NEW PERSPECTIVES ON THE BULLWHIP EFFECT

by
ÖZLEM ÇOBAN

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Master of Science

SABANCI UNIVERSITY
Fall 2010

NEW PERSPECTIVES ON THE BULLWHIP EFFECT

APPROVED BY:

Assist. Prof. Dr. Murat KAYA ……………………..
(Thesis Supervisor)

Assist. Prof. Dr. Tevhide ALTEKİN ……………………..

Assist. Prof. Dr. Gürdal ERTEK ……………………..

Assist. Prof. Dr. Çağrı HAKSÖZ ……………………..

Assist. Prof. Dr. Kemal KILIÇ ……………………..

DATE OF APPROVAL: ………………………………………………………….

NEW PERSPECTIVES ON THE BULLWHIP EFFECT

Özlem ÇOBAN

Industrial Engineering, MSc Thesis, 2010

Thesis Supervisor: Assist. Prof. Dr. Murat KAYA

Keywords: Bullwhip effect, beer game experiments, behavioral operations, supply chain management.

Abstract

In this thesis, we propose a modified version of the Beer Game with two participants at each echelon that have conflicting incentives regarding the order decision. One participant (the sales manager) has backorder cost as his performance measure, whereas the other (the supply manager) has inventory holding cost. We conducted beer game experiments with human participants using the modified and standard game settings. We find that the conflict in the modified game, which reflects the sales/operations conflict in real firms, can dampen the bullwhip effect. We also develop multiple linear regression models to explain participants' order decisions based on factors including incoming demand, backlogs, on-hand inventory levels and outstanding orders. Overall, we identify "supply risk" as an important cause of the bullwhip effect.

KAMÇI ETKİSİ ÜZERİNE YENİ BAKIŞ AÇILARI

Özlem ÇOBAN

Endüstri Mühendisliği, Yüksek Lisans Tezi, 2010

Tez Danışmanı: Yrd. Doç. Dr. Murat KAYA

Anahtar Kelimeler: Kamçı etkisi, bira oyunu deneyleri, davranışsal operasyonlar, tedarik zinciri yönetimi

Bu tezde, her seviyesinde çıkarları birbiriyle çelişen iki oyuncunun bulunduğu modifiye bir "Bira Oyunu Deneyi" üzerinde çalıştık. Bu oyunculardan birinin (satış müdürü) performans ölçütünü bekleyen sipariş maliyeti, diğerinin performans ölçütünü ise stok bulundurma maliyeti olarak belirledik. Modifiye ve standart bira oyununu katılımcılara oynatarak sonuçları karşılaştırdık. Gerçek şirketlerin satış ve operasyon departmanları arasında gözlemlenen çıkar çatışmasını yansıtan modifiye oyunun kamçı etkisini düşürdüğünü gözlemledik. Çalışmamızda ayrıca, oyuncuların sipariş miktarlarını gelen talep, bekleyen sipariş, eldeki stok ve tedarik sürecindeki ürünler gibi faktörler kullanarak tahmin etmeyi amaçlayan çoklu doğrusal regresyon modelleri geliştirdik. Özellikle "tedarik riski" faktörünün kamçı etkisinin önemli bir sebebi olduğunu gözlemledik.

ACKNOWLEDGEMENTS

It is a great pleasure to extend my gratitude to my thesis advisor Assist. Prof. Dr. Murat Kaya for his precious guidance and support. I am greatly indebted to him for his supervision and excellent advice throughout my master study.

I would gratefully thank Assist. Prof. Dr. Tevhide Altekin, Assist. Prof. Dr. Gürdal Ertek, Assist. Prof. Dr. Çağrı Haksöz and Assist. Prof. Dr. Kemal Kılıç for their feedback and for spending their valuable time to serve as my jurors.

I would like to acknowledge the stipend support provided by TÜBİTAK under the BİDEB scholarship, and Sabanci University for waiving the tuition throughout my master study.

My sincere thanks go to all my friends from Sabancı University. In particular, I would like to express my thanks to Mahir, Nimet, Gizem, Merve, Elif, Ezgi, Semih and Nükte and Taner.

I would like to thank my family for all their love and support throughout my life. Finally, I wish to express my deepest gratitude to Hakan Ertaş for providing me the necessary motivation and being my source of strength and happiness in the hardest/most stressful times.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION AND MOTIVATION 1
CHAPTER 2 LITERATURE SURVEY 8
2.1 STUDIES ON OPERATIONAL CAUSES AND REMEDIES 9
2.1.1 Literature on Operational Efficiency 9
2.1.2 Literature on Information Sharing 10
2.1.3 Literature on Channel Alignment (Strategic Alliances) 13
2.2 STUDIES ON BEHAVIORAL CAUSES 15
2.3 OPERATIONS AND SALES INCENTIVE CONFLICT 20
2.4 GROUP VERSUS INDIVIDUAL DECISION MAKING 20
2.5 MEASUREMENT OF THE BULLWHIP EFFECT 22
CHAPTER 3 THE FIRST STUDY: BEER GAME WITH TWO PARTICIPANTS AT EACH ECHELON 24
3.1 EXPERIMENTAL DESIGN AND IMPLEMENTATION 24
3.2 EXPERIMENTAL RESULTS AND ANALYSIS 30
3.2.1 Outlier Analysis 30
3.2.2 Preliminary Observations 34
3.2.3 Comparison of the Standard and the Modified Experiments 37
3.2.3.1 Oscillation Comparison 39
3.2.3.2 Amplification Comparison 40
3.2.3.3 Time Lag Comparison 41
3.2.3.4 Mean Order Comparison 43
3.2.3.5 Cost Comparison 44
3.2.3.6 Analysis with Median Values 50
CHAPTER 4 THE SECOND STUDY: DETERMINING THE BEHAVIORAL FACTORS AFFECTING ORDER DECISIONS 53
4.1 THE CANDIDATE FACTORS 55
4.2 THE REGRESSION MODELS 56
4.2.1 Observations on Model 3 58
4.2.2 Observations on Model 11 60
4.2.3 Observations on Stepwise Regression Models 62
CHAPTER 5 CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH 63

BIBLIOGRAPHY 131

LIST OF FIGURES

Figure 1-1: A Typical Supply Chain ... 1
Figure 1-2: Order and Inventory Levels over Time ... 2
Figure 3-1: The Beer Game ... 25
Figure 3-2: One of Our Experiments ... 28
Figure 3-3: Sample Box Plot ... 32
Figure 3-4: Box Plot of Variance of Orders for Standard Experiments ... 32
Figure 3-5: Orders Placed over Periods of the Experiment (for team 39) ... 34
Figure 3-6: Effective Inventory Levels over Periods (for team 39) ... 35
Figure 3-7: Order and Effective Inventory Levels of the Echelons (for team 39) ... 36
Figure 3-8: Order Variances in the Standard Experiments ... 38
Figure 3-9: Order Variances in the Modified Experiments ... 38
Figure 3-10: Orders Placed by Each Retailer in Standard Experiments ... 47
Figure 3-11: Orders Placed by Each Retailer in Modified Experiments ... 48
Figure 4-1: Predicted Customer Demand Drawn by One of Factory Participants ... 54
Figure 5-1: Record Sheet of One of Our Participants ... 67
Figure 5-2: Post Experiment Survey ... 70
Figure 5-3: Box Plot of Amplification Ratios for Standard Experiments ... 88
Figure 5-4: Box Plot of Order Variances for Modified Experiments ... 88
Figure 5-5: Box Plot of Amplification Ratios for Standard Experiments ... 89
Figure 5-6: Normal *P-P* Plot of Residuals of One of Our Regressions ... 122

LIST OF TABLES

Table 1-1: Causes of the Bullwhip Effect.......... 3
Table 2-1: Channel Alignment through Strategic Alliances.......... 14
Table 2-2: Categorizing the Literature.......... 15
Table 2-3: Types of Measures 23
Table 3-1: Design of Experiments.......... 28
Table 3-2: Order Variance Comparison.......... 40
Table 3-3: *P* Values of Hypothesis Tests for Order Variances.......... 40
Table 3-4: Amplification Ratio Comparison 41
Table 3-5: *P* Values of Hypothesis Tests for Amplification Ratios 41
Table 3-6: Peak Orders Comparison.......... 42
Table 3-7: Peak Backlogs Comparison.......... 42
Table 3-8: First Backlogs Comparison 43
Table 3-9: Mean Order Comparison.......... 43
Table 3-10: *P* Values of Hypothesis Tests for Mean Orders.......... 44
Table 3-11: Cost Comparison.......... 45
Table 3-12: *P* Values of Hypothesis Tests for Total Costs.......... 49
Table 3-13: *P* Values of Hypothesis Tests for Inventory Costs 49
Table 3-14: *P* Values of Hypothesis Tests for Backlog Costs.......... 50
Table 3-15: Median Order Variance Comparison 51
Table 3-16: Median Amplification Ratio Comparison.......... 51
Table 3-17: Median Peak Orders Comparison 51
Table 3-18: Median Peak Backlogs Comparison 51
Table 3-19: Median First Backlogs Comparison.......... 52
Table 3-20: Median Orders Comparison 52
Table 3-21: Median Cost Comparison.......... 52
Table 4-1: Regression Models Summary.......... 57
Table 4-2: Standardized Beta Coefficients for Model 3 59
Table 4-3: Standardized Beta Coefficients for Model 11 61
Table 5-1: Participants Information.......... 68
Table 5-2: Attitude towards Risk and Service 68
Table 5-3: Order Variances.......... 90
Table 5-4: Amplification Ratios 91

Table 5-5: The Period of Peak Order Levels ... 92
Table 5-6: Peak Order Magnitudes ... 93
Table 5-7: The Period of the Peak Backlog Level ... 94
Table 5-8: Peak Backlog Magnitudes ... 95
Table 5-9: First Backlog Periods ... 96
Table 5-10: Mean Orders ... 97
Table 5-11: Mean Inventory Costs ... 98
Table 5-12: Mean Backlog Costs ... 99
Table 5-13: Mean Total Costs ... 100
Table 5-14: Inventory Cost Variances ... 101
Table 5-15: Backlog Cost Variances ... 102
Table 5-16: Total Cost Variances ... 103
Table 5-17: Results for the Supply Chain (R, W, D, F) ... 104
Table 5-18: Results for Downstream Echelons (R, W) ... 106
Table 5-19: Results for Upstream Echelons (D, F) ... 108
Table 5-20: Results for Retailer Echelons ... 110
Table 5-21: Results for Wholesaler Echelons ... 112
Table 5-22: Results for Distributor Echelons ... 114
Table 5-23: Results for Factory Echelons ... 116
Table 5-24: Results for Amplification Ratio Comparisons ... 118
Table 5-25: Regression Results for Model 3 ... 123
Table 5-26: Regression Results for Model 11 ... 125
Table 5-27: Results for SRM1 ... 127
Table 5-28: Results for SRM2 ... 129

Chapter 1

Introduction and Motivation

“Supply chain, which is also referred to as the logistics network, consists of suppliers, manufacturing centers, warehouses, distribution centers and retail outlets, as well as raw materials, work-in-process inventory and finished products that flow between facilities” (Simchi-Levi et al. 2007). Figure 1-1 illustrates a typical supply chain with four echelons: retailer, wholesaler, distributor and factory. Each echelon’s ordering decision affects the performance and profit of the other echelons. This situation leads managers to face major, real time difficulties in managing dynamic systems. In the process of decision making, across all echelons of the supply chain, managers may deviate from optimal or rational decisions. Managers, being individuals, possess unique human attributes which effect their decision making process.

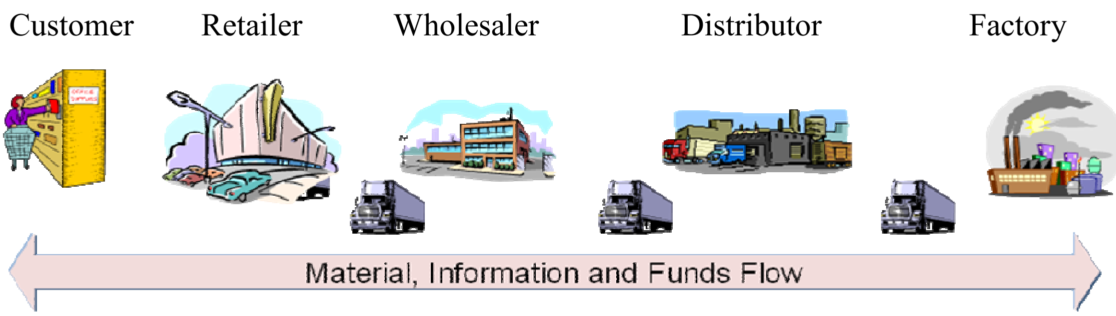


Figure 1-1: A Typical Supply Chain[1]

“Bullwhip effect” defines order variability increases when one goes from “downstream echelons” (i.e., the echelons closer to end customers) of a supply chain to “upstream echelons” (i.e., the echelons closer to raw material sources). Forrester (1958) first identified the effect, but did not refer to it with the term “bullwhip effect”. Croson and

[1] Simchi-Levi et al. (2007)

Donohue (2003) state that the effect is described by *oscillation, amplification and time lag*. As seen in Figure 1-2, oscillations of orders mean that at each supply chain echelon, fluctuation occurs over time. Amplification means that when one goes from downstream to upstream echelons, oscillations increase. Time lag means that amplifications of oscillations propagate with a time lag when one goes from downstream to upstream echelons.

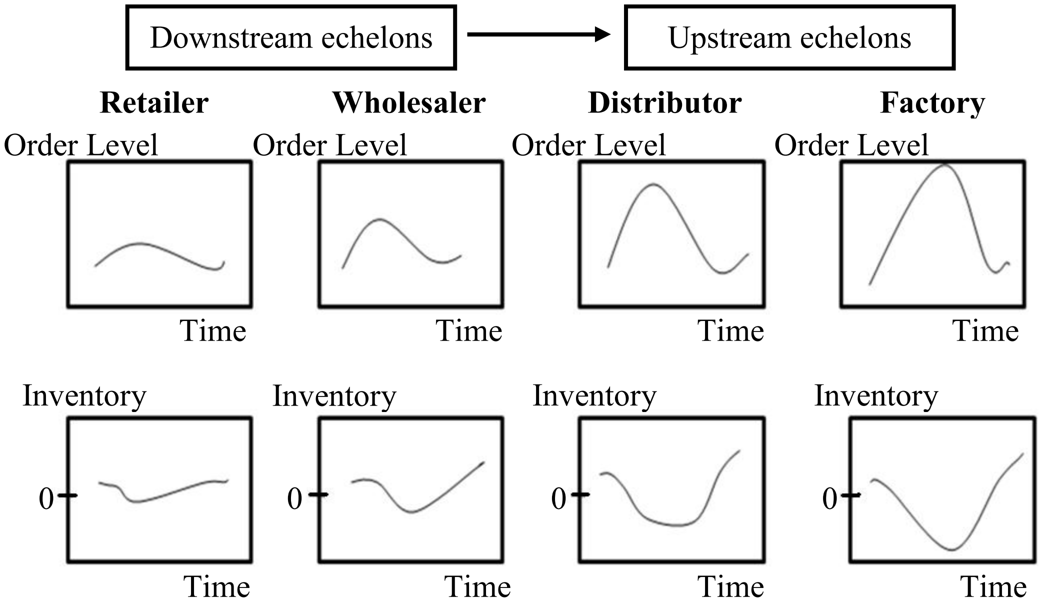


Figure 1-2: Order and Inventory Levels over Time

The term "bullwhip effect" was first coined by Procter & Gamble (P&G) in 1990s (Lee et al. 1997a). The company observed that the diaper orders given by the distributors exhibit a degree of variability that cannot be explained by consumer demand fluctuations alone. Likewise, Hewlett-Packard (HP) observed that the orders placed to the printer division by resellers have a much higher variation than the variation in customer demands (Lee et al. 1997b). Other examples include Eli Lilly and Bristol-Myers Squibb from pharmaceutical industry (Lee et al. 1997b), and Barilla SpA from pasta industry (Hammond 2008). Chen and Lee (2010) reports that bullwhip effect is observed in automobile (Blanchard 1983), cement (Ghali 1987), basic metal (Fair 1989), perishable foods (Fransoo and Wouters 2000) and electronics (De Kok et al. 2005) industries. Bullwhip effect was also known to be a major reason behind Cisco's well-known $2.2 billion inventory write-off in 2001.[2] A recent (January 27, 2010) Wall Street Journal article about Caterpillar, the world's largest manufacturer of construction and mining machines, illustrates that the bullwhip effect continues to affect supply chains even today.[3]

As these industry examples and theoretical studies (for example, Machuca and Barajas 2004, Metters 1997, Disney and Lambrecht 2007, Munson et al. 2003) illustrate, the bullwhip effect causes high supply chain costs. This is because each firm observes high variability in its demand, leading to difficulties in forecasting and production planning. Firms need to have extra capacity and hold extra inventory in order to accommodate high demand variation. In the end, as seen in Figure 1-2, inventory shortages or excess inventory occurs, and utilization level of workers and equipment will be low. Consequently, reduction of the bullwhip effect in a supply chain is critical for its performance.

