to girl shoes than boy shoes! In other words, girl shoes may face more severe competition than boy shoes.

- ## 4.2.1 Optimal pricing - LR

Let **Error! Reference source not found.** $y = a_0 + a_1 x$, where x is the price, so

$\mathrm{re}\,venue = xy = a_0 x + a_1 x^2$. If we would like to maximize the revenue, then

$\max \pi = (x - \mathrm{c})(a_0 + a_1 x)$, given c < x < upper limit. Therefore,

$\frac{d\pi}{dx} = a_0 + a_1 x + a_1 (x - c) = 0$ and $x^* = \frac{a_1 c - a_0}{2 a_1}$.

Figure 3.28 The Average Price for DC28603 under Linear and Log Units Sold

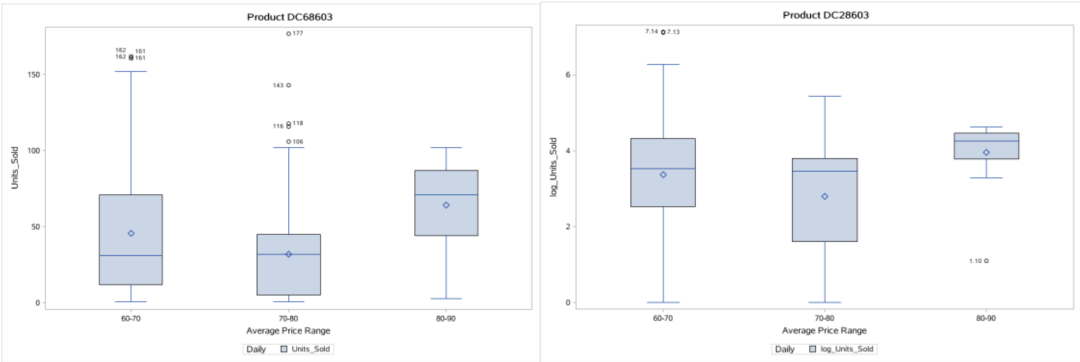
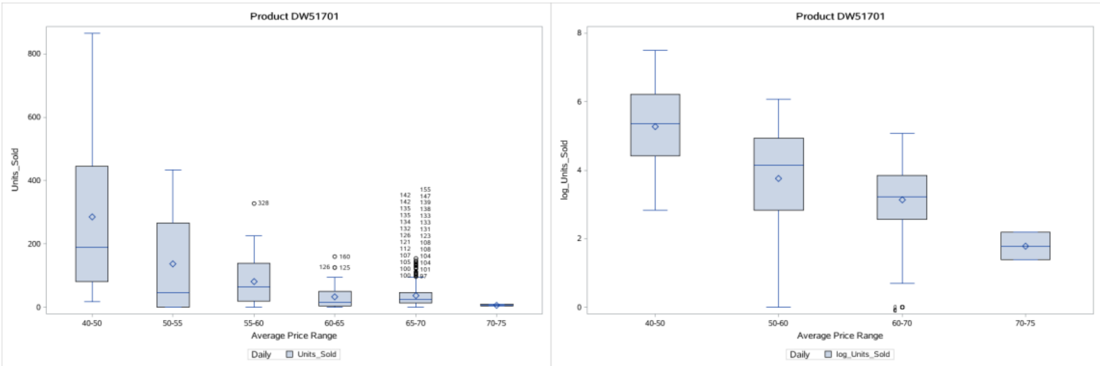


Figure 3.29 The Average Price for DW51701 under Linear and Log Units Sold



Due to the noisy data from the holidays information, we will try out to move the special dates from two products so that the models may perform better in the normal sales pattern. For product DC280603, we compare the data with and without the holidays: 11/11/2015 (1265, ￥65.28), 12/12/2015 (535, ￥63.57), 1/11/2016 (231 , ￥71.30), and 3/16/2016 (1251, ￥64.67); while for product DW51701, we compare the data with and without the holidays: 11/11/2015 (1799), 12/12/2015 (347), 2/14/2016(433), 10/21/2016 (328).

Figure 3.30 Units Sold Prediction for DW51701 with/without Holidays

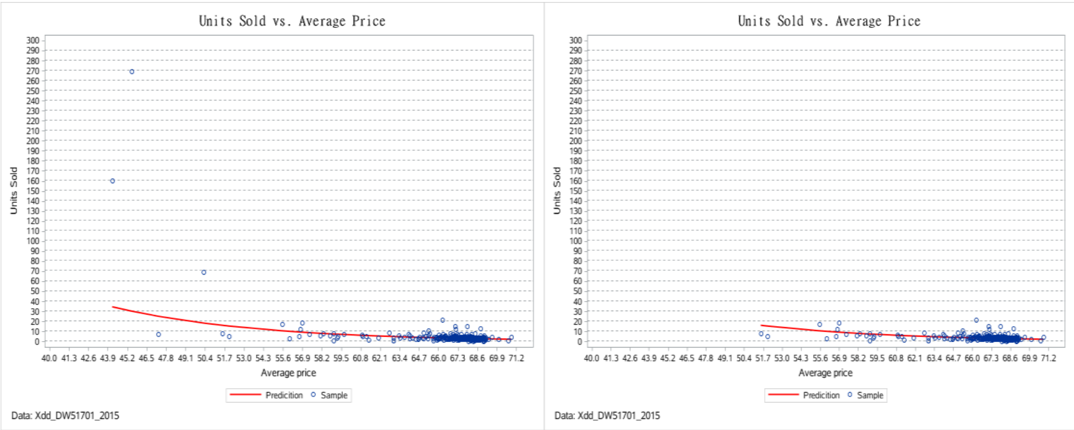
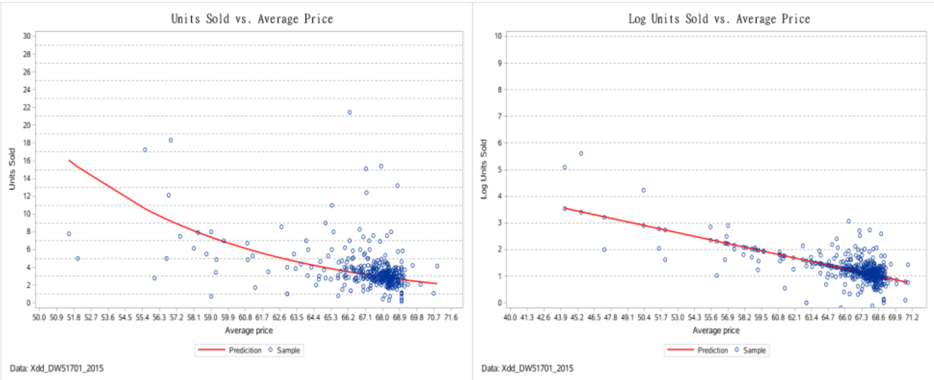


Figure 3.31 Linear and Log Regression for DW51701 without Holidays



4.3. Testing Results for Linear and Log Regressions

For both two products, the log regression works better than linear regression, due to the nonlinear impact of price and seasonality on demand. The log model is valid (by residual analysis) and significant both in R^2 and variable p-values. The impact of price on units sold is negative. Season dependent price elasticity (for different seasons, we have different price elasticity, for instance, the price elasticity in November is particularly high.

To make the testing results more consistent, we not only compare linear and log regression results in 2015, we also combine the data in 2014 and 2015. From the comparison, there are three interesting findings: 1) log models perform much better in both two products. The model accuracy is around 80%, which is up to four times higher than the linear models; 2) data in one year performs 15% better than two years, which indicates the quick change at a yearly basis in the fashion industry; 3) boy shoes prediction is 4% more accurate than that of girl shoes.

Table 3.9 Models Results Comparison for DC 28603

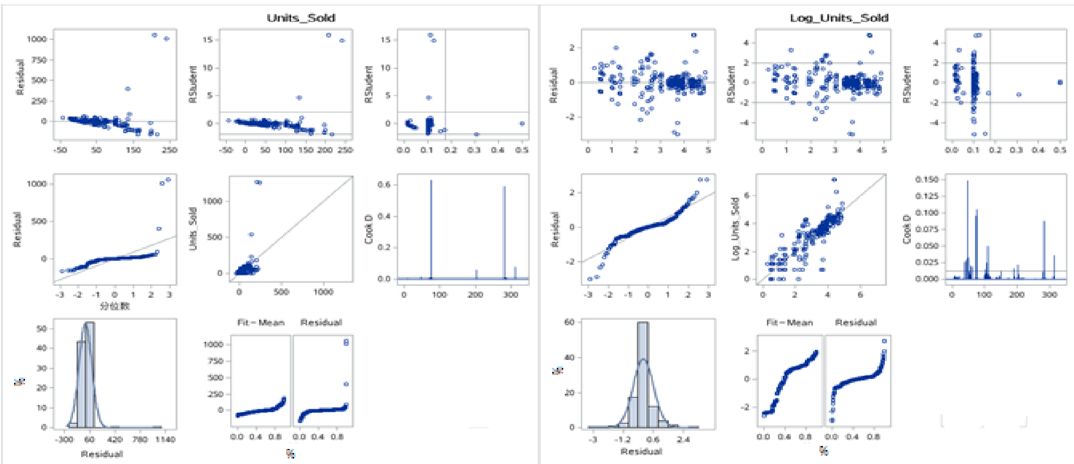
Linear Regression for DC28603 in 2015							
Source	DF	SS	MS	F	P	R^2	Adj. R^2
Regression	28	815977	29142	3.21	<.0001	0.23	0.16
Residuals	302	2741515	9078				
Total	330	3557492					
Log Regression for DC28603 in 2015							
Source	DF	SS	MS	F	P	R^2	Adj. R^2
Regression	28	633	23	55.67	<.0001	0.83	0.82
Residuals	302	123	0.41				
Total	330	756					

Linear Regression for DC28603 in 2014 and 2015							
Source	DF	SS	MS	F	P	R^2	Adj. R^2
Regression	33	1391299	42161	4.23	<.0001	0.17	0.14
Residuals	639	6363696	9959				

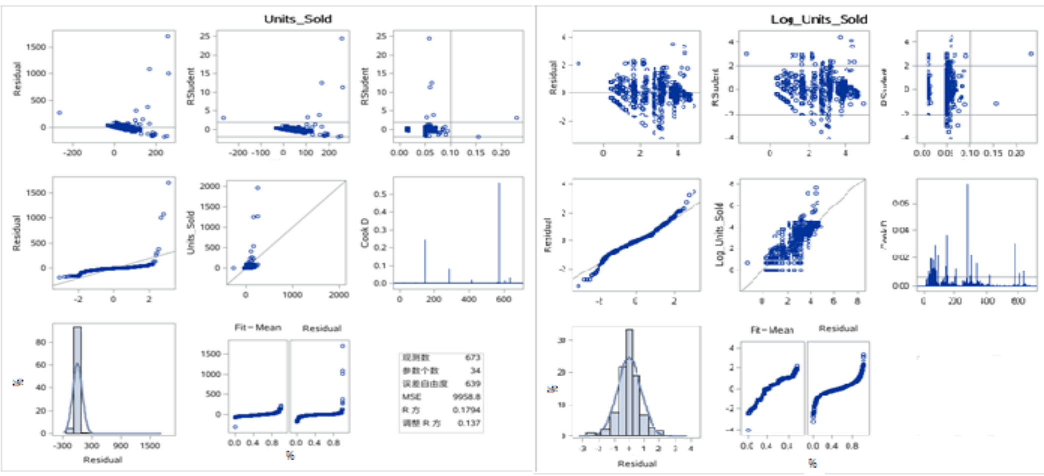
Total	672	7754994					

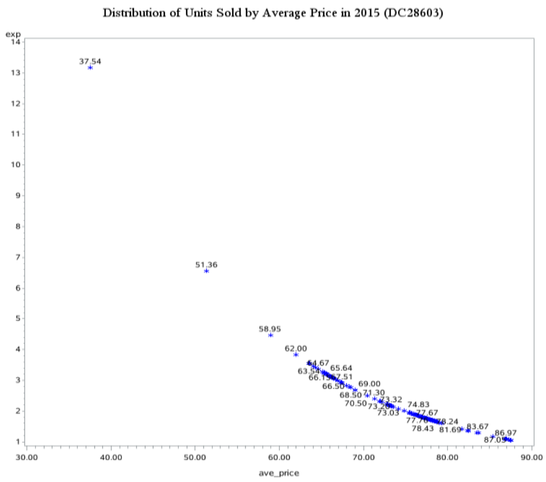
Log Regression for DC28603 in 2014 and 2015							
Source	DF	SS	MS	F	P	R^2	Adj. R^2
Regression	33	1005	30	46.22	<.0001	0.7	0.69
Residuals	639	420	0.66				
Total	672	1425					