Two main groups of causes can explain occurrence of the bullwhip effect. One group refers to operational causes; while the other group refers to behavioral causes as briefly listed in Table 1-1 (Lee et al. 1997a, Croson and Donohue 2006).

Table 1-1: Causes of the Bullwhip Effect

Operational Causes	**Behavioral Causes**
Demand signal processing Order batching Rationing game Price fluctuations	Visibility of supply chain Coordination problem Underweighting the supply line Psychology of decision makers

Lee et al. (1997a) determine the four common "operational causes" of the bullwhip effect as demand signal processing, order batching, rationing game, and price variations. Demand signal processing means that managers use past demand information to update their forecasts. That is, if demand goes up in a time, it is used as a signal of forthcoming high demands in forecasting. Order batching means that managers have a tendency to

---

2 http://www.cio.com/article/30413/What_Went_Wrong_at_Cisco_in_2001

3 http://online.wsj.com/article/SB10001424052748704509704575019392199662672.html

batch orders if fixed ordering and transportation costs are nonzero. When supply shortage is anticipated in the chain, the strategic ordering behavior of buyers is referred to as shortage gaming. In the case of expected shortages, if the supplier allocates products to buyers in proportion to the order of each buyer, buyers order more than they need to achieve the actual quantity they need. Price fluctuations are generally results of promotions on the purchasing prices of products. When there is a promotion, the buyers tend to order more than needed, which is also known as forward buying. These factors cause sudden increases or decreases in order levels, which causes fluctuation.

In addition to operational causes, the bullwhip effect is also known to have "behavioral causes" that are related to human decision-making in dynamic systems. These were first mentioned by Forrester (1958). Then, Sterman (1989a) explained the main behavioral reasons of the bullwhip effect as "misperceptions of feedback" and "participants' tendency to underweight the supply line". Misperception of feedback means that when decisions have delayed and indirect effects on each other, participants find it challenging the control the dynamics. Underweighting the supply line means that participants often undervalue the orders that were previously made and that are still in the supply line. Consequently, they place higher orders than necessary.

The bullwhip effect can be observed in the well-known "Beer (distribution) Game" experiments. The beer game was invented by Sloan's system dynamics group in the early 1960s as part of Jay Forrester's research on industrial dynamics.[4] Sterman (1989a) was the first to use the beer game to test the existence of the bullwhip effect in an experimental context. The standard beer game experiments (see Chapter 3 for details) are played by four participants, representing four echelons of a beer supply chain similar to the one presented in Figure 1-1. Each participant determines how much to order from his upstream echelon at each period. The orders arrive at the upstream echelon after a specific "ordering delay", and that echelon fulfils the order if he has sufficient inventory on hand. Unmet order is backlogged. The shipments arrive at the requesting echelon after a "shipping delay".

[4] http://web.mit.edu/jsterman/www/SDG/beergame.html

In his ordering decisions, the participant at each echelon faces the fundamental trade-off between over-ordering and under-ordering. At the end of each period:

- If he has inventory on hand, he incurs an inventory holding cost.
- If he has backlog, he incurs a backlog cost.

Each participant's individual performance measure is the total inventory holding and backlog costs over all periods. This requires him to strike a balance between the two sides of the trade-off. However, the time lag due to the ordering and shipping delays (which is 4 periods in the standard beer game) makes it difficult to handle the trade-off. In addition, supply is not guaranteed. If the upstream echelon does not have sufficient inventory on hand when the order arrives, he will not be able to meet the order. The time lag and supply uncertainty make it difficult to judge the trade-off. Due to the operational and behavioral factors we mentioned, participants generally over-order. This over-ordering propagates through the supply chain, leading to the bullwhip effect.

Given this discussion, the main research question we ask in this thesis is: Can the bullwhip effect be mitigated, if there exists two participants at each echelon whose performance measures represent the two sides of the trade-off ?

To address this question, we conducted a *modified version of the beer game* in which there are two participants at each echelon with the following roles:

- The supply manager whose performance measure is the inventory holding cost.
- The sales manager whose performance measure is the backlog cost.

At each period, these two participants make a single joint order decision for their echelon. Note that the two participants have conflicting incentives. The supply manager would prefer lower order quantities leading to lower inventory holding cost, whereas the sales manager would prefer higher order quantities leading to lower backlog cost (due to higher product availability). With focused incentives and different performance measures representing the two sides of the trade-off, we expect the order decisions in this modified beer game to cause less bullwhip effect than a standard game. For instance, because the supply manager's performance is measured solely on the inventory holding cost, he would react to "over-ordering" attempts of the other participant. Likewise, the supply manager is more likely to keep track of orders that are in the pipeline. The sales manager, on the other hand, can better focus on forecasting. Our modified beer game captures the well-known operations/ sales conflict observed in real firms. In a firm, an operations manager aims to match supply with demand by deciding how much of a product to supply, whereas a sales manager aims to create and satisfy customer demand. Firms perceive the operations department as a cost center and the sales department as a revenue center (Jerath et al. 1997, Harps 2002). Hence, the incentive of operations people are towards cutting costs by minimizing inventories, whereas the incentive of the sales people is towards increasing revenue by having sufficient stock on hand (Ackoff 1967, Oliva and Watson 2007). The performance measures of the operations and sales managers reflect these incentives.

The joint decision making process at each echelon of our modified game is somehow similar to the sales and operations planning processes (S&OP) applied by firms. S&OP refers to the integrated supply chain management planning process across all departments of a firm. Despite having incentive conflicts, sales, operations and finance departments regularly hold meetings to update sales plans, customer demand forecasts, inventory plans or other strategic plans together. In meetings, making forecast decisions together with shared information increases the trust among the departments and improves the demand forecast accuracy of the firm.

When two human beings make a joint decision, one needs to consider the “group decision making” dynamics. We mention related research in Section 2.4. The two participants in our modified beer game experiments have conflicting incentives and they need to come to an agreement at each period. Another aspect of having two participants at each echelon is that “Two heads are better than one”. That is, one might expect improvements in the beer game outcomes when the single decision maker is replaced with two decision makers simply because two people can make better decisions. This may be because of their higher total “attention” or “intelligence”. To analyze this effect in isolation, one can design an experiment with two participants at each echelon that share the same performance measure of total inventory holding and backup costs minimization (i.e., no different roles, and no incentive conflict). We leave this to further study. In this thesis, our objective is to observe the joint effects of “incentive conflict” and “two heads better than one” factors.

In the second study we report, we aim to determine the behavioral factors that affect the ordering decision of the participants in the standard beer game experiments. Given the role of the behavioral factors, we wanted to assess their relative magnitudes in participant's decision making. The factors that we consider include the on hand inventory (or backlog) level, whether the echelon is in backlog or not, the demand faced at the period, outstanding order quantity, whether there is an increase in demand over the last two periods, and whether the upstream firm has been able to meet previous orders. We conduct multiple linear regression analysis to determine how much weight, if any, the participants place on such factors in determining their order quantity in a period.

This thesis is organized as follows: In Chapter 1, we discussed the causes and the consequences of the bullwhip effect and we explain the beer game experiments. Next, in Chapter 2, we provide a review of the related literature. In Chapter 3, we first explain the beer game experiment procedure. We then present our experimental data analysis, focusing on the comparison between the standard and modified beer games. In Chapter 4, using regression analysis, we analyze the behavioral factors affecting the participants' ordering decisions. We discuss the implications of our work, conclude and provide future research directions in Chapter 5.

Chapter 2

Literature Survey

The bullwhip effect has been studied extensively using empirical, theoretical and experimental methods. In empirical studies, researchers generally collect industry level sales and inventory data to measure the strength of the bullwhip effect. In theoretical studies, researchers quantify and generalize the effects of causes and improvements of proposed systems through, for example, game-theoretic models or simulation models.

In experimental studies, researchers (such as Croson and Donohue 2003, Cantor and Macdonald 2009, Wu and Katok 2006) conduct variations of the beer game experiments to study the bullwhip effect in laboratory settings. The game can be conducted either on a physical board or with computers (see Chapter 3 for detailed discussion). Kaminsky and Simchi Levi (1998) designed a computerized version of the game, which allows playing different modes. Jacobs (2000) designed a web implementation of the game that allows an easier conduct. In the standard beer game, manufacturing capacity is infinite, prices are constant over time and setup times are zero. Therefore, the game alleviates the operational causes of the bullwhip effect that Lee et al. (1997a) mention except demand signal processing.

Next, we present the literature that studies the operational and behavioral causes of the bullwhip effect.

2.1 Studies on Operational Causes and Remedies

Lee et al. (1997b) observe causes of the bullwhip effect and how the companies cope with these causes. Then, according to coordination mechanism of echelons, they classify remedies for causes under the categories of operational efficiency, information sharing and channel alignment. Operational efficiency refers to the practices that aim at reducing the costs as well as lead times of information and materials. Examples include computer aided ordering (CAO) and echelon-based inventory control systems. Information sharing refers to activities which enable quick information flow from downstream echelons to upstream echelons of the supply chain. Under information sharing category, sharing sales (POS), inventory and capacity data through electronic data interchange (EDI) and other internet technologies are proposed. Channel alignment is the coordination of all echelons' planning, delivery, pricing processes. The most known alignment processes are everyday low pricing (EDLP), vendor managed inventory (VMI) and continuous replenishment program (CRP). Next, we present related literature based on this classification.

2.1.1 Literature on Operational Efficiency

Lead time reduction for materials or information, order batching, and computer aided ordering are some of the methods that increase the operational efficiency of a supply chain. Increased operational efficiency might provide less volatile demand through the supply chain. Cantor and Katok (2008) show that shorter lead times decrease the bullwhip effect.

Holland and Sodhi (2004) are the first to quantify the effects of the three causes (order batching, price fluctuations and rationing) of the bullwhip effect. Their results suggest that manufacturers should give incentives to retailers to minimize order batching. Following Holland and Sodhi (2004), in Potter and Disney (2006)'s simulation study, orders are placed in multiple of fixed order batch size under deterministic and stochastic demand conditions. They show that the bullwhip effect is mitigated if the batch size is a multiple of the average demand.

2.1.2 Literature on Information Sharing

Information sharing is the most recommended solution to mitigate the bullwhip effect. If sales or inventory information is not shared among supply chain echelons, upstream echelons may make production, capacity and ordering decisions based on distorted and delayed demand information. Such inefficient decisions result in excess inventories (due to high safety stock levels) or shortages at each echelon of the supply chain. Firms and researchers have been studying the role of real time demand or inventory data for efficient production planning of upstream echelons. For instance, IBM, Apple and HP started to access sell-through data of their retailers (Lee et al. 1997a). Next, we mention the literature on demand and inventory information sharing.

Demand Information Sharing

Theoretical studies of Chen et al. (2000a,b) show that accessing the POS data can reduce the bullwhip effect when customer demand information is unknown to the upstream echelons of the supply chain. When customer demand is stationary and known to suppliers, Chen (1999) states that bullwhip effect should not exist. Croson and Donohue (2003) observe that even in a stationary demand environment, firms invest in information sharing systems. For instance, Home Depot from retail industry invested in POS data sharing systems in a relatively stable customer demand environment.

By conducting experiments, Croson and Donohue (2003) investigate the impact of point of sales (POS) data sharing in reducing the bullwhip effect in a stationary demand environment. They also investigate whether the bullwhip effect still occurs when all operational causes are removed. In their research, different from other studies, they control and eliminate the demand signaling process. They announce the demand distribution to participants, which is stationary and uniform between 0 and 8. Their research indicates that the bullwhip effect still exists, even though demand information is shared through POS data. Similar to Chen et al. (2000b)'s result, however, the effect is dampened. The order oscillations at all echelons of the supply chain, specifically at the distributor and factory echelons are reduced. The amplification of the orders are also decreased significantly.

Steckel et al. (2004) investigate the impacts of changes in order and delivery cycles (lags), availability of POS information and pattern of customer demand in an experimental context. The authors show that reduction in time lags decrease supply chain costs, however the amount of reduction depends on the pattern of demand (step up, S-shaped without error, S-shaped with error). POS data sharing is found useful only with the step up demand pattern. Contrary to theoretical studies (such as Chen 1999, Chen et al. 2000b, Lee et al. 2000 and Raghunathan 2001), sharing POS data is not found to be beneficial in terms of total echelon costs.

In a theoretical study, Gaur et al. (2005) analyze the effects of time series structure of demand processes on the value of demand information sharing in a supply chain. They study a two-echelon supply chain in which the downstream echelon (i.e., the retailer) faces autoregressive moving average (ARMA) demand. Autoregressive processes are generally similar to the real life demand processes in terms of reflecting seasonality. Gaur et. al. (2005) show that safety stock requirement of the upstream echelon (i.e., the manufacturer) decreases when he could forecast the demand from the retailer's orders or access demand information through information sharing. However, the safety stock requirement of the manufacturer increases when he could only use the most recent orders of the retailer in his planning.

Inventory Information Sharing

Theoretical research on inventory management (Bourland et al. 1996, Gavirneni et al. 1999) suggests that inventory information sharing improves supply chain performance in a one supplier, multiple retailers two-echelon supply chain. Chen (1998) compares two inventory policies (echelon stock and installation stock) in a N-echelon supply chain to obtain the value of centralized demand information. The cost difference between echelon and installation stock policies refers to the value of centralized information. The authors find that when the numbers of echelons, lead times or batch sizes increase, value of information has a tendency to increase.

Cachon and Fisher (2000) study a setting which includes one supplier and multiple retailers under stochastic stationary customer demand. They show that information sharing provides two additional benefits: faster and cheaper order processing that leads to shorter lead times and smaller batch sizes. They compare the value of information sharing and the value of two benefits of information sharing. Results show that information sharing reduces supply chain costs by 3.14% whereas reducing lead times (or batch sizes) to half decreases supply chain costs by 21%. The authors propose that using information sharing technology to smooth and speed up the physical flow of materials through a supply chain is more valuable than using information technology to expand the flow of information.

In addition to theoretical studies, researchers are also conducting experiments to investigate the impacts of inventory information sharing. In their web-based experimental study, Machuca and Barajas (2004) show that implementing electronic data interchange (EDI) for information transmission along the echelons of a supply chain decreases the bullwhip effect and mean inventory costs. This finding is consistent with theoretical results.

Croson and Donohue (2005) analyze the effects of sharing the upstream and downstream inventory information across supply chain echelons, separately. They compare these treatments with their baseline treatment in which the participants cannot see other echelons' inventory information. The authors find that sharing downstream information results in a significant reduction in order oscillations. Croson and Donohue (2006) also investigate the impacts of inventory data sharing across the supply chain. Similar to Croson and Donohue (2003), they eliminated all operational causes. They show that inventory data sharing decreases the oscillation of orders at each echelon of the supply chain, specifically at the distributor and factory echelons. Inventory information sharing also decreases the amplification between the distributor and wholesaler echelons.

The results of implementing inventory information sharing in practice are in line with experimental and theoretical studies. Firms in some industries, especially in grocery industry, utilize advanced information sharing to share real time inventory information throughout their supply chains.