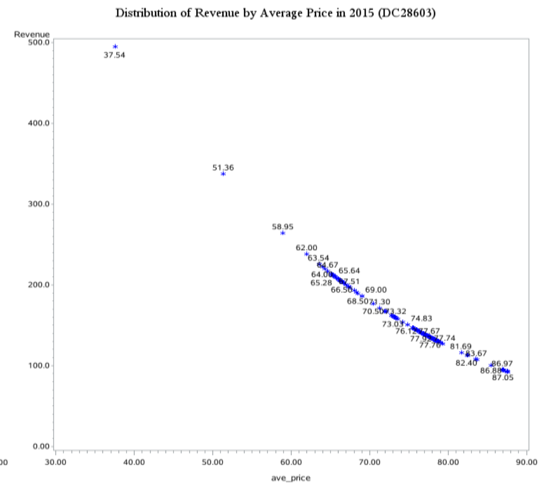
Linear vs. Log Regression for DC28603 in 2015

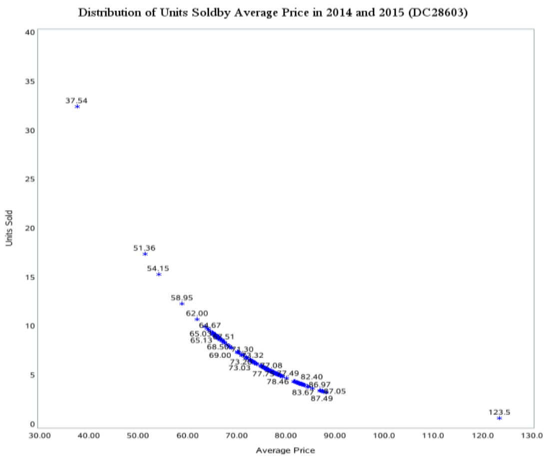


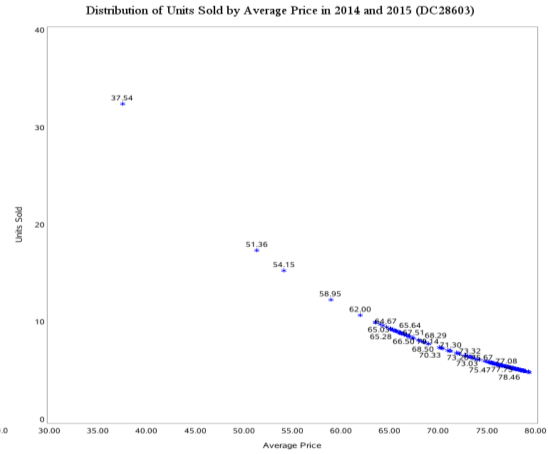
Linear vs. Log Regression for DC28603 in 2014 and 2015











Linear vs. Log Regression for DC28603 in 2015

Linear Regression						Log Regression					
Variable	Free-dom	Estimate	Standard Variance	T-Value	Pr>[t]	Variable	Free-dom	Estimate	Standard Variance	T-Value	Pr>[t]
Intercept	1	443.57882	100.72501	4.40	<.0001	Intercept	1	4.47033	0.67383	6.63	<.0001
ave_price	1	-5.82026	1.32077	-4.41	<.0001	ave_price	1	-0.05038	0.00884	-5.70	<.0001
S1	1	111.55213	35.68522	3.13	0.0019	S1	1	3.87796	0.23873	16.24	<.0001
S2	1	111.86314	32.39930	3.45	0.0006	S2	1	3.99248	0.21674	18.42	<.0001
S3	1	84.94485	32.21094	2.64	0.0088	S3	1	3.77570	0.21548	17.52	<.0001
S4	1	112.73736	32.35076	3.48	0.0006	S4	1	4.07004	0.21642	18.81	<.0001
S5	1	26.64465	32.21209	0.83	0.4088	S5	1	2.07040	0.21549	9.61	<.0001
S6	1	31.76041	32.37096	0.98	0.3273	S6	1	1.94363	0.21655	8.98	<.0001
S7	1	61.48055	32.31419	1.90	0.0580	S7	1	3.38684	0.21617	15.67	<.0001
S8	1	173.43993	32.21134	5.38	<.0001	S8	1	3.16153	0.21549	14.67	<.0001
S9	1	53.52896	32.30265	1.66	0.0985	S9	1	3.23662	0.21610	14.98	<.0001
S10	1	42.97444	32.43814	1.32	0.1862	S10	1	2.83569	0.21700	13.07	<.0001
S11	1	26.59748	32.38100	0.82	0.4121	S11	1	1.68922	0.21662	7.80	<.0001
S12	1	16.17295	32.37581	0.50	0.6178	S12	1	0.80616	0.21659	3.72	0.0002
S20	1	18.06386	32.34711	0.56	0.5770	S20	1	1.38943	0.21639	6.42	<.0001
S21	1	-9.73090	32.94376	-0.30	0.7679	S21	1	2.10247	0.22039	9.54	<.0001
S22	1	31.89002	32.33739	0.99	0.3248	S22	1	2.48310	0.21633	11.48	<.0001
S23	1	48.75612	32.31927	1.51	0.1325	S23	1	3.10207	0.21621	14.35	<.0001
S24	1	51.09147	32.35191	1.58	0.1153	S24	1	3.15214	0.21643	14.56	<.0001
S25	1	49.19139	32.38555	1.52	0.1298	S25	1	3.07545	0.21665	14.20	<.0001
S26	1	61.84616	32.34884	1.91	0.0568	S26	1	3.37763	0.21641	15.61	<.0001
S27	1	57.84154	32.34457	1.79	0.0747	S27	1	3.31221	0.21638	15.31	<.0001
S28	1	44.87203	32.40198	1.38	0.1671	S28	1	2.92207	0.21676	13.48	<.0001
S29	1	144.21498	32.50398	4.44	<.0001	S29	1	3.21832	0.21744	14.80	<.0001
S30	1	35.67086	32.22475	1.11	0.2692	S30	1	2.89543	0.21558	13.43	<.0001
S31	1	43.94662	32.22391	1.36	0.1736	S31	1	3.12053	0.21557	14.48	<.0001
S32	1	61.21303	33.71717	1.82	0.0704	S32	1	3.21541	0.22556	14.26	<.0001
S33	1	14.76266	34.76773	0.42	0.6714	S33	1	3.15923	0.23259	13.58	<.0001
S34	1	23.55641	69.64357	0.34	0.7354	S34	1	3.28025	0.46590	7.04	<.0001

Linear vs. Log Regression for DC28603 in 2014 and 2015

Linear Regression						Log Regression					
Variable	Free-dom	Estimate	Standard Variance	T-Value	Pr>[t]	Variable	Free-dom	Estimate	Standard Variance	T-Value	Pr>[t]
Intercept	1	464.66356	70.32185	6.61	<.0001	Intercept	1	5.17661	0.57190	9.05	<.0001
ave_price	1	-5.85780	0.88883	-6.59	<.0001	ave_price	1	-0.04517	0.00723	-6.25	<.0001
S1	1	80.22209	26.33126	3.05	0.0024	S1	1	2.66696	0.21414	12.45	<.0001
S2	1	53.98039	25.04591	2.16	0.0315	S2	1	2.15246	0.20369	10.57	<.0001
S3	1	34.36672	25.57739	1.34	0.1795	S3	1	1.11673	0.20801	5.37	<.0001
S4	1	56.34402	26.74422	2.11	0.0355	S4	1	1.55440	0.21750	7.15	<.0001
S5	1	7.68680	25.06707	0.31	0.7592	S5	1	0.51248	0.20386	2.51	0.0122
S7	1	17.64618	25.04935	0.70	0.4814	S7	1	1.08998	0.20372	5.35	<.0001
S8	1	82.87965	25.06182	3.31	0.0010	S8	1	1.44015	0.20382	7.07	<.0001
S9	1	22.08112	25.08833	0.88	0.3791	S9	1	1.55953	0.20403	7.64	<.0001
S10	1	15.28755	25.04964	0.61	0.5419	S10	1	1.15107	0.20372	5.65	<.0001
S11	1	6.33535	25.04573	0.25	0.8004	S11	1	0.64198	0.20369	3.15	0.0017
S14	1	-4.47232	28.16503	-0.16	0.8739	S14	1	-0.63202	0.22906	-2.76	0.0060
S15	1	-36.05037	27.84295	-1.29	0.1959	S15	1	-0.92944	0.22644	-4.10	<.0001
S16	1	-7.89875	27.42246	-0.29	0.7734	S16	1	-0.97053	0.22302	-4.35	<.0001
S17	1	-9.63557	30.98418	-0.31	0.7559	S17	1	-1.10755	0.25198	-4.40	<.0001
S18	1	-3.99441	27.41949	-0.15	0.8842	S18	1	-1.27574	0.22299	-5.72	<.0001
S19	1	-17.95087	27.51835	-0.65	0.5144	S19	1	-0.82356	0.22380	-3.68	0.0003
S20	1	-13.03182	27.44579	-0.47	0.6351	S20	1	-0.99069	0.22321	-4.44	<.0001
S21	1	-4.15497	25.56343	-0.16	0.8709	S21	1	-0.84228	0.20790	-4.05	<.0001
S23	1	6.35887	25.05224	0.25	0.7997	S23	1	1.04384	0.20374	5.12	<.0001
S24	1	-2.95928	25.30881	-0.12	0.9070	S24	1	1.44096	0.20583	7.00	<.0001
S25	1	19.16203	25.04551	0.77	0.4445	S25	1	1.57930	0.20369	7.75	<.0001
S26	1	44.43407	25.04553	1.77	0.0765	S26	1	2.26732	0.20369	11.13	<.0001
S27	1	34.40831	25.04585	1.37	0.1700	S27	1	2.05087	0.20369	10.07	<.0001
S28	1	49.35087	25.04574	1.97	0.0492	S28	1	2.34635	0.20369	11.52	<.0001
S29	1	40.95342	25.04563	1.64	0.1025	S29	1	2.20143	0.20369	10.81	<.0001
S30	1	31.11869	25.04569	1.24	0.2145	S30	1	1.96252	0.20369	9.63	<.0001
S31	1	178.48272	25.08921	7.11	<.0001	S31	1	2.17735	0.20404	10.67	<.0001
S32	1	34.79680	25.09461	1.39	0.1660	S32	1	2.16806	0.20408	10.62	<.0001
S33	1	51.68040	25.07106	2.06	0.0397	S33	1	2.22616	0.20389	10.92	<.0001
S34	1	66.04142	25.04849	2.64	0.0086	S34	1	2.20151	0.20371	10.81	<.0001
S35	1	41.87558	25.21285	1.66	0.0972	S35	1	2.41928	0.20505	11.80	<.0001
S36	1	31.14392	27.77021	1.12	0.2625	S36	1	2.31072	0.22584	10.23	<.0001

Table 3.10 Models Results Comparison for DC51701

Linear Regression for DW51701 in 2015							
Source	DF[2]	SS[3]	MS[4]	F[5]	P[6]	R^2	Adj. R^2
Regression	34	1516994	44617	6.28	<.0001	0.39	0.33
Residuals	329	2335764	7099				
Total	363	3852759					

Log Regression for DW51701 in 2015							
Source	DF	SS	MS	F	P	R^2	Adj. R^2
Regression	34	365	10.7	43.77	<.0001	0.81	0.8
Residuals	329	80	0.24				
Total	363	446					