In a survey study, Nienhaus et al. (2003) analyze the value of information about a downstream echelon (including sales forecasts and promotions) to upstream echelons. They ask to operations managers of 200 European companies whether information on their downstream echelon (i.e., customer) is valuable for the production planning of their own upstream echelon (i.e., supplier). Results indicate that operations managers estimate that the customer information is less valuable for their suppliers. Therefore, they share their customer information with their suppliers not as frequently as their customers share this information with them.

Wu and Katok (2006) study the impact of learning and communication on the bullwhip effect. They test the effects of bounded rationality, experiential learning, systems learning and organizational learning with six different treatments. They find that training or communication separately cannot alleviate the bullwhip effect. However, communication and system-wide information sharing together can improve the supply chain performance.

2.1.3 Literature on Channel Alignment (Strategic Alliances)

In Section 2.1.2, we discussed the effects of information sharing in reducing the bullwhip effect. Real life implementations, however, show that in order to gain great improvements in supply chain performance, both information sharing and collaborative planning (such as quick response (QR), continuous replenishment program (CRP) or vendor managed inventory (VMI)) are needed (Kurt Salmon Associates 1993, Clark and Hammond 1997, Kulp et al. 2004). For example, by implementing information sharing and continuous replenishment together, Campbell soup is reported to reduce average retail inventories by 66% and cost of products by 1.2% (Cachon and Fisher 1997).

Collaborative planning enables firms to use each other's knowledge. Suppliers become closer to end consumer demand information through retailer's point of sales data; whereas retailers get insight into lead times of products and supply availability. Empirical studies mention "strategic alliance" type solutions that provide long term benefits for firms. Firms would gain benefits by improving replenishment process of goods which leads to decrease inventory levels at the retailer in the long run as observed in Campbell Soup example (Cachon and Fisher 1997). In a quick response relationship, the supplier utilizes sales information to improve production plans and to reduce lead times. In this type of alliance, orders are determined by the retailer. One step further, in a continuous replenishment program, according to sales data, the supplier organizes shipments in determined intervals to maintain specific inventory levels. Under a vendor-managed inventory (VMI) agreement, the supplier manages the inventory levels and replenishment policies of the retailer. These alliances require the supplier to employ forecasting; inventory control and retail management skills (see Table 2-1). Through information sharing and alliances, forecasting quality increases due to the use of real sales data, and average inventory levels and order fluctuations decrease because of centralized control. All of these contribute the reduction of the bullwhip effect.

Table 2-1: Channel Alignment through Strategic Alliances[5]

Criteria / Type	Ordering Decision Maker	Inventory Ownership	New Skills Employed by the Supplier
Quick Response	Retailer	Retailer	Forecasting
Continuous Replenishment	Contractually agreed levels	Either party	Forecasting and inventory control
Vendor managed inventory	Supplier (vendor)	Either party	Retail management

Next, we summarize the literature on the operational causes of the bullwhip effect in Table 2-2. The vertical axis classifies the studies according to Lee (1997b)'s framework. The horizontal axis classifies the studies based on their methodologies as being empirical, theoretical or experimental.

[5] Simchi-Levi et al. (2007)

Table 2-2: Categorizing the Literature[6]

	Empirical	**Theoretical**	**Experimental**
Operational Efficiency		Holland and Sodhi (2004), Potter and Disney (2006)	Cantor and Katok (2008), Steckel et al. (2004)
Information Sharing	Lee et al. (1997a), Kurt Salmon Associates (1993), Clark and Hammond (1997), Kulp et al. (2004), Cachon and Fisher (1997)	Chen et al. (2000a,b), Chen (1999) , Lee et al. (2000), Raghunathan (2001), Chen (1998), Bourland et al. (1996), Gavirneni et al. (1999), Cachon and Fisher (2000)	Croson and Donohue (2003), (2005), (2006), Steckel et al. (2004), Machuca and Barajas (2004)
Channel Alignment		Simchi-Levi et al. (2007)	

2.2 Studies on Behavioral Causes

Operations management (OM) is large field that includes product development, forecasting, inventory management, process design and supply chain management. Within the field, there exists a gap between the concepts defined in the theory and the rules of thumb applied in the real life. One reason for this gap is that the tools proposed by the theory may not take into consideration some important dynamics of real life. Another reason is that trust issues, misaligned incentives, or lack of information regarding the decision makers may make implementation difficult (Bendoly et al. 2006).

Behavioral research in the field of operations management is highly relevant because operating systems are not fully automated, and human behavior has significant

[6] Lee et al. (1997b)

influence on implementation of tools and techniques in practice. Human beings decide how operating systems will function. Behavioral research in operations management field has been conducted since 1920s. Recently, some researchers have started to conduct human experiments to analyze the effects of human decision making in OM areas including quality management, production control and supply chain management (Bendoly et al. 2006). Within the supply chain management area, experiments are mostly conducted on the bullwhip effect, the newsvendor problem and supply contracting.

In their experimental study, Croson and Donohue (2003, 2006) show that even all operational causes of the bullwhip effect are removed from the supply chain; the effect persists due to behavioral factors. Next, we discuss examples of the behavioral causes of the bullwhip effect mentioned in literature.

Underweighting the Supply Line

Recall that the beer game has ordering and shipping delays (see Figure 3-1 for details). These delays represent the "supply line" for a particular echelon. Sterman (1989a) observed that participants of the beer game often undervalue the orders that are still in the supply line. Therefore, they place orders more than necessary. Sterman (1989a) identified this phenomenon as "underweighting the supply line".

Supply line underweighting is a specific example of misperception of feedback (or time delay) in stock management. Misperception of feedback means that when decisions have delayed and indirect effects on each other, people find it challenging to control the dynamics. Consequently, when making decisions in a dynamic environment, people have tendency to ignore the time delays and feedback. Researchers have shown that in general this effect is not eliminated by information availability, financial incentives or learning opportunities before making decisions (Sterman 1989b, Paich and Sterman 1993, Brehmer 1992, Diehl and Sterman 1995, Kampmann and Sterman 1998, Sterman 2006).

It is important to understand whether sharing the sales and inventory (including supply line) information eliminates the underweighting the supply line effect, because most studies in the literature propose information sharing methods to reduce the bullwhip effect. In the standard beer game, end customer demand is nonstationary and unknown to the echelons except the retailer. Sterman (1989a) report the underweighting of supply line effect under this setting. Croson and Donohue (2006) show that underweighting still occurs when the customer demand is stationary and its distribution is announced to all echelons. In addition to this, Croson and Donohue (2006) also analyze sharing of dynamic inventory information. Contrary to expectations, underweighting the supply line effect is found to be robust to inventory position information of other echelons. However, this result is consistent with the robustness (regarding information availability) of the tendency to ignore time delay and feedbacks (Sterman 1989b, Diehl and Sterman 1995).

One might think that "learning" over time may mitigate the underweighting of supply line. However, Sterman (2006) mentions experimental results of Diehl and Sterman (1995), Croson et al. (2005), Wu and Katok (2006) which show that learning is slow in dynamic environments. Also note that operational remedies that reduce the lead time would mitigate the underweighting the supply line effect through shortening the supply line itself.

Coordination risk

Croson et al. (2005) report that even customer demand is constant and known to participants, supply line underweighting and the bullwhip effect still exist. They propose "coordination risk" as a new behavioral cause. Coordination risk refers to the tendency of participants to build inventory by deviating from the equilibrium to protect themselves against the intuitive risk that other echelons will not behave optimally. Croson et al. (2005) show that holding additional on hand inventory and common knowledge of optimal policy can decrease the coordination risk but cannot eliminate it completely.

Safe harbor & Panic strategies

Over periods of the experiments, participants follow some strategies to seek their goals. Nienhaus et al. (2006) report two extreme behaviors called “safe harbor” and “panic strategy” that increase the bullwhip effect. The authors develop an online beer game that computers and humans can play together. During the experiments, some human participants order more than actually needed to protect themselves from future demand increases. This strategy is known as “safe harbor”, which causes high safety stock costs at these echelons. This strategy also pushes upstream echelons to increase their orders or to incur stock out costs. One echelon that follows safe harbor strategy negatively affects the other echelons of the supply chain.

Contrary to the safe harbor strategy, in the “panic strategy”, some participants continue to decrease their stock levels until they face an increase in their customer’s demand. This strategy also affects all echelons negatively, because when the customer demand increases, a participant that follows the panic strategy needs to order more than a participant that has enough safety stock. The authors also show that when the number of human players in the experiment increases, the average and range of the total supply chain cost increase. When the all players are human in the chain, they find that information sharing through the supply chain is beneficial.

Safe harbor and panic strategies proposed by Nienhaus et al. (2006) lead Ruel et al. (2006) to study the impacts of personality characteristics related to risk taking on supply chain performance. Experimental results show that when all echelons of the supply chain consists of low-risk-taking participants, lower inventory costs and higher backlog costs are incurred compared to the supply chain in which middle and high-risk-taking participants are found. This is because low-risk-taking people react the demand changes slower than high-risk-taking people. This late response causes high backlog costs when all echelons include low-risk-taking participants.

Problem solving approach: Abstract versus concrete

Similar to Ruel et al. (2006), Cantor and Macdonald (2009) analyze the impact of personality characteristics on supply chain performance in a beer game setting. Specifically, they investigate the effects of abstract versus concrete problem solving approaches. A person who has abstract problem solving approach generally asks why-oriented questions and is concerned with strategic implications. These lead him to adapt changes in an environment easily. A person who has concrete problem solving approach, on the other hand, asks how-oriented questions, and considers more specific details and operational concerns. These lead him to follow given tasks easily. Experimental results show that abstract-thinking participants perform better than concrete-thinking participants when information sharing is not allowed in the beer game setting. However, when information sharing is allowed through the supply chain, the effects of problem solving approaches on supply chain performance become negligible.

Overreaction to backlogs

Oliva and Gonçalves (2007) analyze the participants' reactions to backlog and positive inventory situations separately. In the standard beer game, the backlog cost is twice the holding inventory cost, which leads one to expect that participants may overreact to backlogs. Contrary to Oliva and Gonçalves (2007)'s expectations, but consistent with Delhoum and Reiter (2009)'s results, Oliva and Gonçalves (2007) show that participants do not order more when in backlog.

Counterintuitive decision-making patterns

Following Sterman (1989a) and Oliva and Gonçalves (2007), Delhoum and Reiter (2009) study behavioral causes of the bullwhip effect such as bounded rationality and misperceptions of feedback. Inspired by the beer game, they develop a new simulation game (the supply net game) in which four manufacturers produce four distinct products each, where some products are jointly produced by two manufacturers. Their experiments, containing 130 participants, show that a novel behavioral cause of the bullwhip effect is “counterintuitive decision-making pattern of participants”. Even though backlog is building up, some participants do not order, and even though inventory level is high, some keep ordering high quantities.

2.3 Operations and Sales Incentive Conflict

Shapiro (1977) discusses the incentive conflicts between operations and sales managers in some areas such as planning the capacity for uncertain sales, determining the breadth of product line, introducing new products, and coordinating supply decisions with marketing decisions. Among various areas, our study is related to the incentive conflict in coordination of supply and demand decisions.

Oliva and Watson (2007) illustrate the benefits of the S&OP process in the case of an electronics company. Prior to the S&OP approach, the sales department forecasted the sales and shared this information with the operations and finance departments. These departments mistrust the sales department’s forecast due to that department’s incentive to exaggerate the demand. Hence, the operations department came up with its own stable demand forecast using only past sales data, and the finance department forecasted the demand according to its own revenue goals. The lack of coordination resulted in inventory write offs that amounted to approximately 15% of their annual revenue in 2002.

2.4 Group versus Individual Decision Making

Here we mention the literature on “group decision making”. This is relevant because our primary research question is concerned with replacing the single decision maker with a group of two decision makers. Groups of individuals such as teams, partners, families and committees make many important decisions in the society. In a survey study, Osterman (1995) determines that work teams exist in 54.5% of U.S. American firms. Consistent with Osterman (1995), Dumain (1994) estimates that two-thirds of U.S.

firms include work teams. Various companies like P&G, General Motors, Motorola, Ford, General Electric and Caterpillar attribute their cost savings and success stories to their team-based approach (Manz and Sims 1993).

Groups are expected to make better decisions than individuals (Kocher et al. 2006, Ambrus et al. 2009, Blinder and Morgan 2010). In a complex and dynamic world, it is not possible for one to know all facts and a human being has limited information processing while making decisions. However, individuals in a group can share their information with each other, leading to a broader perspective. This allows the group to propose more alternative solutions than a single decision maker.

In the literature, various experimental studies in different contexts demonstrate that there exist systematic differences between the choices of groups and individuals. In some experiments, qualities of decisions are evaluated according to a normative criterion. Tasks in these experiments are named as intellective tasks (Laughlin 1980). Conversely, non-intellective tasks refer to tasks in which only the personal preferences should dictate choice. Increase in quality of decisions made by groups is expected in intellective tasks. At first, the differences between decisions of groups and individuals observed in non-intellective tasks are surprising. However, various experimental studies determine that people act more selfishly in a group than when making a decision individually, and groups have tendency to take risky decisions (Ambrus et al. 2009). Kocher et al. (2006) report that in their beauty contest game experiments, 60% of the participants preferred to make decision in a team.

Experiments including intellective tasks demonstrate that “two heads are generally better than one head” in different contexts. Kocher and Sutter (2005) show that groups learn faster, have ability to better anticipate and make better judgments in beauty contest games. Cooper and Kagel (2005) determine that groups play more strategically than individuals in signaling games. By conducting two experiments in different settings, Blinder and Morgan (2010) show that groups are not slower than individuals in reaching decisions, and that without additional information, groups make better decisions than individuals.

2.5 Measurement of the Bullwhip Effect

Here we outline the ways researchers measure the three characteristics of the bullwhip effect:

1) Oscillation: Generally, to measure the oscillation of orders within each echelon, one may calculate the variance of orders placed over the periods of the experiment.

2) Amplification: To measure the amplification of orders, one calculates the amplification ratio by dividing an upstream echelon's variance by downstream echelon's variance (see, for example Croson and Donohue 2006). As such, three amplification ratios are calculated for a four echelon supply chain as follows:

$$Amplification\ ratios: \quad \frac{\sigma^2_{wholesaler}}{\sigma^2_{retailer}} \quad \frac{\sigma^2_{distributor}}{\sigma^2_{wholesaler}} \quad \frac{\sigma^2_{factory}}{\sigma^2_{distributo\ r}}$$

An amplification ratio greater than 1 indicates that orders are amplified by the echelon. These are not the only measures of the bullwhip effect. Fransoo and Wouters (2000), for example quantify the amplification effect as the ratio of coefficient of variation (CV) out and in. "Out" refers to orders placed to upstream echelon and "in" refers to orders received from downstream echelon.

3) Time lag: The third component of the bullwhip effect, time lag, is somewhat more difficult to characterize. Sterman (1989a) compares the periods of the peak order level at each echelon.

While the bullwhip effect itself can be measured in terms of "orders placed", its consequences show themselves as inventory/ backlog levels at each echelon. Alternatively, one can measure the costs of inventory/ backlog at each echelon and use this as a measure of the detrimental effect of the bullwhip effect (see, for example Machuca and Barajas 2004). After all, one of the major reasons to control the bullwhip effect is to control the underage/ overage costs that it causes.

Table 2-3 illustrates the different measures that researchers use to quantify the bullwhip effect.