[2] DF = Degrees of freedom

[3] SS = Sum of square

[4] MS = Means squares

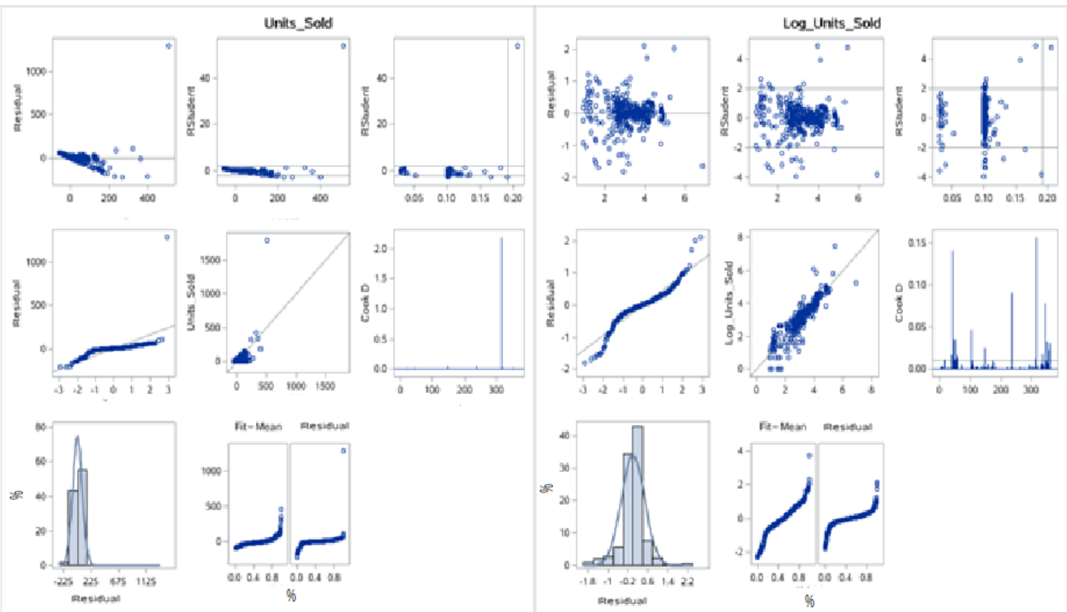
[5] F = F statistics

[6] P = p-value

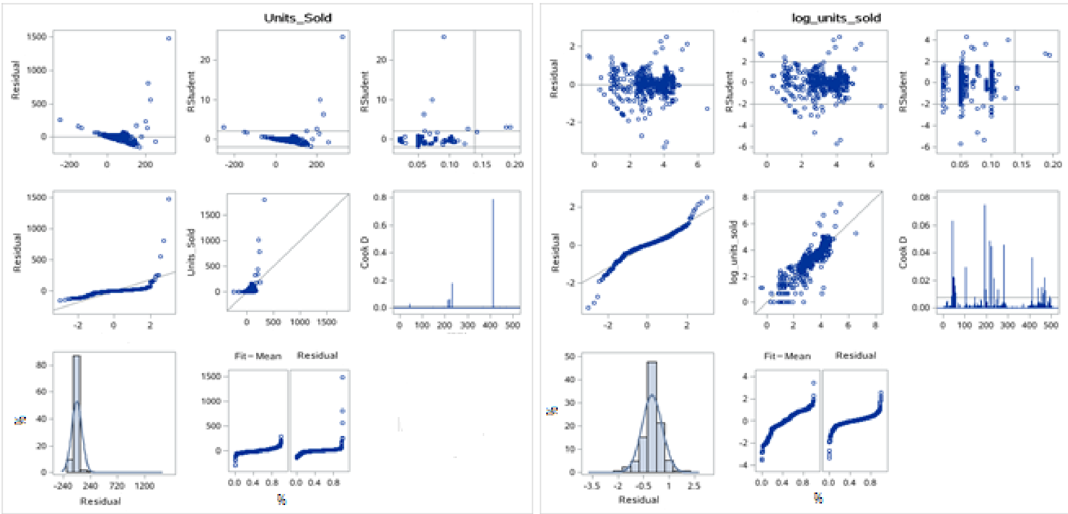
Linear Regression for DW51701 in 2014 and 2015							
Source	DF	SS	MS	F	P	R^2	Adj. R^2
Regression	34	1434229	42183	4.87	<.0001	0.26	0.2066
Residuals	471	4082281	8667				
Total	505	5516510					

Log Regression for DW51701 in 2014 and 2015							
Source	DF	SS	MS	F	P	R^2	Adj. R^2
Regression	34	599	17.6	45.94	<.0001	0.77	0.75
Residuals	471	180	0.38				
Total	505	779					

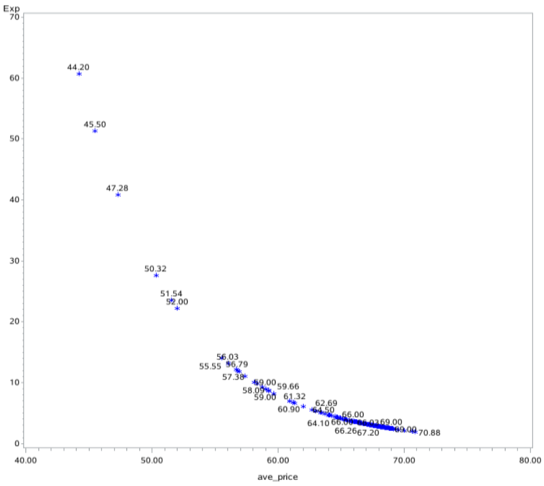
Linear vs Log Regression for DW51701 in 2015



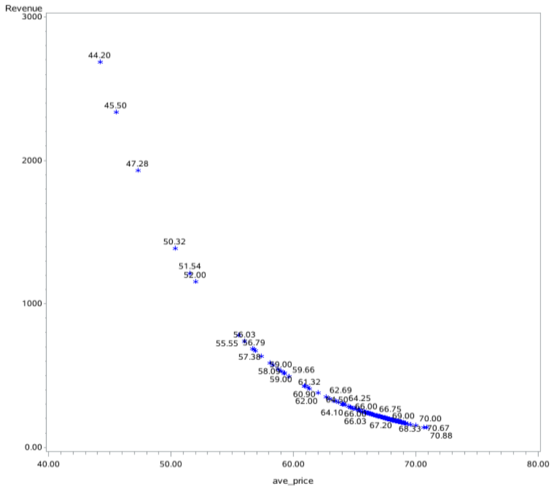
Linear vs. Log Regression for DW51701 in 2014 and 2015



Distribution of Units Sold by Average Price in 2015 (DW51701)



Distribution of Revenue by Average Price in 2015 (DW51701)



Linear vs. Log Regression for DW51701 in 2015

Linear Regression						Log Regression					
Variable	Free-dom	Estimate	Standard Variance	T-Value	Pr>[t]	Variable	Free-dom	Estimate	Standard Variance	T-Value	Pr>[t]
Intercept	1	1108.90143	96.17646	11.53	<.0001	Intercept	1	9.80046	0.56567	17.33	<.0001
ave_price	1	-16.91232	1.46959	-11.51	<.0001	ave_price	1	-0.12882	0.00864	-14.90	<.0001
S3	1	36.90228	30.45109	1.21	0.2264	S3	1	1.64803	0.17910	9.20	<.0001
S4	1	62.69793	30.67687	2.04	0.0418	S4	1	2.05868	0.18043	11.41	<.0001
S5	1	66.66465	30.44098	2.19	0.0292	S5	1	0.63580	0.17904	3.55	0.0004
S6	1	70.30528	30.63413	2.29	0.0224	S6	1	2.00202	0.18018	11.11	<.0001
S7	1	93.10098	30.57761	3.04	0.0025	S7	1	2.95805	0.17985	16.45	<.0001
S8	1	86.02771	30.58251	2.81	0.0052	S8	1	2.75926	0.17987	15.34	<.0001
S9	1	53.46335	30.39028	1.76	0.0795	S9	1	2.46298	0.17874	13.78	<.0001
S10	1	69.37618	30.60909	2.27	0.0241	S10	1	2.44020	0.18003	13.55	<.0001
S11	1	71.63796	30.84232	2.32	0.0208	S11	1	1.90640	0.18140	10.51	<.0001
S12	1	70.94526	30.70753	2.31	0.0215	S12	1	2.29399	0.18061	12.70	<.0001
S13	1	55.94421	30.67215	1.82	0.0691	S13	1	1.71322	0.18040	9.50	<.0001
S14	1	67.54560	30.75918	2.20	0.0288	S14	1	2.06401	0.18091	11.41	<.0001
S15	1	30.33441	30.38700	1.00	0.3189	S15	1	1.64446	0.17872	9.20	<.0001
S16	1	11.63291	30.31383	0.38	0.7014	S16	1	1.49118	0.17829	8.36	<.0001
S17	1	30.89646	30.38314	1.02	0.3100	S17	1	1.48989	0.17870	8.34	<.0001
S18	1	67.91333	30.86299	2.20	0.0285	S18	1	1.71434	0.18152	9.44	<.0001
S19	1	74.99634	30.83047	2.43	0.0155	S19	1	2.21014	0.18133	12.19	<.0001
S20	1	78.50877	30.74109	2.55	0.0111	S20	1	2.51118	0.18081	13.89	<.0001
S21	1	75.31718	30.72117	2.45	0.0147	S21	1	2.43483	0.18069	13.48	<.0001
S22	1	79.20519	30.65977	2.58	0.0102	S22	1	2.64288	0.18033	14.66	<.0001
S23	1	101.49076	30.58672	3.32	0.0010	S23	1	3.09090	0.17990	17.18	<.0001
S24	1	91.45757	30.31274	3.02	0.0028	S24	1	3.17882	0.17829	17.83	<.0001
S25	1	165.19644	30.61542	5.40	<.0001	S25	1	3.77425	0.18007	20.96	<.0001
S26	1	164.98040	30.62756	5.39	<.0001	S26	1	3.77571	0.18014	20.96	<.0001
S27	1	111.57454	30.64133	3.64	0.0003	S27	1	3.20414	0.18022	17.78	<.0001
S28	1	96.46873	30.69680	3.14	0.0018	S28	1	2.95619	0.18055	16.37	<.0001
S29	1	99.10834	30.66902	3.23	0.0014	S29	1	3.03213	0.18038	16.81	<.0001
S30	1	90.34883	30.40388	2.97	0.0032	S30	1	2.80266	0.17882	15.67	<.0001
S31	1	67.48795	30.75686	2.19	0.0289	S31	1	2.02814	0.18090	11.21	<.0001
S32	1	168.60216	30.32211	5.56	<.0001	S32	1	1.52451	0.17834	8.55	<.0001
S33	1	39.39079	30.40723	1.29	0.1972	S33	1	1.60021	0.17931	8.92	<.0001
S34	1	34.71716	30.50851	1.14	0.2560	S34	1	0.95751	0.17944	5.34	<.0001
S36	1	-54.65528	30.57146	-1.79	0.0747	S36	1	-0.21621	0.17981	-1.20	0.2301

Linear vs. Log Regression for DW51701 in 2014 and 2015

Linear Regression						Log Regression					
Variable	Free-dom	Estimate	Standard Variance	T-Value	Pr>[t]	Variable	Free-dom	Estimate	Standard Variance	T-Value	Pr>[t]
Intercept	1	618.09838	63.58920	9.72	<.0001	Intercept	1	9.24167	0.42298	21.85	<.0001
ave_price	1	-9.24928	0.95268	-9.71	<.0001	ave_price	1	-0.12068	0.00634	-19.04	<.0001
S3	1	16.69322	32.51488	0.51	0.6079	S3	1	1.66415	0.21628	7.69	<.0001
S4	1	33.06629	32.58776	1.01	0.3108	S4	1	2.06480	0.21677	9.53	<.0001
S5	1	47.01637	32.51187	1.45	0.1488	S5	1	0.65252	0.21626	3.02	0.0027
S6	1	42.16555	32.57342	1.29	0.1961	S6	1	2.00972	0.21667	9.28	<.0001
S7	1	67.09279	32.55476	2.06	0.0399	S7	1	2.96802	0.21655	13.71	<.0001
S8	1	59.82617	32.55637	1.84	0.0667	S8	1	2.76902	0.21656	12.79	<.0001
S9	1	37.03270	32.49747	1.14	0.2551	S9	1	2.48312	0.21617	11.49	<.0001
S10	1	42.15585	32.56511	1.29	0.1961	S10	1	2.44888	0.21662	11.31	<.0001
S11	1	36.91342	32.64459	1.13	0.2587	S11	1	1.90710	0.21714	8.78	<.0001
S12	1	40.29535	32.59815	1.24	0.2170	S12	1	2.29902	0.21684	10.60	<.0001
S13	1	26.47284	32.58617	0.81	0.4170	S13	1	1.71951	0.21676	7.93	<.0001
S14	1	35.26436	32.61580	1.08	0.2802	S14	1	2.06731	0.21695	9.53	<.0001
S15	1	14.14342	32.49659	0.44	0.6636	S15	1	1.66485	0.21616	7.70	<.0001
S16	1	4.54999	32.48357	0.14	0.8887	S16	1	1.52124	0.21607	7.04	<.0001
S17	1	14.99453	32.49557	0.46	0.6447	S17	1	1.51058	0.21615	6.99	<.0001
S18	1	32.61092	32.65181	1.00	0.3184	S18	1	1.71443	0.21719	7.89	<.0001
S19	1	40.60783	32.64046	1.24	0.2141	S19	1	2.21120	0.21712	10.18	<.0001
S20	1	32.04323	29.31804	1.09	0.2750	S20	1	1.79471	0.19502	9.20	<.0001
S21	1	69.63389	31.75949	2.19	0.0288	S21	1	2.59093	0.21126	12.26	<.0001
S22	1	75.72769	31.73144	2.39	0.0174	S22	1	2.79119	0.21107	13.22	<.0001
S23	1	144.85615	26.95156	5.37	<.0001	S23	1	2.99242	0.17928	16.69	<.0001
S24	1	74.90057	24.95272	3.00	0.0028	S24	1	3.00529	0.16598	18.11	<.0001
S25	1	95.69053	25.05200	3.82	0.0002	S25	1	3.10097	0.16664	18.61	<.0001
S26	1	109.44170	25.05400	4.37	<.0001	S26	1	3.45919	0.16665	20.76	<.0001
S27	1	95.24514	25.06269	3.80	0.0002	S27	1	3.36658	0.16671	20.19	<.0001
S28	1	68.31581	25.08058	2.72	0.0067	S28	1	2.99983	0.16683	17.98	<.0001
S29	1	68.47776	25.08913	2.73	0.0066	S29	1	2.99985	0.16689	17.98	<.0001
S30	1	58.92376	25.00870	2.36	0.0189	S30	1	2.64847	0.16635	15.92	<.0001
S31	1	41.19672	25.11174	1.64	0.1016	S31	1	2.26466	0.16704	13.56	<.0001
S32	1	126.86182	24.93556	5.09	<.0001	S32	1	1.62816	0.16587	9.82	<.0001
S33	1	21.02063	25.51136	0.82	0.4104	S33	1	1.23767	0.16970	7.29	<.0001

5. Conclusions

In conclusion, we have three important findings for further supply chain analytics, in particular, in dynamic pricing or inventory optimization. The first main finding is on price elasticity. The price elasticity of girl shoes is two times higher than that of boy shoes, and one reason is due to more severe competition. And also price elasticity by seasons by product is different. In the holidays, customers are more sensitive to the price, however the sensitivity extent for the boy shoes and girl shoes are slightly different. Besides, the sales lift is different for different products. Some products are sensitive in certain seasons; other products are sensitive in a different season.

The second main finding is on inventory. It is a mess, since XiaoDingDang will order 3 months of inventory and 10 days later order another 3 months. There may be a significant chance for improvement in inventory as well. For the future work, we may focus on inventory optimization and optimal procurement policies.

The third main finding is on coordinating demand and supply. For the future work, we may consider the dynamic pricing and inventory optimization together. For each season and each product (across multiple years), we first run a regression for units sold on average price to get the statistics summary (# of samples, average units sold, etc.) and the line fit plot. If we get the large absolute coefficient, small p-value, high R^2, line fit plot and residuals look no problem, then the results seem significant. If significant, then we can determine the optimal pricing. Once it's not the price sensitive, we then can raise price. Otherwise, we suggest ways to do experiments: fire at well or increase price range to experiment.

Chapter 4. Scheduling the supply chain operations with external contract and resource constraints.

1. Background

Outsourcing has played an important role in low skilled but highly labor-intensive industries such as the apparel and shoe industries. According to the survey conducted by Kremic et al. (2006), decisions on outsourcing are driven by three major factors - cost, strategy, and political considerations. Players in the fashion industry such as Zara and Nike, have their own strategies to respond to rapid market changes. Ferdows et al. (2004) claimed that when Zara, a fast fashion retailer, produces a garment in-house, it uses local subcontractors to undertake the simple and labor-intensive steps of the production process, such as sewing. In these small workshops, they only have a few dozen employees, and Zara is their primary customer. Nike, as another example, has outsourced their total production to developing countries, including China, Brazil and Indonesia, in order to benefit from lower wages and management costs, lower raw materials costs, hence lower overall production costs. Nike is only responsible for product design and marketing, and also for the quality control of the products. Except for the Nike Air series, which is the core competency of the business, Nike has outsourced all of its product series. Therefore, in a very short amount of time, Nike has captured almost 75% of the sneaker market. By putting most of its manufacturing sites in China, proximity to customers and major markets is another good reason for the outsourcing strategy. Brun et al. (2008) surveyed 12 Italian luxury fashion retailers, and of these, 8 companies are partially outsourcing their products to local suppliers, who are providing leather, accessories or shoes components. Kahraman et al. (2010) confirm that, strategically speaking, manufacturers are not always able to fulfill all their customers' orders internally. Therefore, outsourcing some non-core business is one way to maintain a high customer service level but also to maintain a relatively low profit margin. According to Federico et al. (2011), in a sample of 15 Italian luxury product companies, they all focus their efforts on resource-intensive processes (e.g., design, quality management and branding) with regard to the core product; consequently, the production of accessories and licensed products is usually outsourced to external suppliers.

The real life challenge that has motivated our study in terms of this coordination problem was encountered in the daily operations of QianFang Holding Company (Zhejiang, China). As a Top Ten girl's leather shoe manufacturer in China, QianFang produces more than 3 million pairs of shoes annually in order to meet not only the domestic market needs of thirty offline cities and five online channels, but also the international customer demand from Janie and Jack, a premium brand of Gymboree in USA, mainly as a OEM. Janie and Jack is QianFang's new customer with extremely high requirements in terms of deadline and quality. However, it takes up a quite high weight among those three groups, as QianFang holds a private agenda of developing a better understanding the US kids' shoemaking standards in order to pave a way into the US market. The on-line channel demand is increasing dramatically, with a relatively high profit margin, but associated with exclusive designs, which may have a longer production time. Last but not least, 30 off-line domestic customers are the basis of the business, making up more than 70% of sales revenue.

In order to make one pair of shoes, more than 200 operations are involved in three different plants – the Cutting Plant, the Sewing Plant and the Packing Plant. In the Cutting Plant, due to technical limitations in that some shoes need laser printing or computerized embroidering on the upper leather parts, they have to be sent out for external processing. In the Sewing Plant, accessories related to the tops of shoes, such as three-dimensional cloth cartoon figures or flowers, are outsourced to third parties in order to keep the company's core business for the company to concentrate on its core production processes. More than 400 workers sitting in 12 assembly lines in the Sewing Plant will then mass sew the accessories that have been supplied by third parties onto the upper sections above the sole to provide be a complete semi-finished product. Last but not least, equipped with the latest technology and facilities imported from Italy, 4 packing

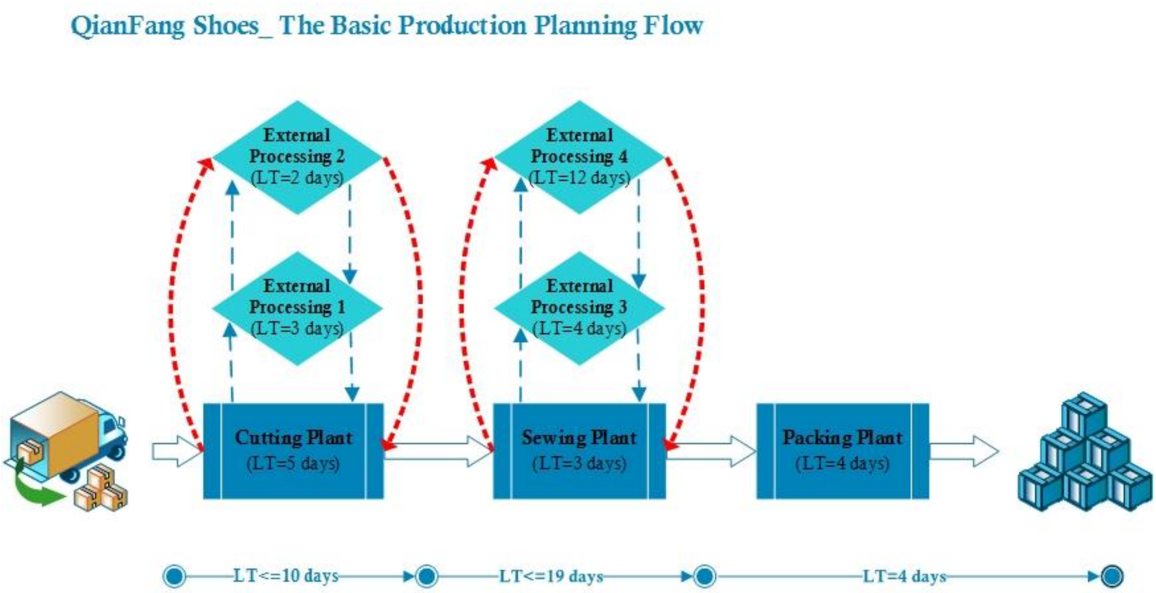
assembly lines in the Packing Plant can semi-automatically pack about 2,500 to 3,000 pairs of shoes per day, which leads to a production capacity of 3 million pairs per year7.

Therefore, the manufacturing process at QianFang takes about 12-33 days, depending on which external processing steps in terms of cutting or sewing are involved. The Packing Plant does not need any external processing. In our case, for

[7] 2,500 (pairs of shoes per day) x 4 (4 assembly line) x 27 (work days per month) x 11 (work months per year) = 2,970,000 pairs of shoes per year.

one pair of shoes, it will involve up to two external processing procedures in both the Cutting Plant and the Sewing Plant.

Figure 4.1 QianFang's production process with the first two plants involving external processes



When external contractors are involved in a supply chain process, the timely fulfillment of customer orders is critically important to ensure a successful operation because of the challenges of achieving close coordination between supply chain partners (Wiendahl, 1995). However, in the apparel and shoe-making industry, due to rapid seasonal changes with regard to fashion trends, meeting the customers' due dates is the bottom line when it comes to the manufacturers' survival. For example, the due dates for printing orders for periodical magazines in a competitive market are not negotiable (Lee et al., 1997). When external contractors are involved, and when the business operates in a due-date based made-to-order process, the coordination between the internal operations and the external operations becomes critical since, once a manufacturer fails to deliver the customer's orders on their specified due-dates, products being produced become markdowns, at a cost to all the business partners in the value chain (Lee and Choi, 2011).

Part I of this research focuses on the analysis and on the solution methodologies required to solve the operations planning/scheduling problem involving external (outsourced) operations and resource constraints. The objective is to minimize the total degree of tardiness in terms of customer order fulfillment. In section 2 we review the existing results, and highlight the contribution of this study to the literature. In section 3, we introduce the notations and assumptions, and propose a mixed integer programming model that provides a formal definition of the problem. We then present a structure analysis of the problem, and discuss two solvable cases under certain conditions. The results of this analysis are then used to construct the solution methodology to be proposed in Section 4. A numerical example demonstrating the solution process is also provided.