Table 2-3: Types of Measures

Types Of Measures	**Researchers**
Mean of Orders	Machuca and Barajas (2004)
Standard Deviations of Orders	Cantor and Katok (2008), Machuca and Barajas (2004), Wu and Katok (2006)
Variance of Orders (VO)	Cantor and Macdonald (2009), Croson and Donohue (2003), (2005), (2006)
Amplification Ratio = VO at Upstream / VO at Downstream	Croson and Donohue (2003), (2005), (2006)
Ratio = Factory Order Variance / Customer Demand Variance	Manyem and Santos (1999)
Coefficient of Variation (CV) of Demand	Disney et al. (2004)
CV out / CV in	Fransoo and Wouters (2000)
Standard Deviations of Costs	Machuca and Barajas (2004)

Chapter 3

The First Study: Beer Game with Two Participants at Each Echelon

In the first study, we propose a modified beer game that involves two participants at each echelon of the supply chain. One of the participants is in the role of the supply manager and the other is in the role of the sales manager. These managers generally have incentive conflict in real life. In the modified experiments, these two participants together decide a single order quantity for their echelon at each period of the experiment. We aim to understand whether this modification will decrease the bullwhip effect or not. To this end, we conducted beer game experiments with standard and modified experiment types and made statistical comparisons on the outcomes.

3.1 Experimental Design and Implementation

Our "standard game experiments" follow previous studies with respect to basic protocols of the beer distribution game (Sterman, 1989a) with some minor modifications on initial inventory level and number of periods of the experiment.

The mechanism of the standard game experiments is as follows:

- The game models a four echelon supply chain, as illustrated in Figure 3-1. The echelons are the retailer (R), wholesaler (W), distributor (D) and factory (F).
- The product that moves in this supply chain is beer, which is measured in "cases". The cases are represented by plastic coins in the board game.

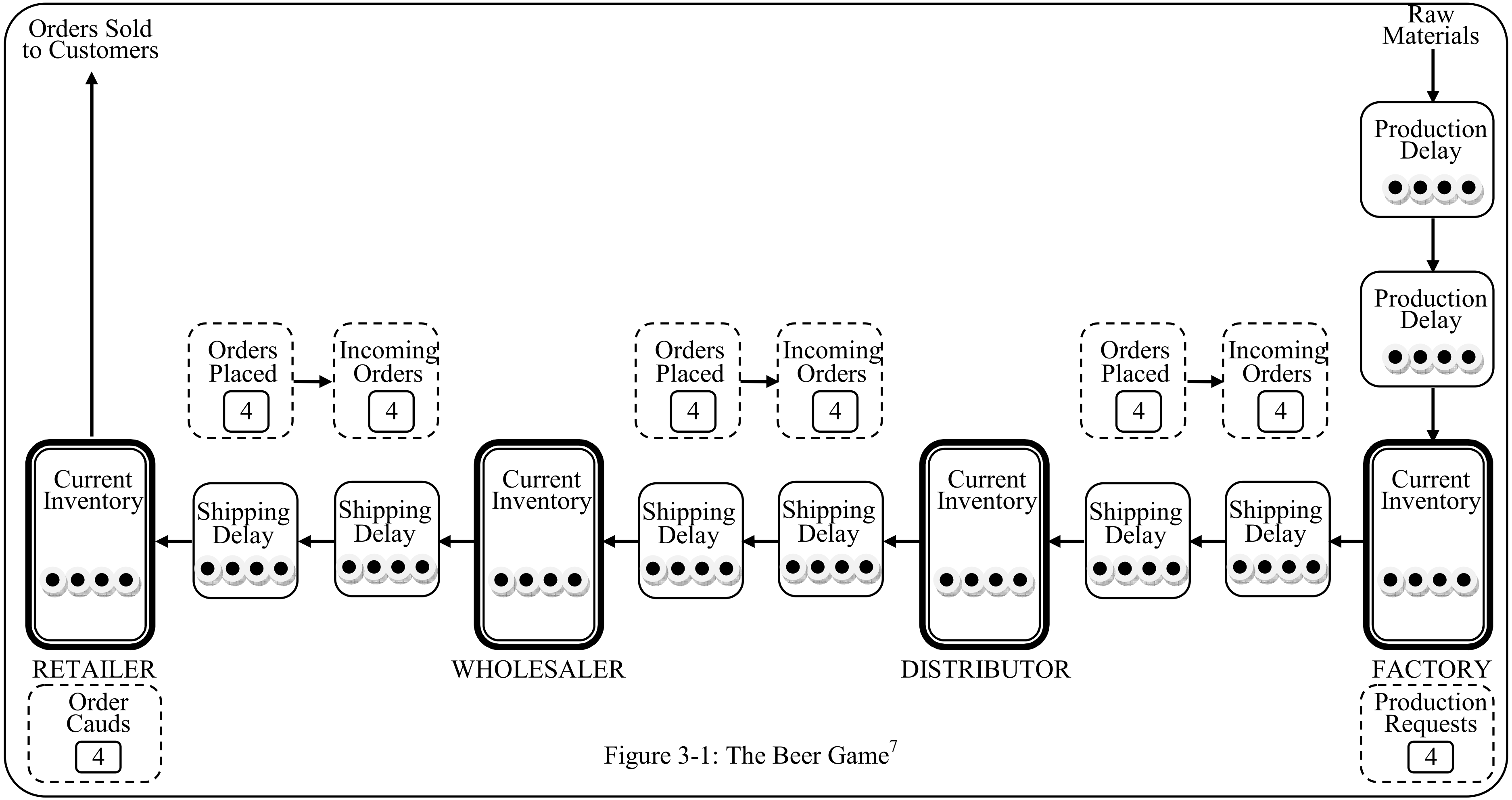


Figure 3-1: The Beer Game[7]

[7] http://web.mit.edu/jsterman/www/SDG/beergame.htm

- The experiment continues for 24 “periods”.
- At each period, each follows a sequential procedure which can be summarized as follows: The echelon receives his incoming orders from his upstream echelon, observes demand from his downstream echelon, tries to fulfill this demand as much as possible from on-hand inventory, records his inventory/ backlog level, and places a new order (which can be zero cases) to his upstream echelon.
- Customer demand at the retailer echelon is exogenously given. It is equal to 4 cases/ period during the first 4 periods, and 8 cases/ period during periods 5-24. This demand stream is unknown to participants and it is revealed to the retailer period by period.
- Demand at each other echelon consists of the orders of the respective downstream echelon. For example, the orders of the retailer become the demand of the wholesaler.
- When an echelon places an order to his upstream echelon, the upstream echelon receives the order two periods later. This “ordering delay” reflects the order processing lead time. To keep track of the cases in ordering delay, the board game has two “ordering delay” boxes between consecutive echelons. These boxes are initialized with 4 cases each to reflect orders in process at the beginning of the experiment.
- When an upstream echelon fulfills the orders received from a downstream echelon, the downstream echelon receives cases two periods later. This “shipping delay” reflects the shipping lead time. To keep track of the cases in shipping delay, the board game has two “shipping delay” boxes between consecutive echelons. These boxes are initialized with 4 cases each to reflect incoming orders in transportation at the beginning of the experiment.
- The factory echelon, which does not have an upstream echelon, places a “production order” to himself. A production order takes three periods to materialize. This “production delay” reflects the production lead time. To keep track of the cases in production delay, the board game has three “production delay” boxes next to the factory echelon. These boxes are initialized with 4 cases each to reflect production in progress at the beginning of the experiment.
- If an echelon cannot meet the demand he faces in a given period, this demand is backlogged. Backlogged demand is met when inventory becomes available.

- Each echelon places his order by writing it in an order card and placing this order card into his “orders placed” box.
- At the end of each period, each echelon records his order quantity into a record sheet (see Appendix A). Inventory on hand incurs a holding cost of $1/ case/ period whereas backlog incurs a backlog cost of $2/ case/ period.
- At the beginning of each period, the cases in ordering delay, shipping delay and production delay are moved in the relevant directions by the participants. This represents the flow of information and materials in the supply chain.
- At the end of the experiment, for each echelon, the sum of the inventory holding and backlog costs over all periods is calculated. The team-objective of each four-participant team is to minimize the total supply chain cost, corresponding to the sum of the four echelons’ costs.

At each period of the experiment, every echelon has to follow the following sequential procedure. It is critical that all participants follow these steps simultaneously to avoid confusion in the experiment. This process received special attention of our experiment facilitators.

- Receive cases from shipping delay.
- Fulfill the orders of the downstream echelon as much as possible.
- Record the backlog or inventory in the record sheet (see Appendix A).
- Retailer, Wholesaler, Distributor echelons: Move the order cards.
  Factory echelon: Move the production card.
- Place a new order to upstream echelon and record in the sheet.

The beer game can be conducted in a laboratory or classroom environment either with computers or as a board game. We run the board version. Figure 3-2 presents a photo taken during one of our experiments. The board game provides a more realistic environment for participants to feel the atmosphere and understand the dynamics of the supply chain. On the other hand, the board game has the disadvantage of being open to human errors in moving cases and in recording data in sheets.

Figure 3-2: One of Our Experiments

We conducted two types of experiments as summarized in Table 3-1. The standard experiments followed the procedure we explained. Each standard experiment is played by a four-participant team. At each echelon the participant (manager) who is responsible for both inventory holding and backlog costs determines the order quantity at each period.

Table 3-1: Design of Experiments

Experiment Type	**Number of Participants at Each Echelon**	**The Role(s) of Participants**	**Incentives of Participants**
Standard	1	Manager	Minimize the sum of inventory and backlog costs
Modified	2	Supply Manager	Minimize inventory costs
		Sales Manager	Minimize backlog costs

The modified experiments are different only in one aspect: In each of the four echelons, there are two participants instead of one (adding up to eight participants in an experiment). They determine the order quantity together at each period. One of these participants plays the role of supply manager, whose performance measure is the inventory holding cost. The other participant plays the role of the sales manager whose performance measure is the backlog cost. Naturally, the supply manager prefers smaller order sizes whereas the sales manager prefers larger ones. We are interested in determining the effect of this incentive conflict (at each echelon) on the bullwhip effect. We expect that the discussions between the two managers will make it less likely to place large orders (because the supply manager will object to this) leading to a decrease in the bullwhip effect.

The participants in the experiments were Sabanci University students. Four groups of senior students between 2008 and 2010 helped us as "experiment facilitators", as part of their graduation project. Detailed participant information can be found in Appendix B. We paid attention to make sure that no participant has prior experience with the beer game. Data acquisition process details are presented in Appendix C.

At the beginning of the experiment, participants are randomly assigned to echelons and roles. We go over the mechanics of the game and explain each participant's role in detail. In particular, we explain that the inventory/ backlog level should be recorded as cumulative (that is, it is carried over from one period to the next). For the modified experiments, we explain the two managers' incentives in detail. The participants know that the overall goal of the team is to minimize the total supply chain cost.

After we make sure that all participants understand the goals and the mechanics of the game, we conduct a pilot experiment that takes 3-4 periods. During the pilot periods, our facilitators answer questions from participants and check whether they are playing correctly. The pilot period results are not recorded. After the pilot experiment, we start the real experiment. We announce there will be no communication between echelons during the experiment.

During the experiment, our facilitators observe the participants and intervene if they see something wrong. In particular, they make sure that all participants follow the sequential procedure we described. We announce that the experiment will take 32 periods, however, we end the experiment at 25th period to avoid "end of experiment" behavior. At the end of the experiment, we calculate total supply chain costs for each team. We announce the winner team and the winner sales and supply managers separately. We also made the participants fill in a post-experiment survey. This survey is provided in Appendix D.

We compare the modified and standard experiments in terms of the orders, total cost, inventory cost and backlog cost. At the end of each period, an echelon incurs either an inventory holding cost or backlog cost. We sum these costs over periods to determine the inventory cost and the backlog cost of the echelon. The total cost refers to the sum of these two costs. We calculate and report both the mean values and the variances of these measures.

3.2 Experimental Results and Analysis

After conducting the beer game experiments, we entered the experimental data from record sheets into MS Excel. Next, we checked the data against invalid entries. We eliminated some team's data due to inconsistencies at this stage. Then, we further eliminated data using outlier analysis. Finally, we compared the standard and modified experiments through descriptive analysis and hypothesis testing, and applied formal statistics test to observe significance of difference.

Before explaining the details of our experimental data analysis, we first present our outlier elimination process and the hypothesis tests we use.

3.2.1 Outlier Analysis

Before conducting statistical analysis on data, we determined and eliminated the outliers. Grubbs (1969) defines an outlier as: "An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs".

Eliminating outliers is crucial for our study, because we measure the bullwhip effect through the variance of orders, which is very sensitive to large data values. Therefore, we considered teams that have very high variance or amplification ratio as an outlier.

Various methods are found to detect outliers. In the bullwhip effect literature, Wu and Katok (2006) conduct Grubbs' outlier detection method for each echelon and each experiment type separately. Machuca and Barajas (2004) detect outliers by observing box plots of variables. Massart et al. (2005) states that box plots are more robust to the presence of outliers than classical methods based on normal distribution, such as Grubb's method. Similar to Machuca and Barajas (2004), we used box plots according to the variance of orders and amplification ratio variables for each echelon and experiment type separately.

A box plot allows one to observe important features of data like spread, center and outliers. It represents batches of data through five values (McGill et al. 1978): As seen in Figure 3-3, the bottom of the box shows the lower quartile (25th percentile), the top of the box shows the upper quartile (75th percentile) and the line near the middle of the box shows the median (50th percentile) of the data. Interquartile range (IQR) is the range between the lower and upper quartiles. The ends of the whiskers (vertical lines) represent the lowest and highest values that are within 1.5 times the IQR (box width). Values that are between 1.5 and 3 times the IQR are named as outliers and values that are more than 3 times the IQR are named as "extremes".

Figure 3-4 presents the box plot for the order variance data for teams in our standard experiments. The stars denote extremes and the circles denote possible outliers. The numbers denote the team numbers. We created such box plots for the order variance and amplification values. We marked the teams that cause extreme values in any one of their four echelons. We eliminated a team if it causes two or more extreme values in total (according to the variance of orders or the amplification ratios, combined). Other box plots are presented in Appendix G.

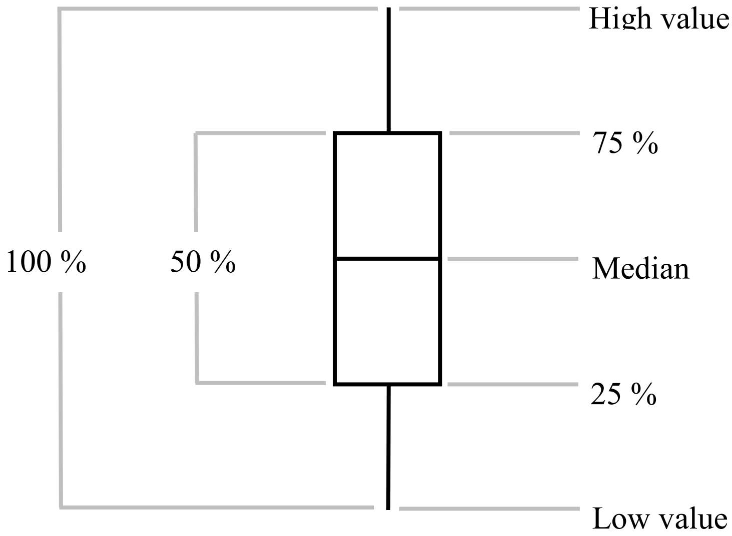


Figure 3-3: Sample Box Plot[8]

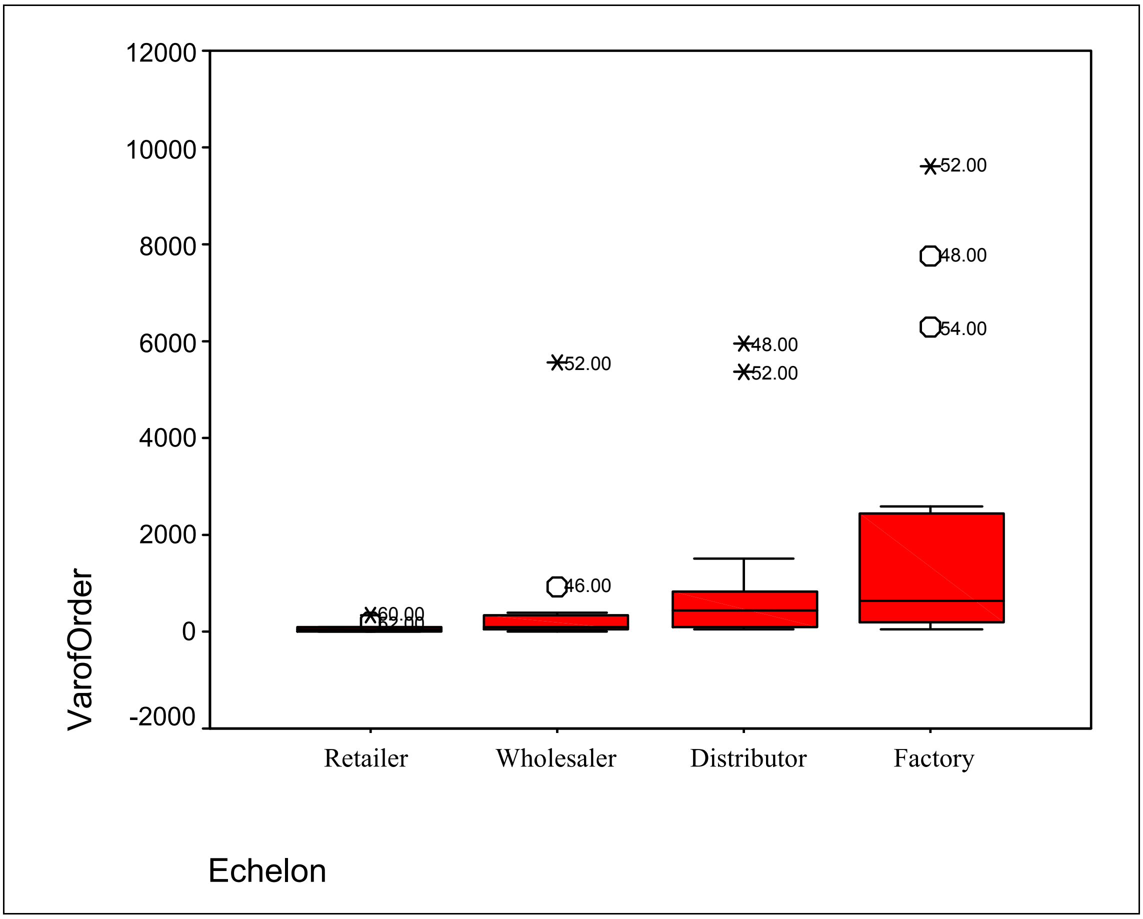


Figure 3-4: Box Plot of Variance of Orders for Standard Experiments

---

8 http://www.information-management.com/issues/20050801/1033566-1.html

The data for the analysis was carried out through a long and strenuous effort that spanned over a period of 24 months. Four student groups conducted 39 experiments for the modified game, and 23 experiments (with a total of 62 teams) for the standard game as a part of their senior projects. The groups handed in the collected data, together with the hardcopies of the record sheets to the supervisors, before the thesis started. For the 2008-2009 academic year, the data and the sheets were handed in by the two groups at the end of the semester, after all the experiments were completed. A very detailed data validation procedure was carried out by the supervisors. Unfortunately, the data for all 12 experiments (teams) of one of the groups was found to be unusably dirty unreliable. The other group of the 2008-2009 academic year had conducted 16 experiments, but only 10 of these teams were found to be recorded correctly. Thus, out of all the experiments carried out in 2008-2009 academic year, only the data of 10 experiments was judged to be reliable and valid. The main reason of unreliability in the dirty data were the unavailability of hard copy record sheets to cross-check with the data in the Excel spreadsheets for validation. Another reason was the re-entry of the data of one group by the other in their Excel sheets. Other sources of errors include wrong levels of initial inventory, inconsistency between columns, and data that was "too regular", giving the impression of being generated, rather than being collected.

The failure in data collection in the first year of the project guided the data collection and validation procedure in the second year. In the 2009-2010 academic year, data was validated by the supervisors and the author of this thesis as it was collected. This new procedure resulted in much more reliable data, initially resulting in the collection of data of 34 experiments. Since the validation of the data cannot be done during the experiments in the board game version of the beer game, some game data was found to be unreliable in this academic year, as well. Out of the 34 valid experiments, 29 of them were found to be reliable and included in the analysis. Eventually, a total of 39 teams were considered to be included in the data analysis for this thesis, and a final validation check was carried out. Hence, our outlier analysis started with 23 modified and 16 standard teams. After outlier elimination (Appendix G), we are left with a total of 33 (19 modified and 14 standard) teams for further analysis.

3.2.2 Preliminary Observations

Here, we present our preliminary observations regarding the existence of the bullwhip effect. To observe whether bullwhip effect exists for an experiment, we first plot the order data for each team and echelon. These are presented in Appendices E and F. A typical example is provided in Figure 3-5, which shows the orders placed by each echelon of one of our teams over the periods of the experiment.

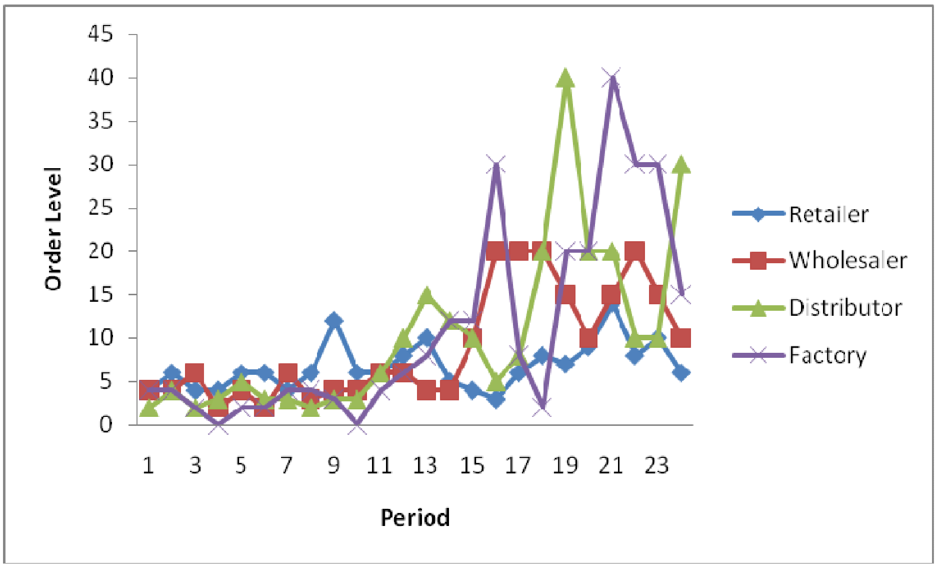


Figure 3-5: Orders Placed over Periods of the Experiment (for team 39)

The figure exhibits the three characteristics of the bullwhip effect as mentioned in Chapter 1. We observe "oscillation of orders": Each echelon's orders have a zigzagging pattern. We observe "amplification of oscillations": The variance of oscillations increase as one goes from downstream to upstream echelons. We observe the "time lag" between order increases: The order pattern shifts with a time lag as one moves towards the upper echelons. For instance, we observe the peak orders for retailer, wholesaler, distributor and factory at periods 9, 13, 19, 21 respectively.

In order to understand the nature of oscillations of orders, we graph the "effective inventory level" of each echelon of the same team in Figure 3-6. Recall that at each period of the experiment, an echelon either is in backlog (negative effective inventory) or has on hand inventory (positive effective inventory). For the team in the figure, the retailer and the wholesaler experienced their first backlog at period 8, whereas the distributor and factory fell in backlog at periods 11 and 13 respectively.

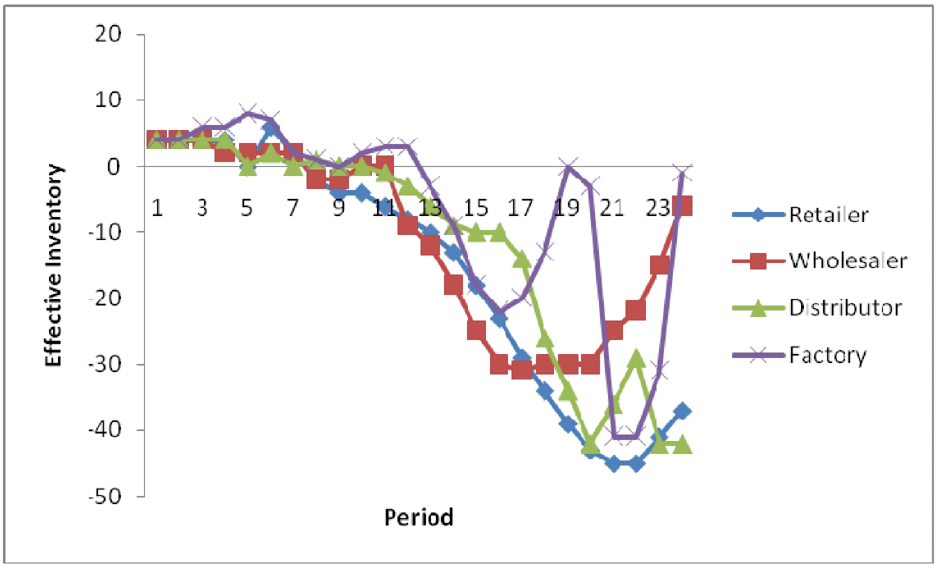


Figure 3-6: Effective Inventory Levels over Periods (for team 39)

Next, in Figure 3-7, we present the order and effective inventory plots together for each echelon. We observe that when effective inventory levels decrease very much, the order quantities increase at the same and at the following periods. In other words, for each echelon, peak effective inventory levels and peak order quantities occur around the same time period. The participants react to their backlog and increase their order levels to compensate their backlogs.

It is interesting to observe that the retailer experiences a huge backlog. Being exogenous, the retailer's demand is the most stable of all; hence, one does not expect the retailer to experience high levels of backlog. The explanation lies in "supply uncertainty". The retailer increases his orders over time, but that does not guarantee that the orders will be delivered by its upstream echelon, the wholesaler. In fact, because the wholesaler himself is in deep backlog, the retailer's supply is highly uncertain. The same is true for the wholesaler, who is supplied by the distributor. Among all echelons, the only one that does not experience this supply risk is the factory. The factory is sure that once he places an order, the order will be produced by himself in three periods. This "supply uncertainty" factor turned out to be very strong in our experiments, and we will mention its effects later in the thesis.

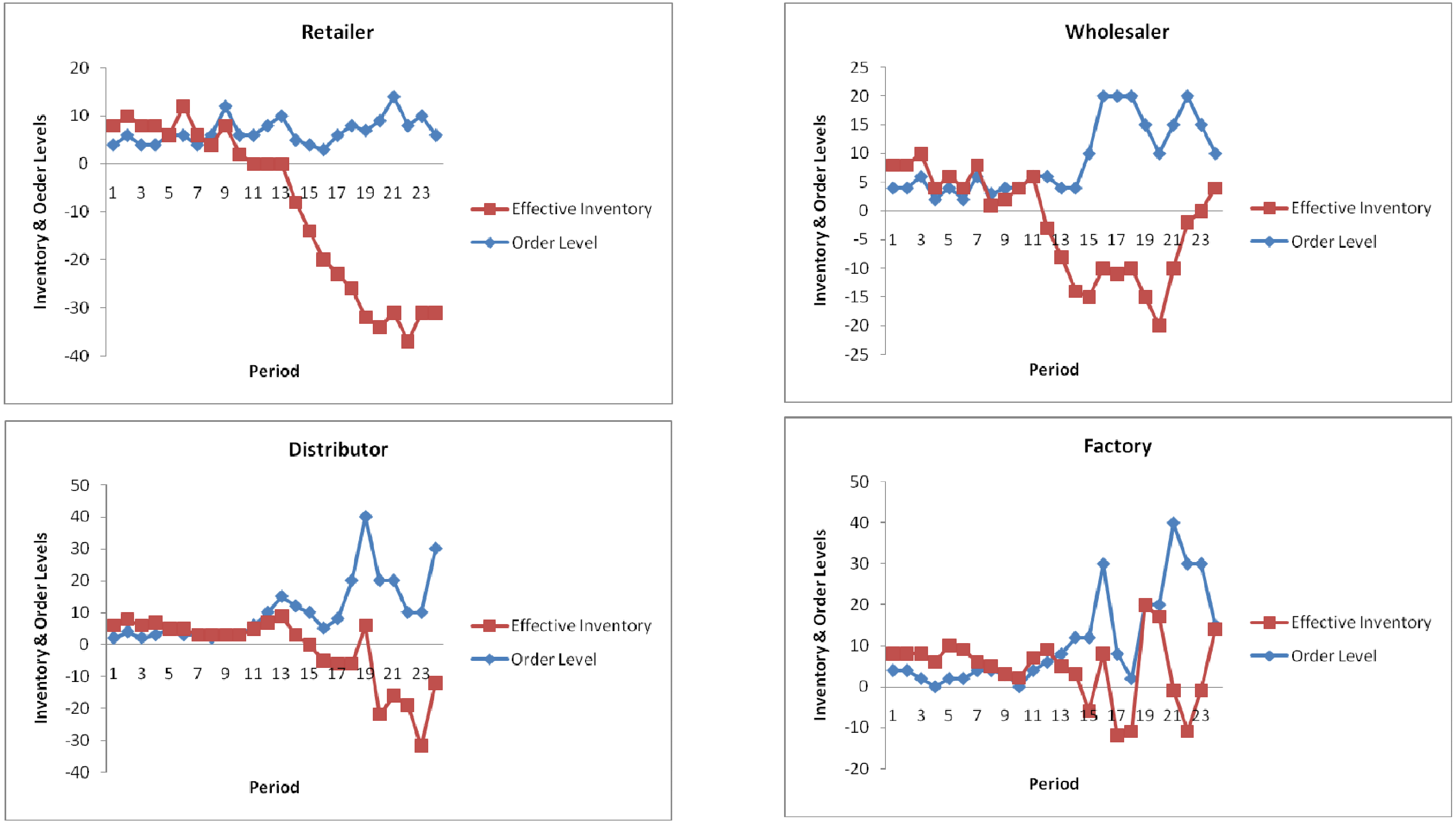


Figure 3-7: Order and Effective Inventory Levels of the Echelons (for team 39)

We discussed these observations using a single team data. Most of the other teams' data also exhibit the bullwhip effect, as can be seen in Appendices E and F. We should note that there are also some teams for which the bullwhip effect was not pronounced.

Before comparing the standard and modified experiments, we also provide a support on the existence of amplification in our experiments. Similar to Croson and Donohue (2006), we perform a sign test to measure differences in order variances between adjacent echelons. The sign test is applicable to compare two related samples when one wants to show that two populations are different. The test assumes that the variable has a continuous distribution and does not make any assumptions on the type of the distribution. The test focuses on the direction of the differences. Under the null hypothesis, one expects half of the differences to be negative and half to be positive. One can reject the null hypothesis if too few differences of one sign occur (Siegel 1956). We state our hypothesis as follows:

Hypothesis 1: Amplification occurs in both the standard and modified experiments.

We assign a positive sign for an increase in order variances between adjacent echelons (retailer/ wholesaler, wholesaler/ distributor, distributor/ factory pairs), and a negative sign for a decrease. Data reveals that the rate of the positive signs is 90% in the standard ($N= 42, x= 4, p< 0.001$)[9] and 89% in the modified experiments ($N= 57, x= 6, p< 0.0001$) which supports our hypothesis. Thus, the increase in the variance of orders between adjacent echelons is significant. There is statistical evidence for the existence amplification in both standard and modified experiments.

3.2.3 Comparison of the Standard and the Modified Experiments

Here, we compare the standard and modified experiments visually, as well as with descriptive analysis and hypothesis tests.

---

[9] We obtain one-sided p values from sign test table D (Siegel 1956)
N refers to number of pairs, x refers to number of fewer signs.

First, we make a visual comparison. Figure 3-8 and Figure 3-9 illustrate the order variances at each echelon for standard and modified experiments respectively. Each color represents one team. In both figures, we observe that the order variance increases from downstream to upstream echelons. We also note that the variance values within an echelon exhibit strong difference from team to team. Comparing the two figures, we observe that the average order variance in modified experiments is less than the average order variance in standard experiments.