2. Literature Review

The existing results with regard to production scheduling with outsourcing options can be classified into that of a single-machine/process environment, and that of a multiple-machine/process environment. Representative works for the single-machine/process environment are as follows. Lee and Sung (2008a) discussed a single machine-scheduling problem with an outsourcing option. The objective function is to minimize the total outsourcing costs and the total completion time, subject to an outsourcing budget constraint. The authors proposed two heuristic methods and also a branch-and-bound approach to solve this NP hard problem and used numerical examples to validate the performance. Qi (2008) considered a single-machine scheduling problem, with the option of outsourcing some orders to one machine subcontractor. All the outsourced orders are returned to the facility in a batch, and thus the outsourcing/transportation costs is one of the objective function will be taken into account in any case, when comparing four different objective functions, including 1) minimizing total completion time **Error! Reference source not found.**; 2) minimizing the makespan (the completion time of the last job) **Error! Reference source not found.**; 3) minimizing the maximum lateness **Error! Reference source not found.** and 4) minimizing the number of late jobs **Error! Reference source not found.**, where **Error! Reference source not found.** if **Error! Reference source not found.** and **Error! Reference source not found.** if **Error! Reference source not found.**. After developing dynamic programming algorithms, the authors concluded that the first objective function (i.e., minimizing the sum of outsourcing/transportation costs and total completion time) could be solved in polynomial time, by making decisions with regard to: 1) the subset of jobs to be outsourced; 2) the processing sequence for the jobs on the in-house machines; 3) the processing sequence of the outsourced jobs; 4) the transportation bathing for the outsourcing jobs. Based on the Shortest Processing Time (SPT) property, the time complexity of the algorithm is in **Error! Reference source not found.**. The second problem is reduced to partitioning all jobs into two subsets, with one part for in-house production, and one part for outsourcing, and the sequences in both sets being really arbitrary. Moreover, the third and fourth problems are both based on the Earliest Due Date (EDD) property. Overall, the last three problems are all NP-hard in the ordinary sense.

For the multiple-machine/process environment, Lee and Sung (2008b) discussed two single-machine problems with an outsourcing option, where the objective function is to minimize the weighted sum of outsourcing costs and scheduling measures, involving minimizing maximum lateness and minimizing total tardiness, subject to the outsourcing budget constraints. The authors assumed unlimited capacity for the subcontractors, and hence they ignored the scheduling of outsourced orders. Some researchers also considered the two-stage scheduling problem with an option of outsourcing. Qi (2009) studied a two-stage production-scheduling problem in with each job has two sequential operations. Both stages involved in-house production with only one machine. However, in stage one, the manufacturer has an option of either producing the first operation of orders in-house, or outsourcing some orders to a remote subcontractor. All outsourced orders should be batch delivered back to the facility for stage two of the second production operation. The objective function is to optimally balance the outsourcing costs and the makespan. The author then developed an optimal algorithm according to a Johnson schedule. By partitioning the jobs into three subset, in which 1) all outsourced jobs are processed consecutively in a time interval [a, b] on the stage two machine; 2) all in-house jobs with **Error! Reference source not found.**; and 3) all in-house jobs with **Error! Reference source not found.**, the problem is solvable in pseudo-polynomial time. However, this algorithm is too complex to implement in practice. Therefore, the author proposed another heuristic algorithm with reduced complexity. Here, he only considered a schedule with all the in-house jobs having **Error! Reference source not found.** and its worst–case bound in terms of the makespan is 2, i.e., **Error! Reference source not found.**. In the computational tests, the author found that the job processing time distribution is a critical factor that affects the outsourcing benefits. When the number of jobs increases, the computational errors decrease, which shows the scalability of the algorithm. Lastly, the author extended the basic model into two more complex scenarios: 1) the different production facility of the outside supplier; 2) multiple parallel outside supplier's machines with infinite capacity. Under both scenarios, the two dynamic programming formulations are still valid, and only some small changes are needed. Choi (2013) analyzed the computational complexity of the two-machine ordered flow shop problem, where each job is processed in-house or outsourced to a subcontractor. The objective function is to minimize the sum of the makespan and the total outsourcing cost. Since this problem is NP-hard, they present an approximation algorithm. They consider three special cases in which job j has a processing time requirement p_j, and machine i a characteristic q_i. The first case assumes that the time job j occupies machine i and is equal to the processing requirement divided by a characteristic value of machine i, that is, pj/qi. The second (third) case assumes that the time job j that occupies machine i is equal to the maximum (minimum) of its processing requirements and a characteristic value of the machine, that is, $\max\{p_j,q_i\}$ ($\min\{p_j,q_i\}$). Finally, they show that the first and the second cases are NP-hard, and the third case is polynomial-time solvable.

Up till now, very few studies have been presented for scheduling with an outsourcing option in a complex shop scheduling environment with multiple stages. A two-stage flow shop scheduling problem with outsourcing options, where each in-house production stage involves one machine, was presented by Qi (2011). Here, the author designed three different scenarios: Model 1) outsourcing both operations for a subset of jobs to a single subcontractor; Model 2) outsourcing the stage-one operation for all jobs to a single subcontractor; Model 3) outsourcing both operations for a subset of jobs to two subcontractors, each focusing on one operation. Given a specific configuration of outsourcing, the author developed an optimization algorithm to find the best schedule, and also conducted computational experiments to validate both models and algorithms at the operational level. In all the scenarios, the author failed to consider the possibilities that the orders will return to the facility for post-processing or any packing work before being sent directly to the customers. In reality, some manufacturers would like to send out the finished goods by themselves, as they pay a lot of attention to customer service levels by providing more personalized packages, or better controlling product quality, or closely monitoring the whole product and information flow.

Recently, Kangbok and Choi (2011) discussed a two-stage production scheduling problem in which each activity requires two operations to be processed in stages 1 and 2. In each stage, the company can either produce in-house or outsource to subcontractors. The objective function is to minimize the total weighted sum of makespan, and also to minimize outsourcing costs. The authors presented the approximation algorithm for one NP-hard case with regard to the worst-case analysis, and also summarized the computational complexity of the problem in the cases of both a general flow shop and a proportionate flow shop, depending on the values of the unit time outsourcing cost for the operations at different stages.

More recently, Mokhtari et al. (2012) considered a scheduling problem, where multiple stages, multiple machines, multiple operations and multiple subcontractors are involved, and transportation costs are also considered. Each order can be either produced by in-house production or by subcontractors at relatively higher cost, in order to meet the customer's due date. By minimizing the outsourcing costs as well as the weighted flow mean time, they not only should determine the sequence of in-house production orders, but also need to decide which orders should be outsourced, and which vendor to choose from. Two different outsourcing policies are presented: Policy I) each order has to be processed either using in-house resources as a whole, or outsourced to one the subcontractors; and Policy II) each order is divisible and each operation of an order can be separately processed. The authors developed Mixed Integer Programming models under four scenarios in consideration of subcontractor's capacities, including 1) capacitated problem with outsourcing Policy I; 2) incapacitated problem with outsourcing Policy II; 3) capacitated problem with outsourcing Policy II; 4) incapacitated problem with outsourcing Policy II. The authors developed mixed integer nonlinear programming for each scenario, and in order to solve the proposed models, an effective approach based on the Team Process Algorithms (TPA) was devised. Based on two illustrative cases, a comparison between the suggested TPA and the traditional approach was conducted, and it was found that TPA outperforms the other.

Methodologically, in order to solve a scheduling problem with outsourcing options, few researchers have developed a mixed integer programming model. Coman (2000) formulated the production-planning problem with outsourcing options as a linear programming problem so as to maximize the throughput from manufacturing and from outsourcing products regard to manufacturing and outsourcing products. Lee et al. (2002) developed an advanced planning and scheduling model to integrated states include: a) selecting the best machine for each operation, b) deciding the sequence of operations, c) picking the operations to be outsourced, and d) minimizing the makespan for the due date of each order. More recently, Zhen (2012) presented an analytical approach to investigate the optimal decision in terms of a multiple product production-planning problem with stochastic demand, either manufacturing all the parts in-house and assembling them, or outsourcing parts of products and then assembling them. However, the author failed to consider the customer due date constraint, and also customer weighting was ignored.

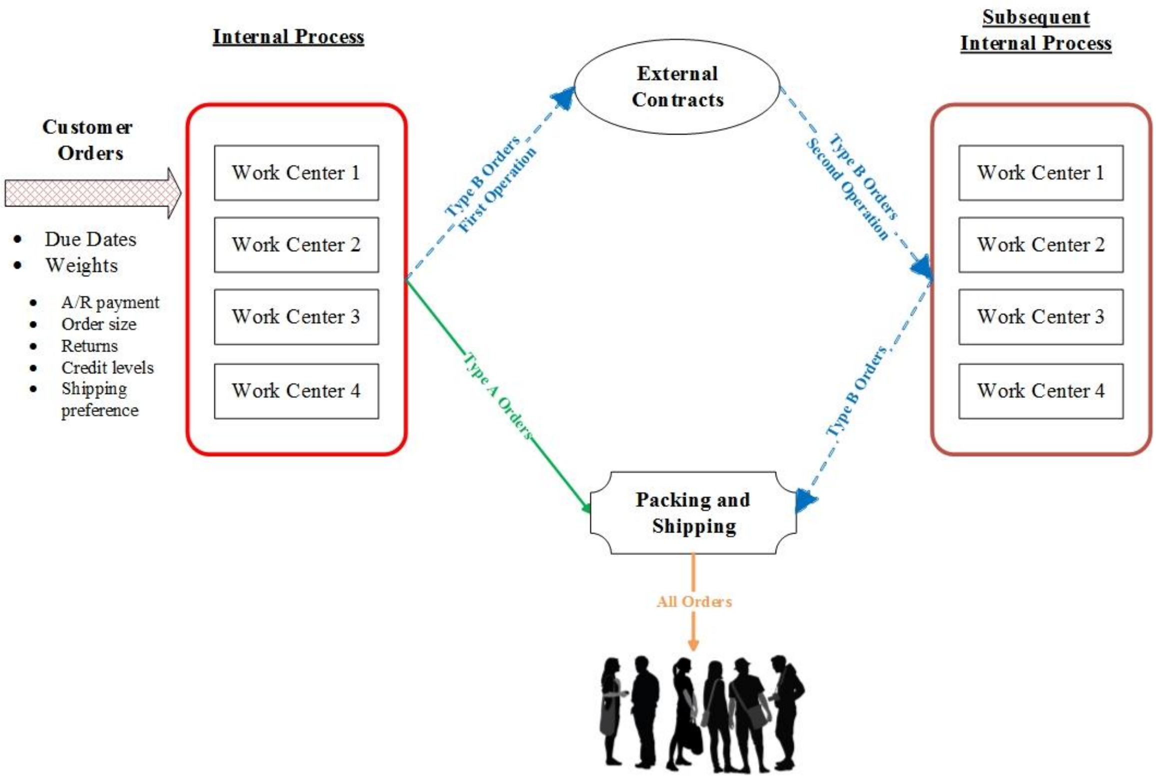
2.1. Contribution of Part 1 Results

Part I of this research contributes to the literature with regard to two aspects. Firstly, from a modeling point of view, customer orders, after the external operations, may return to a subsequent internal process, which makes our modeling more challenging and different from most of those discussed in the literature. Based on our knowledge, most existing results assume a two-stage process: internal operations and then external operations. Secondly, our model is the very first to explicitly include the production time windows of external operations, an aspect which is however, very commonly encountered in practice. Most external contract manufacturers have their own resource and capacity constraints, and can only accommodate the needs UNCLEAR within certain time windows.

3. A Formal Definition of the Problem P1

Figure 4.2 below illustrates the supply chain process with regard to which we shall develop the mathematical programming model that formally defines our problem, P1.

Figure 4.2. Supply Chain Process of Type A and Type B Customer Orders



If one job needs to incorporate external processing, we consider that this order has two operations, with the first production part before external processing referred to as Operation A, while Operation B refers to the parts that need to be done after external processing. Therefore, if one job has any external processing, the whole production time in this plant is divided into three parts: time in Operation A, external processing time, and time in Operation B. However, if one job has no external processing, this job only has a single operation - Operation A. Both the external processing time and time in Operation B are equal to 0.

There are some identical parallel machines in one plant, and jobs need to be assigned to different machines. The processing time on any machine is the same for each job. For any chosen job, the external processing time is given. Each operation can only be processed on one machine at a time. Operations cannot be interrupted, and all machines are available in the planning horizon. The switching time between jobs is ignored. Depending on the value of orders, each job with is given a different weight. The due date of each order is given. The capacity of outsourced operations is unlimited. That is, all required external-processing operations could be outsourced.

To define this problem more formally, we introduce the following notation.

J Set of jobs

M Set of machines

R Set of resources

Error! Reference source not found. The weight of job j for j in J

Error! Reference source not found. Set of orders. Each has only a single internal operation

Error! Reference source not found. The set of orders, each has two internal operations and one external operation, and **Error! Reference source not found.**, where **Error! Reference source not found.**, and **Error! Reference source not found.**, refers to the set of first internal operations, and the set of the second internal operations, respectively.