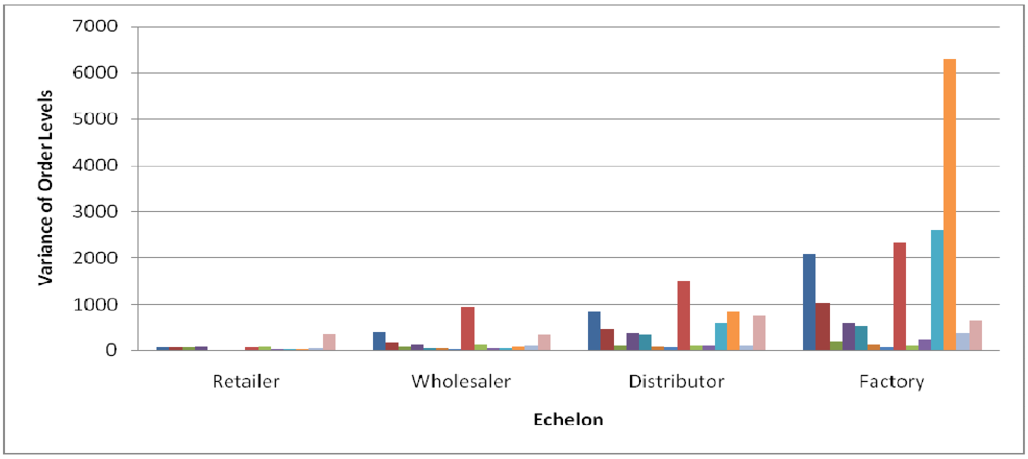


Figure 3-8: Order Variances in the Standard Experiments

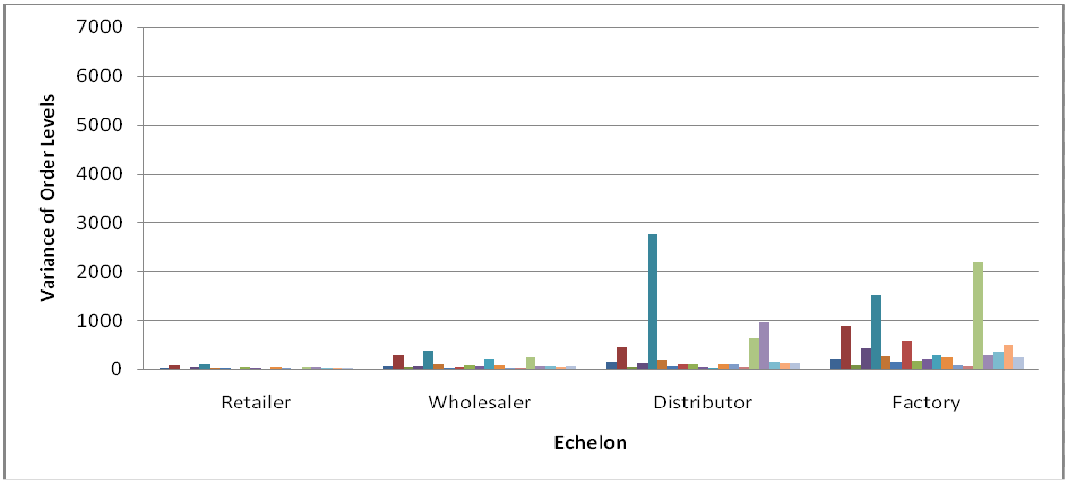


Figure 3-9: Order Variances in the Modified Experiments

Next, we analyze the average values of variables *over different teams* for each experiment type and echelon. For instance, by "average of mean orders over teams" we refer to the average of "mean of orders placed over periods" over different teams.

Following this descriptive analysis, we present our hypothesis-testing results regarding comparisons between the standard and modified games. To this end, we use the "*Mann Whitney U test*" (Siegel 1956).

Mann Whitney is a "nonparametric" test. Nonparametric tests do not assume any particular distribution regarding the population, whereas parametric tests assume that we are testing the random samples based on normally distributed. The cost of this generality is the reduced power of nonparametric tests due to not benefiting from all the information provided by the sample. However, the power loss is not large for small sample sizes. Consequently, nonparametric tests are preferred when the sample size is small and the underlying distribution is not normal.

In the Mann Whitney U test, the null hypothesis suggests that the two populations have the same distribution. To this end, the test combines observations from two samples and ranks these in an increasing order. The test provides a test statistics, *U*, based on the rank-order of the observations. According to sample sizes (*n* and *m*), the test calculates a statistic "*z*" and the related significance level "*p*". If the *p*[10] value is smaller than the selected significance level (α= 0.05), one can reject the null hypothesis.

3.2.3.1 Oscillation Comparison

Table 3-2 presents the average order variances at each echelon. We observe that variance at each echelon in the modified game is lower than its counterpart in the standard game. As seen in the Table 3-2, the largest reduction in the average of order variances is observed at factory echelon with a ratio of 62%. However, the retailer, wholesaler and distributor echelons experienced 44%, 40% and 24% decreases respectively. Observing larger reductions at downstream (retailer, wholesaler) echelons than the distributor is interesting in terms of the oscillation aspect of the bullwhip effect. Appendices H and I provide detailed tables of order comparisons.

[10] ** refers to strongly significant difference in the test ($p < 0.05$).
* refers to weakly significant difference in the test ($0.05 < p < 0.10$).

Table 3-2: Order Variance Comparison

Measurement Unit	Experiment Type	R	W	D	F
Avg. of Order Variances over Teams	Standard	63.26	180.53	444.32	1225.13
	Modified	35.12	107.87	336.03	469.58

Next, we provide the related hypothesis test.

Hypothesis 2: The modified experiments will decrease the order variance **(i.e. oscillations)** compared to the standard experiments.

Table 3-3 presents the *p* values for the order variance comparisons. The reduction in order variance is strongly significant for the whole supply chain (SC) (n= 56, m= 76, U= 1762, p= 0.046), and for the upstream (D, F) echelons (n= 28, m= 38, U= 396, p= 0.039). In contrast, the downstream echelons (R, W) could not enjoy a significant decrease (n= 28, m= 38, U= 442, p= 0.124). This is consistent with Croson and Donohue (2006) who study the effect of inventory information sharing. The only echelon-by-echelon reduction that has significance is the one in factory echelon, and this is weakly significant.

Table 3-3: *P* Values of Hypothesis Tests for Order Variances

Echelon	SC	R, W	D, F	R	W	D	F
Order Variances	0.046^{**}	0.124	0.039^{**}	0.179	0.23	0.152	0.099^{*}

3.2.3.2 Amplification Comparison

Next, we compare the amplification ratios across the experiment types. Recall that, this ratio is calculated by dividing an upstream echelon's order variance by downstream echelon's order variance. From Table 3-4, we observe that the "*average of amplification ratios*" is decreased by 10% and 25% in the wholesaler/ retailer, and distributor/

wholesaler pairs. In the factory/ distributor pair, we observe an increase (11%). See Appendices H and I for details.

Table 3-4: Amplification Ratio Comparison

Measurement Unit	Experiment Type	W / R	D / W	F / D
Avg. of Amplification Ratios over Teams	Standard	4.08	3.76	2.43
	Modified	3.65	2.84	2.69

Next, we provide the related hypothesis test.

Hypothesis 3: The modified experiments will decrease the **amplification ratios** between adjacent echelons of the supply chain compared to standard experiments.

The Mann-Whitney test *p* values presented in Table 3-5 indicate that we could not find support for Hypothesis 3.

Table 3-5: *P* Values of Hypothesis Tests for Amplification Ratios

Echelon Pairs	SC	W/ R	D/ W	F/ D
Amplification Ratio	0.423	0.493	0.179	0.327

3.2.3.3 Time Lag Comparison

Compared to oscillation and amplification, time lag is different to define and quantify. We analyze the time lag using the following measures:

- The periods (and magnitudes) of peak order levels
- The periods (and magnitudes) of peak backlog levels
- The first period to experience backlog

The first two are similar to Sterman (1989a). Here, we present the average values of these measures over all teams. Individual team values are presented in Appendix H. Also recall that the team graphs are found in Appendices E and F.

Table 3-6 shows that the average period of peak order level (and the average magnitude of peak orders) increases as one moves upstream. This is true for both the standard and modified experiments. We observe a slight decrease in the average period of peak order levels from standard to modified experiments.

Table 3-6: Peak Orders Comparison

Measurement Unit	Experiment Type	R	W	D	F
Avg. Period of Peak Orders	Standard	13.57	14.57	16.00	17.29
	Modified	12.26	14.47	15.79	16.47
Avg. Magnitude of Peak Orders	Standard	24.64	40.36	71.07	110.00
	Modified	19.89	34.42	55.79	72.63

The situation is different for the backlogs. The average period of peak backlog values are closer to each other compared to the average period of the peak order values. For standard experiments, there is no particular trend between the four average values, whereas for the modified experiments the average periods decrease as one moves upstream. Combining these observations with Table 3-7 suggests that the downstream echelons increase their orders before the upstream echelons, however, due to supply risk, they cannot recover from the backlog earlier than the upstream echelons. The average peak backlog magnitudes do not indicate any particular ordering between the echelons.

Table 3-7: Peak Backlogs Comparison

Measurement Unit	Experiment Type	R	W	D	F
Avg. Period of Peak Backlogs	Standard	17.93	17.36	18.79	18.36
	Modified	18.89	18.11	17.53	16.95
Avg.Magnitude of Peak Backlogs	Standard	-39.50	-107.29	-94.86	-91.86
	Modified	-47.11	-76.37	-78.00	-52.84

Finally, we observe that the average period of first backlog occurrence increases as one moves upstream. This is consistent with our expectations.

Table 3-8: First Backlogs Comparison

Measurement Unit	Experiment Type	R	W	D	F
Avg. Period of First Backlogs	Standard	6.3	6.4	9.0	9.9
	Modified	5.9	7.1	8.4	9.5

3.2.3.4 Mean Order Comparison

Next, we compare the mean order over period values. Note that the mean order over periods for an echelon is not directly related to the three aspects of the bullwhip effect. As demonstrated in Table 3-9, the “averages of mean orders” at every echelon of the supply chain are less in the modified experiments relative to the standard experiments. However, the reduction is not symmetric. Modified experiment reduced the average of mean orders with a ratio of 29% for upstream echelons, while this ratio becomes 21% at downstream echelons. This result is not surprising, since in the standard experiments, upstream echelons face with high orders which are already amplified by the retailer and the wholesaler. However, in the modified experiments, because the orders are amplified less by the downstream echelons, the upstream echelons do not need to amplify as well.

Table 3-9: Mean Order Comparison

Measurement Unit	Experiment Type	R	W	D	F
Avg. of Mean Orders over Teams	Standard	11.15	14.27	18.13	21.27
	Modified	8.88	11.20	12.86	15.13

Next, we provide the related hypothesis test.

Hypothesis 4: The modified experiments will decrease the mean orders compared to the standard experiments.

Table 3-10 presents the *p* values for the mean order comparisons. The reduction in mean orders is strongly significant for the whole supply chain (SC) (n= 56, m= 76, U= 1521, p= 0.002), for the downstream (R, W) echelons (n= 28, m= 38, U= 365, p= 0.015) and upstream (D, F) echelons (n= 28, m= 38, U= 344.5, p= 0.007). The echelon-by-echelon comparison finds that the reduction in mean orders was strongly significant for the distributor echelon, whereas it was weakly significant for the other three echelons.

Table 3-10: *P* Values of Hypothesis Tests for Mean Orders[11]

Echelon	SC	R, W	D, F	R	W	D	F
Mean Orders	0.002^{**}	0.015^{**}	0.007^{**}	0.07^{*}	0.052^{*}	0.026^{**}	0.065^{*}

In particular, the reduction in mean orders becomes more significant as one goes from the retailer echelon to the distributor echelon. However, the order mean reduction in the factory level is less significant (i.e., has higher p value) than the distributor echelon. This result is somewhat expected. As Croson and Donohue (2006) mention, even though the factory is the uppermost echelon in the supply chain, it is not necessarily the one that is most affected by the bullwhip effect. This is because the factory does not face supply uncertainty whereas the other echelons do. The factory is sure to receive products after a three period delay, once he places a production order. The other echelons depend on the inventory status of their upstream echelon. We observe that for supply chains that experience high backlog, supply uncertainty might become a critical determinant of the bullwhip effect.

3.2.3.5 Cost Comparison

Next, we compare the costs between the standard and modified experiments in Table 3-11. Recall that the total cost consists of backlog and inventory costs. The cost comparison is important because the costs quantify how much the firms suffer from the bullwhip effect. We expect the reductions in order averages to lead to a decrease in total

[11] *P* values are one sided.

costs of each echelon. Before moving on to the detailed results, we briefly list our main observations about the cost changes across the experiment types.

Table 3-11: Cost Comparison

Measurement Unit	Experiment Type	Retailer			Wholesaler		
		Inv. Cost	Backlog Cost	Total Cost	Inv. Cost	Backlog Cost	Total Cost
Avg. of Mean Cost over Teams	Standard	5.59	31.29	36.88	3.82	68.68	72.5
	Modified	1.80	40.61	42.41	4.49	49.85	54.34
Avg. of Cost Variance over Teams	Standard	296.46	978.32	1086.85	175.36	10648.65	10475.32
	Modified	33.50	1208.10	1108.81	155.45	3719.11	3499.54

Measurement Unit	Experiment Type	Distributor			Factory		
		Inv. Cost	Backlog Cost	Total Cost	Inv. Cost	Backlog Cost	Total Cost
Avg. of Mean Cost over Teams	Standard	6.74	51.48	58.22	9.73	36.11	45.84
	Modified	7.22	45.97	53.19	13.18	18.65	31.83
Avg. of Cost Variance over Teams	Standard	494.97	5373.63	5255.32	594.59	4898.53	4873.64
	Modified	297.81	4329.11	4083.11	576.17	1313.59	1487.24

We observe that the backlog cost dominates the inventory cost in both the standard and modified experiments, at every echelon. This happens because the per unit cost of backlog is twice the per unit cost of holding inventory, and also because the echelons in most of our experiments stay in backlog for long periods. In the modified experiments, relative to the standard experiments:

- The total cost decreased at all echelons except the retailer. However, the increase at the retailer is quite small.
- The backlog cost decreased by 27%, 11% and 48% at the wholesaler, distributor and factory echelons respectively. It increased in the retailer echelon, but slightly.
- The inventory cost increased by 18%, 0.07%, 36% at the wholesaler, distributor and factory echelons respectively. It decreased at the retailer echelon.

In the modified experiments, at the wholesaler, distributor and factory echelons, the "averages of mean total costs" are reduced relative to standard experiments by 25%, 9% and 31% respectively. In contrast, the retailer echelon has not experienced a reduction in total costs. This is because the retailer is the closest echelon to the customer and he does not experience the bullwhip effect as much as the other echelons. After a number of periods, the retailer participants possibly figure out that the customer demand they face is flat at 8 cases per period. Understanding this, as Figure 3-10 and Figure 3-11 illustrate, the retailers generally decrease their orders after around 15 periods of the experiment (especially in the modified experiments). This decrease affects the other echelons with some delay.

Anticipating the flat nature of the demand does not isolate the retailer from the bullwhip effect. This is because of his supply uncertainty. Recall that wholesaler often cannot supply the retailer's orders from his stock and falls into backlog. These backlogged orders will be met once the wholesaler obtains sufficient units from the distributor, which also takes time. As soon as these backlogged "high" orders are satisfied, they start pouring on the retailer, subject to the shipping delay. The retailer's late "break" on the orders can only mitigate the bullwhip effect.

We analyze the cost variances as well. The fluctuation of costs would be relevant for a risk averse manager. A "risk neutral manager" considers the mean value of the cost whereas a "risk averse manager" considers not only the mean but also the cost variance. In real life, managers are generally known to be risk averse in making decisions and they are afraid of the cost variances. As demonstrated in Table 3-11, in the modified experiments, "average of total cost variances" and "average of backlog cost variances" are decreased at wholesaler, distributor and factory echelons whereas they are increased at the retailer echelon relative to the standard experiments. Every echelon experienced the reduction in inventory cost variance. Appendices H and I provides detailed tables of cost comparisons.

Next, we provide the related hypothesis tests.

Hypothesis 5: The modified experiments will decrease the mean and variance of the total, inventory and backlog costs compared to the standard experiments.

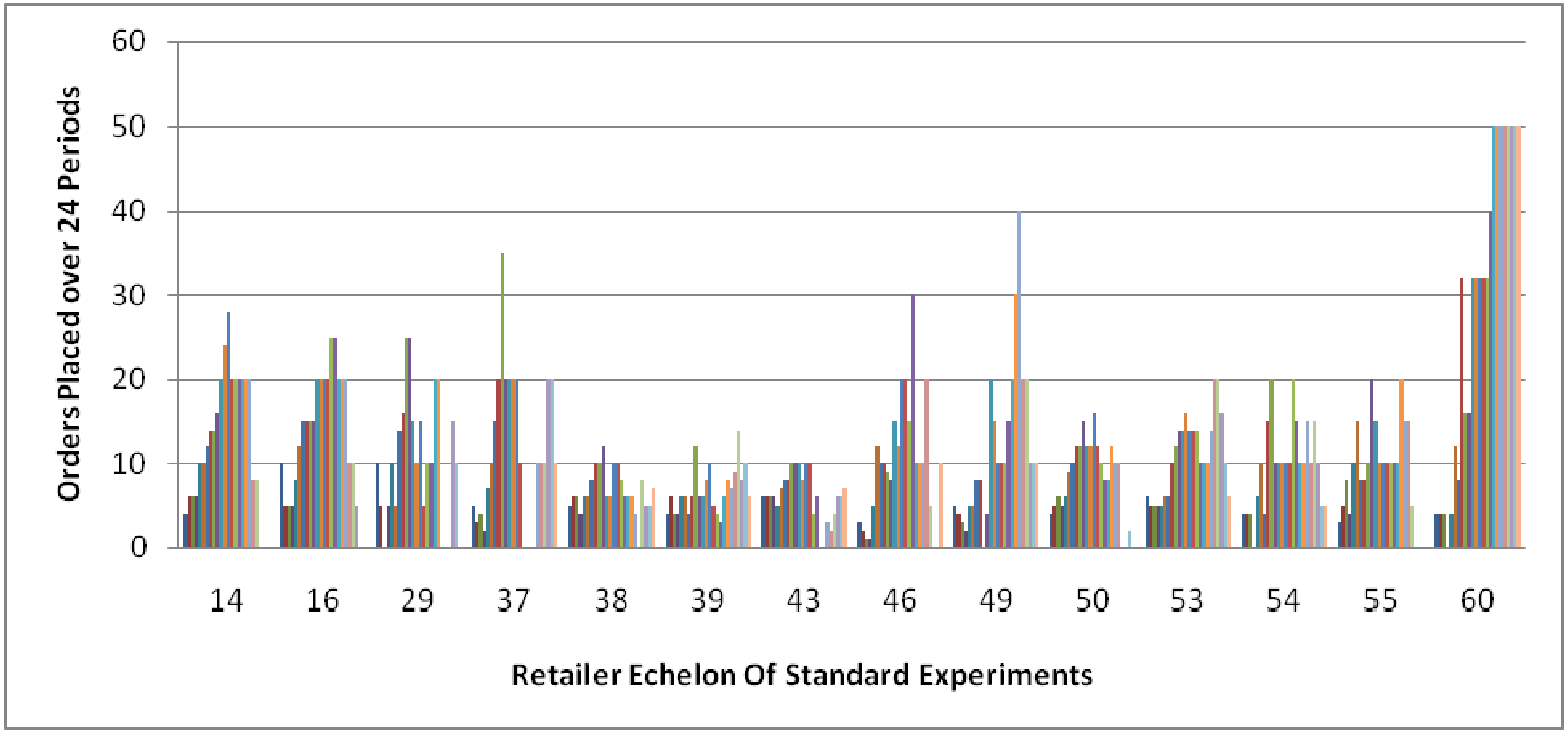


Figure 3-10: Orders Placed by Each Retailer in Standard Experiments

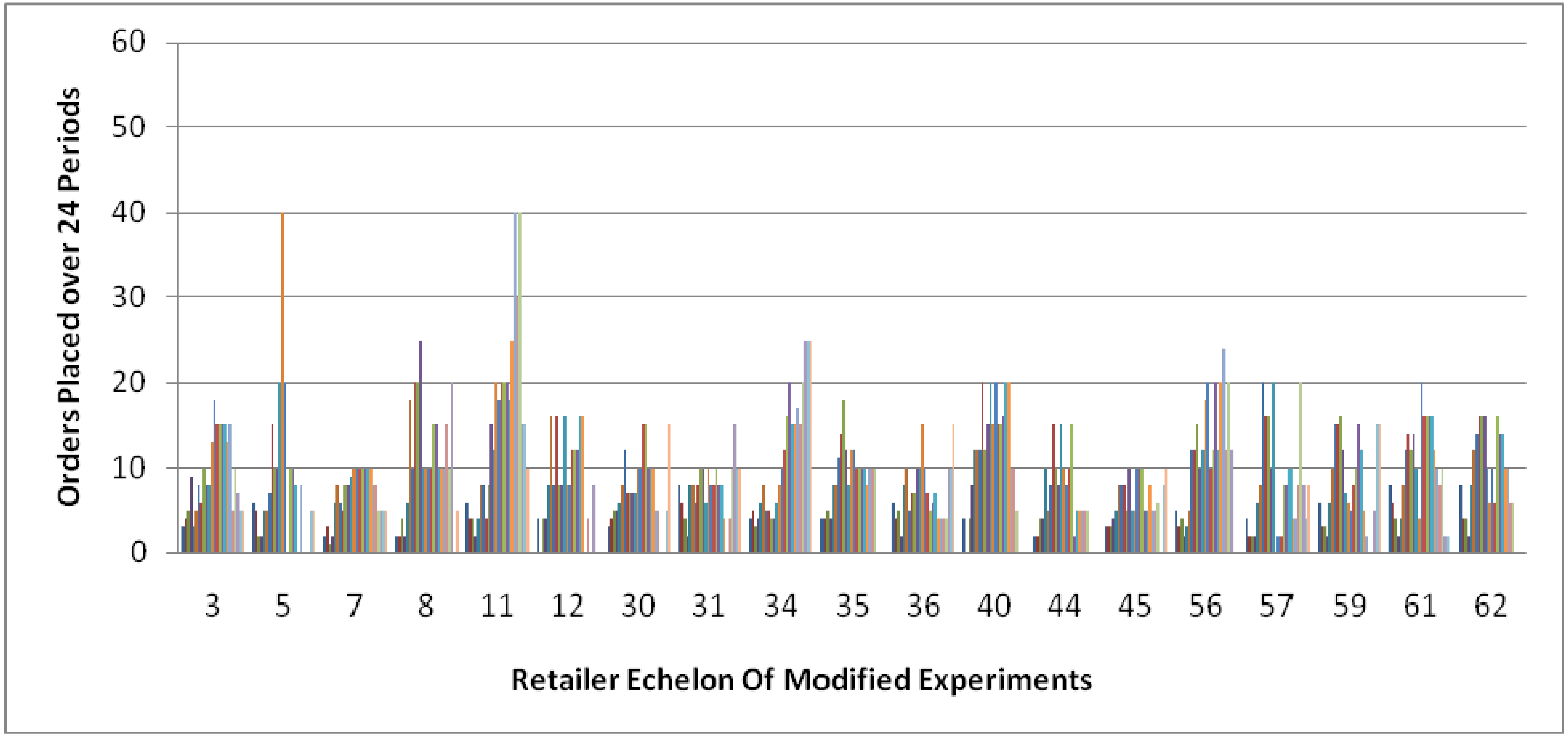


Figure 3-11: Orders Placed by Each Retailer in Modified Experiments

We make comparisons across the total cost as well as the inventory and backlog costs separately. We do not list each hypothesis one by one because indicating the significance of reduction is the basic idea for all of the variables. From Table 3-11, recall that we could not observe reductions in the average values of some cost variables. Considering this, we perform one tailed hypothesis tests to show significance of reductions for variables in which we observed decreases. For variables in which we observed an increase, our aim is to show that the increase is not significant.

Total (the sum of inventory and backlog) cost results are presented in the first row of Table 3-12. We observe that the decrease in the mean total costs of the supply chain (SC) is not significant (n= 56, m= 76, U= 1956.5, p= 0.216), whereas the decrease for the upstream echelons (D, F) is weakly significant (n= 28, m= 38, U= 422, p= 0.078).

Table 3-12: *P* Values of Hypothesis Tests for Total Costs

Variables \ Echelon	**SC**	**R, W**	**D, F**	**R**	**W**	**D**	**F**
Mean Total Costs	0.216	0.287	0.078*	0.099*	0.340	0.253	0.076*
Total Cost Variances	0.094*	0.371	0.024**	0.152	0.314	0.143	0.045**

Table 3-13 presents the p values regarding the inventory cost comparisons. The significant decreases were found at the retailer and wholesaler echelons.

Table 3-13: *P* Values of Hypothesis Tests for Inventory Costs

Variables \ Echelon	**SC**	**R, W**	**D, F**	**R**	**W**	**D**	**F**
Mean Inventory Costs	0.500	0.478	0.417	0.055*	0.072*	0.475	0.411
Inventory Cost Variances	0.375	0.501	0.313	0.103	0.135	0.432	0.353

Table 3-14 presents the p values regarding the backlog cost comparisons. We observe strongly significant reductions for the upper echelon, the factory, and the retailer echelons. Together with Table 3-13 results, we observe that the modified experiments achieve inventory cost reduction in the retailer echelon, and backlog cost reduction in the upstream echelons.

Table 3-14: *P* Values of Hypothesis Tests for Backlog Costs

Variables \ Echelon	SC	R, W	D, F	R	W	D	F
Mean Backlog Costs	0.204	0.244	0.04**	0.045**	0.286	0.209	0.048**
Backlog Cost Variances	0.180	0.199	0.028**	0.057*	0.408	0.219	0.033**

Overall, the supply-chain (SC) *p* values are not illustrating a significant reduction. The significant reductions we observed in the mean orders and order variances did not lead to significant reductions in total, backlog or inventory costs of the total supply chain. We expect that we may observe a significant reduction by increasing our sample sizes through conducting more experiments.

3.2.3.6 Analysis with Median Values

So far, we have reported the "averages" of observed values. An alternative is to use the "median", which is defined as the middle value when the observations are ordered from smallest to largest in magnitude (Devore 1995). One advantage of median over average is that it is less affected by the extremes values in data. We chose to report our main findings using averages, because this is the more common approach in literature. Wu and Katok (2006) and Croson et al. (2005) are among the researchers that report median values in their bullwhip effect studies.

Table 3-15, Table 3-16, Table 3-17, Table 3-18, Table 3-19, Table 3-20 and Table 3-21 report the median values. Comparing these tables with Table 3-2, Table 3-4, Table 3-6, Table 3-7, Table 3-8, Table 3-9 and Table 3-11, we observe that the general results are consistent to what is obtained with the average values. The median values are increasing as one moves upstream in the supply chain in both game types. The values in the modified game are generally lower than their counterparts in the standard game. As expected, the median values are less than their average counterparts because our data has a number of large values even after outlier elimination.

Table 3-15: Median Order Variance Comparison

Measurement Unit	Experiment Type	R	W	D	F
Median of Order Variances over Teams	Standard	43.91	88.47	368.8	555.14
	Modified	28.75	64.81	116.9	292.52

Table 3-16: Median Amplification Ratio Comparison

Measurement Unit	Experiment Type	W / R	D / W	F / D
Median of Amplification Ratios over Teams	Standard	2.58	2.26	1.56
	Modified	2.63	1.99	1.98

Table 3-17: Median Peak Orders Comparison

Measurement Unit	Experiment Type	R	W	D	F
Median Period of Peak Orders	Standard	13.00	15.50	17.00	19.00
	Modified	12.00	14.00	15.00	16.00
Median Magnitude of Peak Orders	Standard	22.50	35.00	75.00	95.00
	Modified	18.00	30.00	40.00	60.00

Table 3-18: Median Peak Backlogs Comparison

Measurement Unit	Experiment Type	R	W	D	F
Median Period of Peak Backlogs	Standard	18.00	17.50	19.00	19.00
	Modified	19.00	18.00	17.00	17.00
Median Magnitude of Peak Backlogs	Standard	-43.50	-72.00	-91.00	-79.50
	Modified	-50.00	-70.00	-66.00	-55.00

Table 3-19: Median First Backlogs Comparison

Measurement Unit	Experiment Type	R	W	D	F
Median Period of First Backlogs	Standard	6.0	6.5	9.0	10.0
	Modified	6.0	8.0	9.0	9.0

Table 3-20: Median Orders Comparison

Measurement Unit	Experiment Type	R	W	D	F
Median of Mean Orders over Teams	Standard	10.06	12.10	17.13	19.65
	Modified	8.04	10.38	11.63	13.92

Table 3-21: Median Cost Comparison

Measurement Unit	Experiment Type	Retailer			Wholesaler		
		Inv. Cost	Backlog Cost	Total Cost	Inv. Cost	Backlog Cost	Total Cost
Median of Mean Cost over Teams	**Standard**	1.29	29.67	33.85	1.40	52.50	53.79
	Modified	0.79	42.42	43.04	2.92	50.08	53.00
Median of Cost Variance over Teams	**Standard**	10.67	927.21	899.17	5.77	2864.43	2658.03
	Modified	2.52	1202.75	1091.01	22.17	2679.82	2182.78

Measurement Unit	Experiment Type	Distributor			Factory		
		Inv. Cost	Backlog Cost	Total Cost	Inv. Cost	Backlog Cost	Total Cost
Median of Mean Cost over Teams	**Standard**	5.48	45.08	59.06	7.92	26.58	35.00
	Modified	5.25	38.75	45.25	7.33	17.33	26.63
Median of Cost Variance over Teams	**Standard**	143.32	3949.61	3638.16	348.51	2322.46	2146.75
	Modified	68.46	2442.72	2086.61	100.93	874.75	1171.64

Chapter 4

The Second Study: Determining the Behavioral Factors Affecting Order Decisions

In the second study, we aim to determine the behavioral factors that affect the ordering decisions of the participants. Understanding these factors might enable supply chain managers to develop effective policies to counter the bullwhip effect. For instance, Croson and Donohue (2006) show that participants often underweight the supply line. That is, they do not value the orders that are already coming, or they forget about them while they are placing a new order. If one can show this effect on participant data, one can then recommend policies to practitioners that would address this behavioral factor. For example, the firms might invest in supply chain visibility software that would remind decision makers what orders are already coming, and when they will come.

To obtain insights about the participants' decision-making strategies, we conducted a post-experiment survey. We asked the participants what their ordering strategy was during the experiment. The most frequent answers were:

- I followed the orders I received from my downstream partner (i.e., I ordered what was demanded from me)
- I tried to simultaneously minimize the backlog and inventory levels.
- I tried to keep some safety stock against backlogs because the cost of a backlog is twice the cost of inventory holding.

Even these responses indicate that the ordering polices of participants might be quite different from each other. The participants are not equal in terms of the importance they place in achieving the trade-off between backlogs and inventory holding. They also seem to react differently to the delays in the system that makes matching supply and demand difficult. These observations suggest that it is not easy to determine generally-applying weights on behavioral factors that determine the ordering decision. Hence, our study will focus on the ordering strategy of each individual separately. We will then try to see if some general conclusions can be drawn.