Φ Set of all operations. That is **Error! Reference source not found.**

Z A sufficiently large number

Error! Reference source not found. The due date of order j

Error! Reference source not found. The external processing time of job j in **Error! Reference source not found.**

Error! Reference source not found. The required processing time of job j on machine m under a given resource R

Decision Variables:

Error! Reference source not found. If operation j is processed on machine m, the starting time of operation j on machine m, and otherwise 0. **Error! Reference source not found.**

Error! Reference source not found. The tardiness of job j, **Error! Reference source not found.**

Error! Reference source not found. Binary variable indicating whether job i precedes job j on machine m,**Error! Reference source not found.**

Error! Reference source not found. Binary variable indicating whether job j is the first job assigned to machine m, **Error! Reference source not found.**. Let {0} be a dummy job.

Hence, the problem could be formatted as:

$$\boldsymbol{Minimize}: \sum_{j\in\Phi_A\cup\Phi_{B2}} w_j \cdot T_j \tag{3.0}$$

Subject to:

1. All the operations j, **Error! Reference source not found.** *must be completed*

$$\sum_{\forall\, k\in\Phi\cup\{0\}}\sum_{\forall m\in M} X_{k,j,m} = 1 \qquad \forall\, j \in \Phi \tag{3.1}$$

$$\sum_{\forall m\in M} X_{0,j,m} = 1 \qquad \forall\, j \in \Phi \tag{3.2}$$

2. Constraints on the operation starting times

$$S_{j,m} \le Z \cdot \sum_{\forall\, k\in\Phi\cup\{0\}} X_{k,j,m} \qquad \forall\, j \in \Phi \quad \forall\, m \in M \tag{3.3}$$

$$S_{i,m} + p^R_{i,m} \le S_{j,m} + Z \cdot X_{i,j,m} \qquad \forall\, i,j \in \Phi_A \cup \Phi_{B1} \quad \forall\, m \in M \tag{3.4}$$

$$\sum_{\forall m\in M}\left[S_{i,m} + p^R_{i,m} \cdot \sum_{\forall\, k\in\Phi\cup\{0\}} X_{k,i,m}\right] + e_i \le S_{j,m'} + Z \cdot (1 - \sum_{\forall\, k\in\Phi\cup\{0\}} X_{k,j,m'}) \qquad \forall\, i \in \Phi_{B1} \quad \forall\, j \in \Phi_{B2} \quad \forall\, m,m' \in M \tag{3.5}$$

3. Constraints on the order tardiness

$$T_j \ge \sum_{\forall m\in M}\left[S_{j,m} + p^R_{j,m} \cdot \sum_{\forall\, k\in\Phi\cup\{0\}} X_{k,j,m}\right] - d_j \qquad \forall\, j \in \Phi_A \cup \Phi_{B2} \tag{3.7}$$

4. Domain constraints

$\forall\ k \in \Phi \cup \{0\}$ (3.8)

$X_{k,j,m} \in \{0,1\}$ $\forall\ j \in \Phi_A \cup\ \Phi_B$

$\forall\ m \in M$

Error! Reference source not found. are non-negative continuous variables (3.9)

The objective function (3.0) minimizes the total tardiness for all the jobs. Constraint (3.1) enforces any chosen job j (job j could be the first operation or the second operation in **Error! Reference source not found.** or job j in **Error! Reference source not found.**) that must be completed using some machines. Constraint (3.2) ensures that job j could also be assigned as the first job on machine m. Constraint (3.3) guarantees that if job j is not processed on machine m, the starting time of job j on machine m is forced to be 0. Constraint (3.4) shows that for any pair of jobs i and j in the first operations, the starting time of job j on machine m is later than the starting time of job i plus any processing time of job i on the same machine m. Constraint (3.5) considers that for any second operation of job **Error! Reference source not found.** in **Error! Reference source not found.**, the starting time of job **Error! Reference source not found.** must be later than the internal production time **Error! Reference source not found.**plus the external processing time **Error! Reference source not found.** and the starting time **Error! Reference source not

found. of any fist operation of job j in **Error! Reference source not found.**. Constraint (3.6) represents the tardiness of job j expressed as the difference between the completion time of job j and the due date of job j. Constraint set (3.7) indicates the binary restrictions of variable **Error! Reference source not found.**, while Constraint (3.8) en ensures that the job starting time and tardiness are non-negative.

Note: the discussions in Sections 4 and 5 use the following notations:

Φ_A The set of jobs/orders that each has only a single operation

$\Phi_B = \{ j =< j^-, j^+ > \}$ The set of jobs, that each job j has two operations, j^- and j^+

4. A Solvable Case of P1

The general version of problem P_1 is computationally intractable (i.e., NP-Hard). Therefore, it merits the use of effective heuristics in problem solving, especially for problems with practical sizes. To facilitate the design of such heuristics, special cases that can be solved in strongly polynomial time, are of great interests. In this section, we introduce such a special case with single operation per customer order, and identical processing times. Meanwhile, we do allow different customer orders to have different due dates, and to require different amount of time for packaging and shipping before the delivery.

To start, let's first consider the scenario where the number of processing lines (or machines) in the system equals to one (i.e., |M|=1), each customer order has only one operation (i.e., $\Phi_B = \varnothing$, $\Phi_A = \Phi$), and all the operations have the same processing time on the given single machine (i.e., $P_{j,.}^R = P, \forall j$). Upon such assumptions, problem P1 reduces to the well known Assignment problem. To see this, consider the fact that each customer order (operation) requires a constant P units of processing time and thus the total processing time performed by the single machine is $|\Phi| \cdot P$. Therefore, the optimization problem P1 becomes the one to sequence the given $|\Phi|$ customer orders, each has a single operation of duration P time units, so that the total weighted tardiness is minimized. Let the cost of assigning order j to position k in the sequence as

$$C_{j,k} = Max.\{0, P \cdot k - d_j\}, \ \forall j \in \Phi_A, k = 1, 2, \ldots, |\Phi_A|$$

where $P \cdot k$ defines the completion time of the *k-th* job in the sequence. If each customer order, *j*, also requires a packaging/shipping time, δ_j, before the delivery, we can redefine the cost of assigning order j to position k in the sequence as

$$C_{j,k} = Max.\{0, P \cdot k + \delta_j - d_j\}, \ \forall j \in \Phi_A, k = 1, 2, \ldots, |\Phi_A|$$

Then, the optimal solution to the reduced P1 in this special case can be obtained by solving the following problem:

$$Min. \sum_{\forall j \in \Phi_A} \sum_{k=1,2,\ldots,|\Phi_A|} C_{j,k} \cdot X_{j,k}$$

$$s.t.$$

1. Each job/operation must be assigned to a position in the sequence

$$\sum_{k=1,2,\ldots,|\Phi_A|} X_{j,k} = 1, \qquad j = 1, 2, \ldots, |\Phi_A|$$

2. Each position in the sequence must be filled with exactly one job/operation

$$\sum_{k=1,2,\ldots,|\Phi_A|} X_{j,k} = 1, \qquad j = 1, 2, \ldots, |\Phi_A|$$

$$X_{j,k} \in \{0,1\}, \ \forall j \in \Phi_A, k = 1, 2, \ldots, |\Phi_A|$$

which is the known Assignment problem, and can be solved in $O(|\Phi_A|^3)$.

This special case can be extended to the situation with multiple process lines (machines) as long as each customer order has only a single operation and the processing times are identical on each given process/machine *m*. To see this, let P_m be the required processing time on machine/process *m* for any operation, and let K_m be the length of the sequence assigned to machine m, $m=1,2,\ldots,|M|$, so that $\sum_{m=1,2,\ldots,|M|} K_m = |\Phi|$, and let

$$C_{j,k,m} = Max.\{0, P_m \cdot k - d_j\}, \ \forall j \in \Phi_A, \ where\ \Phi_A = \Phi, \ k = 1,2,\ldots,K_m$$

be the cost of assigning operation j to position k on the sequence assigned to process/machine *m*, $m=1,2,\ldots,|M|$. Then, the respective version of problem P1 becomes:

$$Min\ . \sum_{m=1,\ldots,|M|} \sum_{\forall j \in \Phi_A} \sum_{k=1,2,\ldots,K_m} C_{j.k.m} \cdot X_{j,k,m}$$

$$s.t.$$

1. Each operation j must be assigned to a position in the sequence of some machine *m*, $1 \le m \le |M|$

$$\sum_{m=1,2,\ldots,|M|} \sum_{k=1,2,\ldots,K_m} X_{j,k,m} = 1, \qquad j = 1,2,\ldots,|\Phi_A|$$

2. Each position in the sequence of machine *m*, $1 \le m \le |M|$, must be filled with exactly one job/operation

$$\sum_{j=1,2,\ldots,|\Phi_A|} \sum_{k=1,2,\ldots,K_m} X_{j,k,m} = 1, \qquad m = 1,2,\ldots,|M|$$

$$X_{j,k,m} \in \{0,1\}, \quad \forall j \in \Phi_A,\ k = 1,2,\ldots,K_m,\ m = 1,2,\ldots,|M|$$

As we can see, for any given set of parameters $K_m,\ m = 1, 2, \ldots, |M|$, with $\sum_{m=1,2,\ldots,|M|} K_m = |\Phi_A| = |\Phi|$, , the respective version of problem P1 becomes a linear 3D-Assignment problem with a totally unimodular coefficient matrix and can also be solved as a linear programming (LP) problem.

5. The Proposed Solution Approach for P1

In this section, we propose a heuristic search algorithm based on the special case introduced in section 4. In summary, this heuristic algorithm solves problem P1 in two steps:

Step 1. Heuristically estimate the average processing time per operation on each machine *m*, P_m, and determine the value of parameter $K_m,\ m = 1, 2, \ldots, |M|$, so that

$$P_m \cdot K_m \approx P_{m'} \cdot K_{m'},\ \forall m, m' \in M .$$

Since some customer orders may involve external contracted manufacturing (E_j) and may require a subsequent (i.e., the 2nd) internal operation ($P_{j^+,m}$), we revise the cost of assigning operation *j* to position *k* on machine *m* as following:

$$C_{j,k,m} = Max.\{0,\ P_m \cdot k - d_j\},\quad \forall j \in \Phi_A,\quad k = 1, 2, \ldots, K_m$$

$$C_{j,k,m} = Max.\{0,\ P_m \cdot k - (d_j - E_j - P_{j^+,m})\},\ \forall j \in \Phi_B\ j = \{j^- . j^+\},\ k = 1, 2, \ldots, K_m$$

Based on the given $\{C_{j,k,m}\}$, we then solve the respective linear 3D-assignment problem to obtain $\{X_{j,k,m} = 0\ or\ 1 \mid \forall j \in \Phi, k = 1, 2, \ldots, K_m,\ m \in M\}$, which defines

a). The assignment of customer orders to each machine;

b). The sequence of orders on each machine; and

c). The completion time of each operation j, and the estimated returning time from its external contracted processing (if any), tt_j , that

$$tt_j = \sum_{i=1,2,\ldots,k} P^R_{[i],m} + E_j,\ [i] \in \Phi_{B1},\ \textit{that}\ \ X_{[i],i,m} = 1,\ \ X_{j,k,m} = 1,\ [k] = j$$

where $P^R_{[i],m}$ stands for the actual processing time of the *i-th* operation processed by machine *m*.

Step 2. Assign the 2nd operations, $\{ j^+ \mid j \in \Phi_B \}$, to the |M| machines to minimize the total tardiness. This respective optimization problem is, however, NP-Hard.

Lemma 1. *Assigning the 2nd operations,* $\{ j^+ \mid j \in \Phi_B \}$ *, to the |M| machines to minimize the total tardiness is NP-Hard.*

Proof. Consider a special case of problem P1 defined by

$$\Phi_A = \varnothing,\ P^R_{j-,m} = 0,\ P^R_{j+,m} > 0,\ E_j = 0,\ \forall j \in \Phi_B$$

and |M|=1. The reduced Step 2 problem now becomes the one to construct a sequence of $\{j^+ | \forall j \in \Phi_B\}$ jobs on the single-machine to minimize the total tardiness, which has been shown by Du and Leung (1990) to be NP-Hard.

Therefore, we shall heuristically assign the 2nd operations $\{j^+ | \forall j \in \Phi_B\}$ to the M-machine schedule obtained in Step 1 by the following process:

Input: $\{tt_j \mid \forall j \in \Phi_B\}$ and $\{ X_{j^-,k,m} \mid \forall j \in \Phi_A \cup \Phi_B,\ 1 \le k \le K_m,\ m = 1,\ldots, |M| \}$

Step 2-1 Arrange $\{tt_j \mid \forall j \in \Phi_B\}$ into a *non-decreasing* sequence

$$\sigma =< tt_{[i]} \mid \textit{where}\ \ [i] = j^+ \in \Phi_B >$$

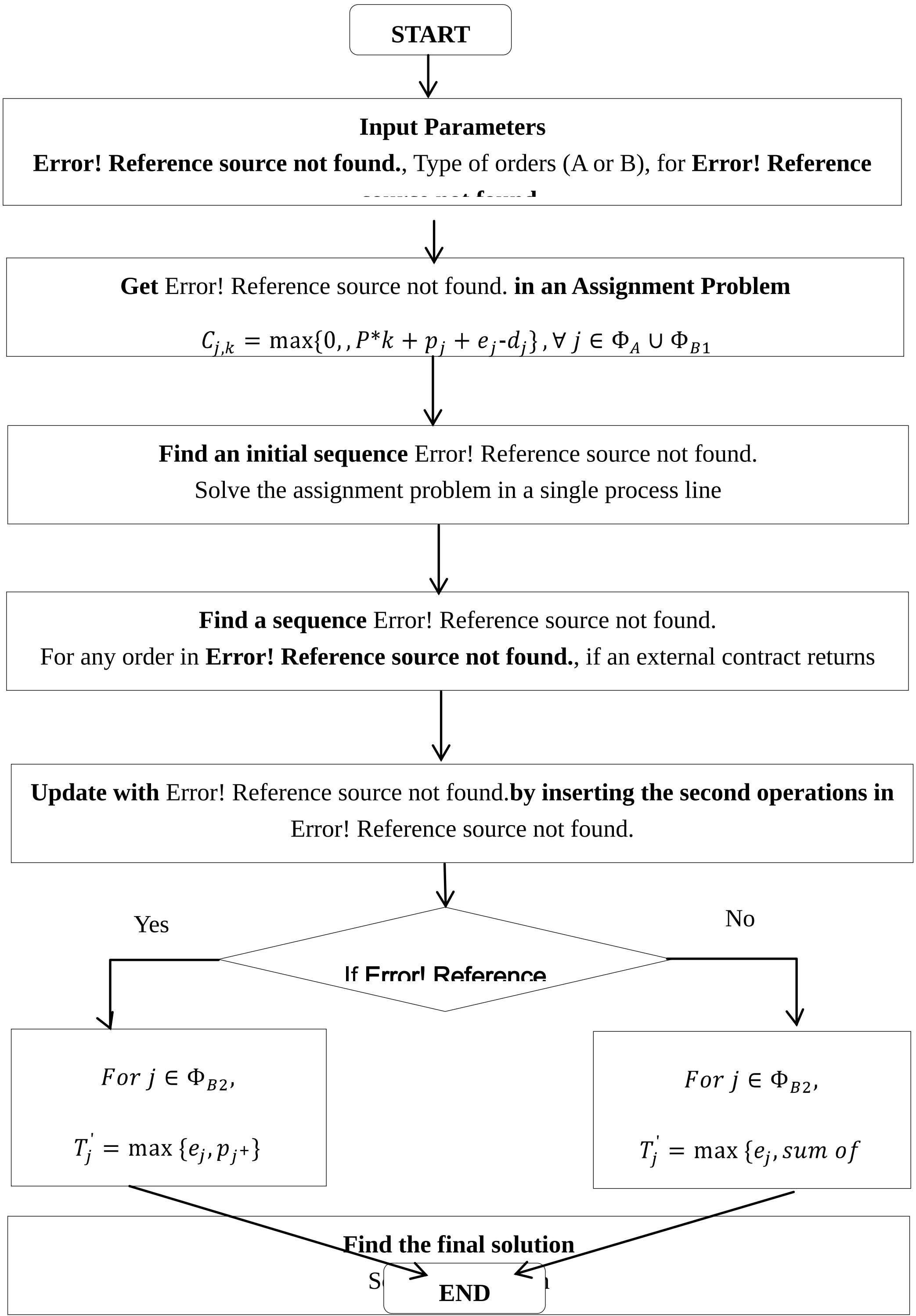
Step 2-2 Remove next job [i] from S and insert it to the M-machine schedule to minimize the increase in total tardiness while preserving $j^{-} \prec j^{+}$, until [i]=$|F_B|$.

Note that the Step 2 may be designed differently, such as choosing the 2nd operation that has the earliest due date (i.e., the most urgent job first) or solving a respective mixed integer program, which will be evaluated in our next step of dissertation research.

6. Preliminary Empirical Results

In this section, we will test a few data sets use the algorithm that we propose in Section 5 to validate if the algorithm can beat the performance of the Python Gurobi Solver with a relatively low error gap and shorter CPU time. Figure 4.3 shows the detailed steps that we will follow to get some preliminary results.

Figure 4.3 Flowchart of the non-interactive algorithm



We will not perform any empirical study in this paper. Instead, we shall demonstrate the step-by-step solution process by this proposed algorithm in deriving the good solution to a special case. In this numerical example, we only consider multiple orders in one processing line and input parameters include each order's weight, due date, external processing time and in-house processing time (for an order in **Error! Reference source not found.**, **Error! Reference source not found.**are involved). For each dataset, we randomly selected some orders with a consideration of production settings in QianFang Shoes. Table 4.1 gives one example of 5 orders with two Type B orders to be produced in one processing line.

Table 4.1 Dataset of 5 customer orders parameters

w_j	d_j	Type of Order	p_j or p_{j^-}	p_{j^+}	e_j
1	35	A	25	0	0
2	25	A	28	0	0
3	19	B	27	18	10
4	29	B	26	20	26
5	40	B	24	20	17

In order to get some unbiased results from numerical examples, for each data set of customer orders, we randomly choose 5 groups sets that consists of 5, 6, 7, 8, 9 and 10 orders. We then compare the results with Python Gurobi Solver with our heuristics, in two dimensions: average/maximum CPU time and also average/maximum error gaps. Table 2 shows the comparisons of two methods and some interesting findings validate that our algorithm is outperform the Gurobi results in both two dimensions.

The experiments results reported in Table 2 highlight the effectiveness of this two-step algorithm. The performance is measured by

$$Error\ Gap = \frac{\left|G^{H}\text{-}G^{*}\right|}{G^{*}}*100\%$$

and **Error! Reference source not found.**, and **Error! Reference source not found.**, stands for the objective function value of the proposed two-step algorithm and the best solution obtained by Gurobi (on Mac Air Interl Core i5 CPU, 1.7 GHz).

Table 4.2 Performance of the heuristic against the order sizes

Order Sizes	**CPU Time (Seconds)**				**Error Gap**		
	Gurobi Optimizer		MIP-based Heuristic				
	Average	Maximum	Average	Maximum	Average	Maximum	Std. Dev.
5	0.158	0.18	0.0032	0.0035	5.84%	11.85%	4.29%
6	0.4040	0.62	0.0051	0.0062	5.49%	8.75%	2.80%
7	2.074	2.67	0.0093	0.0108	7.35%	9.81%	2.23%
8	28.86	42.95	0.0165	0.0229	3.85%	7.83%	2.50%
9	117.1	189.69	0.0251	0.0303	4.19%	7.33%	2.57%
10	696.82	956.59	0.0311	0.035	4.12%	8.70%	2.79%

Each data point in Table 4.2 is the average results from five test cases. From the CPU time and average error gap, it's obvious that the two-step algorithm consistently achieved a better solution quality. That's is, in a shorter amount of time, we are more likely to get near optimal solutions with a very low error gaps less than 7% for a given order sizes from 5 to 10.

Chapter 5. Concluding Remarks

Back to 2010, I was accidentally entering to the world of supply chain and healthcare analytics. From processing data, reading tons of literature, modeling the math problem and coding the very first “Hello, World”, I grew up from a outsider to a insider. Five years of studying supply chain and healthcare analytics, was the biggest treasure in my whole life. Not only because I have a solid understanding of how the data flows and how transfer the real problem into mathematical problem, but also I have established a logic thinking system supported by data. I began to realize the importance of application of the data, management through the data and even plan my own life with the data.

The truth may be hidden in the messy data, however seeking for the truth is a non-stop journey. Inspired with my loved mentors and with strong sense of business world, LineZone Data was born in Hangzhou, China, a company expertized in supply chain optimization and data analytics. Hopefully, what I learnt from Rutgers, will be beneficial to my whole life and also the clients that LineZone Data served.

Thank you, Supply Chain!

Thank you, Rutgers!

Bibliography

1. *Chinitz JS, Vaishnava P, Narayan RL, Fuster V. Atrial fibrillation through the years: contemporary evaluation and management. *Circulation*. 127(3), 408-416 (2013).

2. Kernan WN, Ovbiagele B, Black HR *et al*. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline for healthcare professionals from the american heart association/american stroke association. *Stroke*. 45(7), 2160-2236 (2014).

3. Cheng JW, Barillari G. Non-vitamin K antagonist oral anticoagulants in cardiovascular disease management: evidence and unanswered questions. *J. Clin. Pharm. Ther.* 39(2),118-135 (2014).

4. Visser LE, Penning-van Bees FJ, Kasbergen AA *et al*. Overanticoagulation associated with combined use of antibacterial drugs and acenocoumarol or phenprocoumon anticoagulants. *Thromb. Haemost.* 88(5), 705-710 (2002).

5. Teichert M, van Noord C, Uitterlinden AG *et al*. Proton pump inhibitors and the risk of overanticoagulation during acenocoumarol maintenance treatment. *Br. J. Haematol.* 153(3), 379-385 (2011).

6. van Dijk KN, Plat AW, van Dijk AA *et al*. Potential interaction between acenocoumarol and diclofenac, naproxen and ibuprofen and role of CYP2C9 genotype. *Thromb. Haemost.* 91(1), 95-101 (2004).

7. Knijff-Dutmer EA, van der Palen J, Schut G, Van de Laar MA. The influence of cyclo-oxygenase specificity of non-steroidal anti-inflammatory drugs on bleeding complications in concomitant coumarine users. *QJM*. 96(7), 513-520 (2003).

8. Johnson SG, Witt DM, Eddy TR, Delate T. Warfarin and antiplatelet combination use among commercially insured patients enrolled in an anticoagulation management service. *Chest.* 131(5), 1500-1507 (2007).

9. Johnson SG, Rogers K, Delate T, Witt DM. Outcomes associated with combined antiplatelet and anticoagulant therapy. *Chest.* 133(4), 948-954 (2008).

10. Shikata E, Ieiri I, Ishiguro S *et al*. Association of pharmacokinetic (CYP2C9) and pharmacodynamic (factors II, VII, IX, and X; proteins S and C; and gamma-glutamyl carboxylase) gene variants with warfarin sensitivity. *Blood*. 103(7), 2630-2635 (2004).

11. D'Andrea, G, D'Ambrosio RL, Perna PD *et al*. A polymorphism in the VKORC1 gene is associated with an interindividual variability in the dose-anticoagulant effect of warfarin. *Blood*. 105 (2), 645-649 (2005).

12. Sconce EA, Khan TI, Wynne HA *et al*. The impact of CYP2C9 and VKORC1 genetic polymorphism and patient characteristics upon warfarin dose requirements: proposal for a new dosing regimen. *Blood* . 106(7), 2329-2333 (2005).

13. Schwarz UI, Ritchie MD, Bradford Y *et al*. Genetic determinants of response to warfarin during initial anticoagulation. *N. Engl. J. Med.* 358(10), 999-1008 (2008).

14. Takahashi H, Wilkinson GR, Nutescu EA *et al*. Different contributions of polymorphisms in VKORC1 and CYP2C9 to intra- and inter-population differences in maintenance dose of warfarin in Japanese, Caucasians and African-Americans. *Pharmacogenet. Genomics.* 16(2), 101-110 (2006).

15. Cooper GM, Johnson JA, Langaee TY *et al*. A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood*. 112(4), 1022-1027 (2008).

16. Wadelius M, Chen LY, Lindh JD *et al*. The largest prospective warfarin-treated cohort supports genetic forecasting. *Blood*. 113(4), 784-792 (2009).

17. Nutescu EA, Shapiro NL, Ibrahim S, West P. Warfarin and its interactions with foods, herbs and other dietary supplements. *Expert Opin. Drug Saf.* 5(3), 433-451 (2006).

18. Ford SK, Moll S. Vitamin K supplementation to decrease variability of international normalized ratio in patients on vitamin K antagonists: A literature review. *Curr. Opin. Hematol.* 15(5), 504-508 (2008).

19. Holbrook A. Schulman S, Witt DM *et al*. Evidence-based management of anticoagulant therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 141(2), e152S–84S (2012).

20. Wysowski DK, Nourjah P, Swartz L. Bleeding complications with warfarin use: a prevalent adverse effect resulting in regulatory action. *Arch. Intern. Med.* 167(13), 1414-1419 (2007).

21. Torn M, Cannegieter SC, Bollen WL, van der Meer FJ, van der Wall EE, Rosendaal FR. Optimal level of oral anticoagulant therapy for the prevention of arterial thrombosis in patients with mechanical heart valve prostheses, atrial fibrillation, or myocardial infarction: a prospective study of 4202 patients. *Arch. Intern. Med.* 169(13), 1203-1209 (2009).

22. Gage BF, Fihn SD, White RH. Management and dosing of warfarin therapy. *Am. J. Med.* 109(6), 481-488 (2000).

23. January CT, Wann LS, Alpert JS *et al*. AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. *J. Am. Coll. Cardiol*. 60, 1740-1749 (2014).

24. Witt DM, Delate T, Clark NP *et al*. Twelve-month outcomes and predictors of very stable INR control in prevalent warfarin users. *J. Thromb. Haemost.* 8(4),744-749 (2010).

25. van Walraven C, Forster AJ. Anticoagulation control in the peri-hospitalization period. *J. Gen. Intern. Med.* 22(6), 727-735 (2007).

26. Witt DM, Delate T, Clark NP *et al*. Outcomes and predictors of very stable INR control during chronic anticoagulation therapy. *Blood* . 114(5), 952-956 (2009).

27. Kimmel SE, Chen Z, Price M *et al*. The influence of patient adherence on anticoagulation control with warfarin: results from the International Normalized Ratio Adherence and Genetics (IN-RANGE) Study. *Arch. Intern. Med.* 167(3), 229-235 (2007).

28. Platt AB, Localio AR, Brensinger CM *et al*. Can we predict daily adherence to warfarin?: Results from the International Normalized Ratio Adherence and Genetics (IN-RANGE) Study. *Chest.*137(4), 883-889 (2010).

29. Rose AJ, Ozonoff A, Berlowitz DR, Henault LE, Hylek EM. Warfarin dose management affects INR control. *J. Thromb. Haemost.* 7(1), 94-101(2009).

30. Rose AJ, Hylek EM, Ozonoff A, Ash AS, Reisman JI, Berlowitz DR. Patient characteristics associated with oral anticoagulation control: results of the Veterans AffaiRs Study to Improve Anticoagulation (VARIA). *J. Thromb. Haemost.* 8(10), 2182-2191 (2010).

31. Ansell J, Hirsh J, Hylek E *et al*. Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest*. 133(6), 160S-198S (2008).

32. *Matchar DB, Samsa GP, Cohen SJ, Oddone EZ, Jurgelski AE. Improving the quality of anticoagulation of patients with atrial fibrillation in managed care organizations: results of the managing anticoagulation services trial. *Am. J. Med.* 113(1), 42-45 (2002).

33. Mitra R, Marciell MA, Brain C, Ahangar B, Burke DT. Efficacy of computer-aided dosing of warfarin among patients in a rehabilitation hospital. *American Journal of Physical Medicine & Rehabilitation*. 84(6), 423-427 (2005).

34. Ageno W, Johnson J, Nowacki B, Turpie AG. A computer generated induction system for hospitalized patients starting on oral anticoagulant therapy. *Thrombosis & Haemostasis*. 83(6), 849-852 (2000).

35. Van Spall HG, Wallentin L, Yusuf S *et al*. Variation in warfarin dose adjustment practice is responsible for differences in the quality of anticoagulation control between centers and countries: an analysis of patients receiving warfarin in the randomized evaluation of long-term anticoagulation therapy (RE-LY) trial. *Circulation.* 126, 2309-2316 (2012).

36. Jones M, McEwan P, Morgan CL, Peters JR, Goodfellow J, Currie CJ. Evaluation of the pattern of treatment, level of anticoagulation control, and outcome of treatment with warfarin in patients with non-valvar atrial fibrillation: a record linkage study in a large British population. *Heart.* 91(4), 472-477 (2005).

37. Wallentin L, Yusuf S, Ezekowitz MD *et al*. Efficacy and safety of dabigatran compared with warfarin at different levels of international normalised ratio control for stroke prevention in atrial fibrillation: an analysis of the RE-LY trial. *Lancet.* 376(9745), 975-983 (2010).

38. *Reynolds KK, Valdes R, Hartung BR, Linder MW. Individualizing warfarin therapy. *Personalized Medicine*. 4, 11-31 (2007).

39. *Holford NHG, Sheiner LB. Understanding the dose-effect relationship: Clinical application of pharmacokinetic-pharmacodynamic models. *Clinical Pharmacokinetics*. 6(6), 429-453 (1981).

40. **Holford NHG. Clinical pharmacokinetics and pharmacodynamics of warfarin. Understanding the dose-effect relationship. *Clinical Pharmacokinetics*. 11(6), 483-504 (1986).

41. *Wright DFB, Duffull SB. Development of a Bayesian forecasting method for warfarin dose individualization. *Pharm. Res.* 28(5), 1100-1111 (2011).

42. *Pasterkamp E, Kruithof CJ, van der Meer FJ, Rosendaal FR, Vanderschoot JP. A model-based algorithm for the monitoring of long-term anticoagulation therapy. *Journal of Thrombosis and Haemostasis*. 3(5), 915-921 (2005).

43. *Vadher B, Patterson DLH, Leaning M. Prediction of the international normalized ratio and maintenance dose during the initiation of warfarin therapy. *Br. J. Clin. Pharmacol.* 48, 63-70 (1999).

44. Carter BL, Taylor JW, Becker A. Evaluation of three dosage-prediction methods for initial in-hospital stabilization of warfarin therapy. *Clinical Pharmacy*. 6(1), 37-45 (1987).

45. White RH, Hong R, Venook AP *et al*. Initiation of warfarin therapy – comparison of physician dosing with computer-assisted dosing. *Journal of General Internal Medicine*. 3, 141-148 (1987).

46. *Wright DFB, Duffull SB. A Bayesian dose-individualization method for warfarin. *Clin. Pharmacokient.* 52(1), 59-68 (2013).

47. White RH, Mungall D. Outpatient management of warfarin therapy: comparison of computer-predicted dosage adjustment to skilled professional care. *Therapeutic Drug Monitoring*. 13(1), 46-50 (1991).

48. *Hamberg AK, Wadelius M, Lindh JD *et al*. A pharmacometic model describing the relationship between warfarin dose and INR response with respect to variations in CYP2C9, VKORC1, and age. *Clin. Pharmacol. Ther.* 87(6), 727-734 (2010).

49. Shi L, Olafsson S. Nested partitions method for global optimization. *Operations Research*. 48(3), 390-407 (2000).

50. Shi L, Olafsson S. *Nested Partitions Method, Theory and Applications*. Springer Science + Business Media, 2009.

51. Sheiner LB, Beal SL. Some suggestions for measuring predictive performance. *Journal of pharmacokinetics and biopharmaceutics*. 9(4), 503-512 (1981).

52. Brun, A., F. Caniato, M. Caridi, C. Castelli, G. Miragliotta, S. Ronchi, A. Sianesi, G. Spina, 2008. Logistics and supply chain management in the luxury fashion retail: Empirical investigation of Italian firms. *International Journal of Production Economics* 114 (2), 554–570.

53. Choi, B.C., J.B. Chung, 2011. Two-machine flow shop scheduling problem with an outsourcing option. European Journal of Operational Research 213, 66–72.

54. Caniato, F., M. Caridi, C. Castelli, R. Golini. 2011. Supply chain management in the luxury industry: a first classification of companies and their strategies. *International Journal of Production Economics* 133, 622–633.

55. Chung, D. Y., B. C. Choi. 2013. “Outsourcing and Scheduling for Two-machine Ordered Flow Shop Scheduling Problems.” *European Journal of Operational Research* 226 (1), 46–52.

56. Du, J., J.Y.T. Leung. 1990. Minimizing total tardiness on one machine is NP-hard. Mathematics of Operations Research 15, 483–495.

57. Ferdows, K., M. A. Lewis, J. A. Machuca. 2004. Rapid-Fire Fulfillment. *Harvard Business Review*, 82 (11).

58. Lee, C.Y., L. Lei, 2001. Multiple-project scheduling with controllable project duration and hard resource constraint: Some solvable cases. *Annals of Operations Research* 102, 287–307.

59. Lee, I. S., C. S. Sung. 2008a. Single machine scheduling with outsourcing allowed. *International Journal of Production Economics* 111, 623–634.

60. Lee, I. S., C. S. Sung. 2008b. “Minimizing Due Date Related Measures for a Single Machine Scheduling Problem with Outsourcing Allowed.” *European Journal of Operational Research* 186 (3), 931–952.

61. Lee, K., B.C. Choi, 2011. Two-stage production scheduling with an outsourcing option. *European Journal of Operational Research*, 213 (3), 489–497.

62. Lee, Y.H., C.S. Jeong, C. Moon. 2002. Advanced planning and scheduling with outsourcing in manufacturing supply chain. *Computer & Industrial Engineering* 43, 351–374.

63. Lee, K., K. Kim, K. Shin, J. Park. 1997. A scheduling algorithm considering subcontract cost in MRP environment. *Proceedings of the 14th International Conference on Production Research*, Osaka, Japan, pp. 1068–1071.

64. Lee, Y.H., C.S. Jeong, and C. Moon, 2002. Advanced planning and scheduling with outsourcing in manufacturing supply chain. *Computers and Industrial Engineering*, 43 (1–2), 351–374.

65. Kahraman, C., A. Beskese, and I. Kaya, 2010. Selection among ERP outsourcing alternatives using a fuzzy multi-criteria decision making methodology. International Journal of Production Research, 48 (2), 547–566.

66. Kremic, T., O.I. Take, and W.O. Rom, 2006. Outsourcing decision support: a survey of benefits, risks and decision factors. *Supply Chain Management: An International Journal*, 11 (6), 467–482.

67. Kuhn, H.W. 1955. The Hungarian method for the assignment problem, *Naval Research Logistics Quarterly* 2 (1) 83–97

68. Mokhtari, H., I. N. Kamal Abadi, M. R. Amin-Naseri. 2012. “Production Scheduling with Outsourcing Scenarios: A Mixed Integer Programming and Efficient Solution Procedure.” *International Journal of Production Research* 50 (19), 5372–5395.

69. Öncan, T. 2007. “A Survey of the Generalized Assignment Problem and Its Applications,” *Information Systems and Operational Research*, 45 (3), 123-142

70. Qi, X. T. 2008. “Coordinated Logistics Scheduling for In-house Production and Outsourcing.” *IEEE Transactions on Automation Science and Engineerin*g 5 (1): 188–192.

71. Qi, X. T. 2009. “Two-stage Production Scheduling with an Option of Outsourcing from a Remote Supplier.” *Journal of System Science and System Engineering* 18 (1): 1–15.

72. Qi, X. T. 2011. “Outsourcing and Production Scheduling in a Two-stage Flow Shop.” *International Journal of Production Economics* 129 (1): 43–50.

73. Shmoys, D.B. and Tardos, E. (1993) An approximation algorithm for the generalized assignment problem. *Mathematical Programming*, 62 (3), 461–474.

74. Wiendahl, H.P., 1995. Load-oriented manufacturing control. Berlin, Germany: *Springer.*