In the post-experiment survey, we also asked the participants to draw their prediction of the exogenous customer demand that the retailer echelon faced. The retailer participants knew that demand was 4 cases/ period in the first four periods, and 8 cases/ period afterwards. However, almost all other participants came up with a prediction more or less similar to the one below, which was submitted by a factory participant:

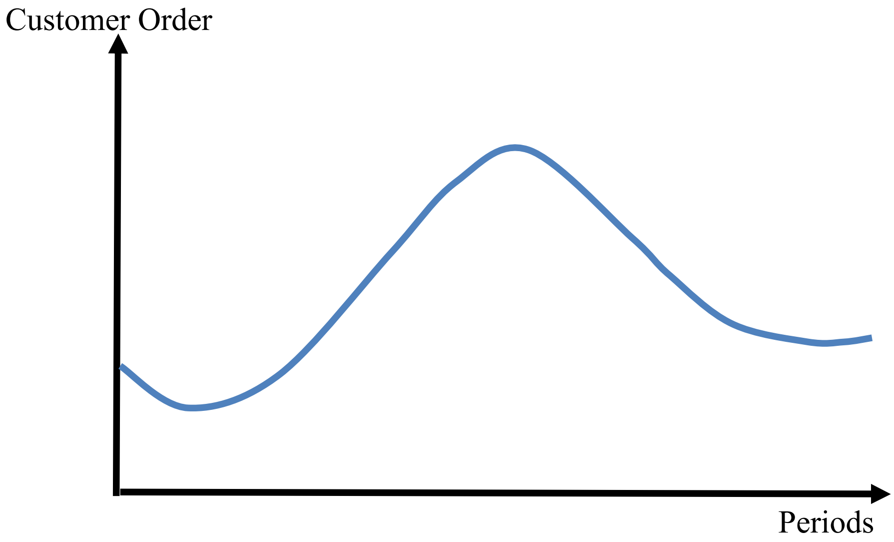


Figure 4-1: Predicted Customer Demand Drawn by One of Factory Participants

This prediction confirms the existence of the bullwhip effect. After the initial 4-5 periods, all echelons thought that the orders will be increasing. Anticipating the true pattern of the customer demand (which is flat at 8 cases/ period) after a number of periods, the retailer stopped placing high orders but this took some time to propagate in the supply chain.

4.1 The Candidate Factors

The key to understanding the bullwhip effect is to determine what factors the participants at different echelons considered in their ordering decisions. To this end, we developed a number of multiple linear regression models to predict *O(t)*, the order placed at the end of period *t*. The following is a list of the candidate independent variables for the regression model, (i.e., the candidate factors):

- *The demand faced at period t, D(t) :* The participant observes the demand that he needs to satisfy before making his own order decision. It is natural to expect that a high demand will positively affect the order quantity.

- *Inventory/backlog level:* This measures the inventory/ backlog level after the incoming shipment is taken in and after the faced demand is met. We predict that in general, the higher the current inventory level, the less the need to place a high order. Because the echelon might either have on hand inventory or may be in backlog, we use three different variables:
    - *Effective inventory at period t, EI(t) :* This is positive when there is on-hand inventory and negative when there is backlog. Using this variable alone ignores the fact that the cost of the backlog is twice the cost of on hand inventory. As a result, we also considered the following two separate variables:
    - *On hand inventory level, I(t),* which is the positive when there is on-hand inventory, and zero in case of backlog
    - *Backlog level, B(t),* which is positive if there is backlog, and zero in case of positive on hand inventory.

- *Whether in backlog or not, (If B(t)>0)*: Independent of the size of the backlog, being in backlog alone might cause the participants to panic and increase their order size. This factor is defined as a 1/ 0 variable.

- *Outstanding orders, O(t-1), O(t-2), O(t-3):* Outstanding orders refer to orders that were placed at the previous periods and that are currently on the supply line of an echelon (i.e., not received yet). When a participant in the retailer, wholesaler, or distributor echelon places an order to his upstream echelon, he receives cases after 4 periods due to the ordering and shipping delays. If the upstream echelon does not have sufficient inventory, the order might be backlogged and further delayed. Because of these delays and uncertainties in supply, the participants discount the value of their outstanding orders. Sterman (1989a) and Croson and Donohue (2006) show that participants undervalue, or simply forget incoming orders in the supply line while making order decisions. We aim to see how much underweighting our participants made. To this end, we used the outstanding order quantities placed three, two and one period ago as independent variables. Recall that the order placed four periods ago arrives at the echelon in the beginning of the period t (if it was filled by the upstream echelon).

- *The increase in demand over the last two periods, (D(t)-D(t-1)):* The long delays and supply uncertainty forces participants to forecast the demand for future periods. In particular, the increases in demand may lead the participants to assume an increasing trend to follow. This would make them increase their order sizes.

- *Whether there is an increase in demand over the last two periods, (If D(t)>D(t-1)):* This is the 1/ 0 version of the variable described above.

4.2 The Regression Models

Similar to Sterman (1989a) and Croson and Donohue (2006), we run regression analysis for each participant individually to detect how much weight, if any, participants place on these factors. To collect sufficient data points for each participant, we conducted *"long experiments",* which take 50 periods. The long experiments are played by four-participant teams, with one participant at each location (i.e., we are not interested in the modified experiment type here). The cost parameters and the experimental procedures are the same as the standard experiments of our first study, explained in Chapter 3. Different from the standard experiment, the long experiments begin with 12 cases on hand at each echelon. We conducted 7 of these long experiments with 28 participants.

Table 4-1 summarizes the 11 multiple linear regression models that we constructed to explain the ordering behavior of the participants. The table shows which independent variables (the factors that we explained above) were in a particular model, and what the average *adjusted* R^2 value of the model is, over 28 participants.

The 11 models can be divided into two. The first 7 models use "effective inventory *EI(t)*". The last 4 models use "On hand inventory *I(t)*" and "Backlog *B(t)*". Next, we provide details on two models from each group.

Table 4-1: Regression Models Summary

Model	**Independent Variables**						**Average *Adj.* R^2**
1	*EI(t)*	*D(t)*					0.524
2	*EI(t)*	*D(t)*	*O(t-1)*	*O(t-2)*	*O(t-3)*		0.623
3	***EI(t)***	***D(t)***	***O(t-1)+O(t-2)+O(t-3)***				**0.574**
4	*EI(t)*	*D(t)*	*O(t-1)+O(t-2)+O(t-3)*	*If D(t)>D(t-1)*			0.589
5	*EI(t)*	*D(t)*	*O(t-1)+O(t-2)+O(t-3)*	*(D(t)-D(t-1))*			0.600
6	*EI(t)*	*D(t)*	*If B(t)>0*				0.555
7	*EI(t)*	*D(t)*	*If B(t)>0*	*O(t-1)*	*O(t-2)*	*O(t-3)*	0.641
8	*B(t)*	*D(t)*	*I(t)*				0.573
9	*B(t)*	*D(t)*	*I(t)*	*O(t-1)*			0.624
10	*B(t)*	*D(t)*	*I(t)*	*O(t-1)*	*O(t-2)*	*O(t-3)*	0.650
11	***B(t)***	***D(t)***	***I(t)***	***O(t-1)+O(t-2)+O(t-3)***			**0.618**

We choose models 3 and 11 as the examples on which to provide analysis details, because the weights obtained from these two models have the best consistency between participants. Also, both of these models consider the sum of all outstanding orders as a factor, rather than each order separately. This may be more realistic because the participants are more likely to remember the total outstanding orders than remembering each individual order separately.

4.2.1 Observations on Model 3

This model aims to explain the orders placed at period *t* by using

1) the effective inventory at period *t*
2) demand faced at period *t*
3) total outstanding orders (the supply line) as of period *t*

Factors 1 and 3 was to give us an idea about the level of supply line underweighting.

Before beginning the analysis, we checked a number of regression assumptions. The details can be found in Appendix J. In summary, we found that regressions are highly significant ($p<0.05$) for 26 out of 28 participants. Normality assumption of residuals does not hold for 6 out of 26 participants. These six participants were eliminated from further analysis. All *VIF* values are found to be less than 10, which indicates that there is no multicollinearity. Durbin Watson tests show that there is no autocorrelation between the residuals.

For the remaining 20 participants, Table 4-2 shows the *adjusted* R^2 values and standardized beta coefficients. The average adjusted coefficient of determination (*adj.* R^2) value over 20 participants is 59.7%. Given the complexity of the game and the number of potential behavioral factors, we believe that this is a reasonable *adjusted* R^2 value. In fact, other researchers have also achieved similar R^2 values (see, for example Croson and Donohue 2006).

Next, we check the signs of the (beta) coefficients. In parallel with our expectations, the demand coefficients are positive for most participants (19 participants). The average demand coefficient of 0.26 seems reasonable. The effective inventory coefficients are negative for most participants (19 participants). The average effective inventory coefficient of -0.40 also looks reasonable.

Table 4-2: Standardized Beta Coefficients for Model 3

Participants	**Echelon**	***Adj. R2***	**Standardized Coefficients**		
			EI(t)	***D(t)***	***∑O(t-i)***
1	Factory	60.71%	-0.39	0.40	0.13
2	Distributor	60.93%	-0.51	0.12	0.23
3	Wholesaler	72.61%	-0.23	0.39	0.46
4	Retailer	23.90%	0.07	0.06	0.55
5	Factory	68.37%	-0.44	0.57	-0.17
6	Distributor	86.90%	-0.58	0.48	-0.05
7	Wholesaler	42.04%	-0.12	0.42	0.30
8	Retailer	55.85%	-0.51	0.12	0.33
11	Wholesaler	83.67%	-0.41	0.21	0.39
12	Retailer	82.76%	-0.61	0.12	0.37
15	Wholesaler	59.61%	-0.56	0.33	0.20
16	Retailer	23.17%	-0.31	0.13	0.27
18	Distributor	24.15%	-0.63	-0.36	0.13
19	Wholesaler	68.60%	-0.24	0.13	0.55
21	Factory	75.85%	-0.26	0.44	0.32
22	Distributor	62.52%	-0.30	0.45	0.35
25	Factory	82.18%	-0.20	0.52	0.32
26	Distributor	81.88%	-0.25	0.50	0.29
27	Wholesaler	61.41%	-0.70	0.13	0.05
28	Retailer	18.17%	-0.73	0.01	-0.43

However, in contrast to expectation, the outstanding orders coefficients are positive for most participants (17 participants). This means that the orders have a tendency to be larger when the outstanding orders are large. Ideally, when one has a high value of outstanding orders, he does not need to order more, given that these orders will be arriving in the following periods. However, the general ordering behavior of the participants is to increase their orders in the first half of the experiment, and decrease in the second half. Thus, higher orders are likely to follow each other. One might think this behavior to cause "autocorrelation" between the residuals in the regression analysis, however, as we mentioned in the beginning, autocorrelation is not found in the data of these 20 participants.

After conducting regression analysis, we try to detect whether there exists a significant difference in adjusted R^2 values across echelons or not. Therefore, we conduct Mann Whitney tests for each echelon pair. Hypothesis tests show that there is not a significant difference in R^2 values across echelons.

4.2.2 Observations on Model 11

This model aims to explain the orders placed at period *t* by using
1) inventory on hand at period *t*
2) backlog at period *t*
3) demand faced at period *t*
4) total outstanding orders (the supply line) as of period *t*

Note that this model is different from Model 3 in that the inventory and backlog values are taken as separate factors. We expected the beta coefficient for factors 1 and 4 to be negative and the coefficient for factors 2 and 3 to be positive. The relation between the magnitudes of factors 1 and 2 shall give us an idea about the powers of the two sides of the inventory/ backlog trade-off.

Similar to analysis of the Model 3, before beginning the analysis, we checked a number of regression assumptions. The details can be found in Appendix J. Regressions are highly significant ($p<0.05$) for 26 out of 28 participants. The two participants with insignificant results are the same ones in Model 3. Normality assumption of residuals does not hold for 8 out of 26 participants. For one participant, (*VIF*) value of the backlog variable is greater than 10 which indicates multicollinearity. These nine participants were eliminated from further analysis. Durbin Watson tests show that there is no autocorrelation between the residuals.

For the remaining 17 participants, Table 4-3 shows the *adjusted* R^2 values and standardized beta coefficients for each participant. The average adjusted coefficient of determination (*adj.* R^2) value over 17 participants is 66.5%.

As expected, the demand and backlog coefficients are positive for most participants (14 and 16 participants out of 17, respectively) and the inventory coefficients are negative for most (16 participants out of 17). Similar to Model 3, the average demand coefficient is 0.26. The average backlog coefficient is 0.32, whereas the average inventory coefficient is -0.24. We observe that contrary to our expectation and similar to Model 3, the outstanding orders' coefficients were mostly positive (14 participants).

Table 4-3: Standardized Beta Coefficients for Model 11

Participants	**Echelon**	***Adj. R^2***	**Standardized Coefficients**			
			B(t)	***I(t)***	***D(t)***	***∑O(t-i)***
2	Distributor	60.26%	0.44	-0.12	0.12	0.23
3	Wholesaler	73.61%	-0.02	-0.22	0.40	0.55
4	Retailer	22.21%	-0.05	0.03	0.06	0.55
6	Distributor	86.63%	0.55	-0.07	0.48	-0.05
7	Wholesaler	41.88%	0.19	-0.03	0.38	0.26
8	Retailer	59.80%	0.24	-0.49	0.12	0.27
11	Wholesaler	85.40%	0.50	-0.39	0.00	0.18
12	Retailer	83.06%	0.16	-0.56	0.12	0.33
13	Factory	82.79%	0.34	-0.03	0.63	0.01
16	Retailer	23.54%	-0.06	-0.37	0.18	0.28
19	Wholesaler	75.51%	0.75	-0.10	0.11	0.03
20	Retailer	41.61%	0.39	-0.48	0.26	-0.15
21	Factory	82.22%	0.39	-0.12	0.29	0.28
22	Distributor	79.45%	0.76	-0.04	0.33	-0.07
25	Factory	85.04%	0.33	-0.20	0.35	0.22
26	Distributor	87.71%	0.43	-0.29	0.39	0.05
27	Wholesaler	60.56%	0.18	-0.59	0.13	0.05

Similar to Model 3, there is not a significant difference in adjusted R^2 values across echelons in Model 11.

4.2.3 Observations on Stepwise Regression Models

We also conduct stepwise regression analysis for each participant in order to show that the factors that affect ordering decisions might change from person to person. The first stepwise regression model (SRM1) uses independent variables of Models 1 to 7 (see Table 4-1). The second stepwise regression model (SRM2) uses independent variables of Models 8 to 11 (see Table 4-1).

Average adjusted R^2 over 28 participants is 60.40% in the SRM1. Consistent with our expectations, we find that different factors are important for different participants. However, the effective inventory (16 out of 28), demand (16 out of 28) and orders placed one period ago (13 out of 28) are the factors that are important for most of the participants as seen in
Table 5-27.

Average adjusted R^2 over 28 participants is 59.69% in the SRM2. Similar to SRM1, backlog (15 out of 28), demand (14 out of 28), inventory (10 out of 28) and the orders placed one period ago (14 out of 28) are the factors that are important for most of the participants as seen in Table 5-28.

Chapter 5

Conclusions and Directions for Future Research

In this thesis, we present two studies related to the bullwhip effect. In the first study, we proposed a version of the beer game with two participants at each echelon with conflicting incentives regarding the order decision. To the best of our knowledge, this has not been studied in the literature. Such a decision structure reflects the well-known incentive conflict between the sales and operations functions of a firm, particularly in the Sales and Operations Planning process. Our expectation was that with two participants that represent the two sides of the order decision trade-off, the bullwhip effect will be dampened relative to a standard beer game. Our observations in this study are as follows:

1) Bullwhip effect exists. We observed the three characteristics of the bullwhip effect in most of our experiments. Order levels are oscillating, these oscillations are amplified towards upstream echelons and there exist time lags between echelons.

2) The results exhibit high level of variation among teams. Hence, generalizations are difficult. Studies that report regression models that are based on multiple participant data should be carefully interpreted.

3) On average, in the modified experiments, relative to the standard experiments

- Order variances (oscillations) reduced at each echelon of the supply chain. The reduction is strongly significant at the supply chain and upstream echelons level.
- The amplification ratio decreased between wholesaler/ retailer and distributor/ wholesaler echelons. However, the reductions are not statistically significant.

- The participants reach their peak order levels faster.
- The total cost and backlog cost decreased at all echelons except the retailer. Most of the decreases are not statistically significant.

Supply risk turned out to be an important factor for the bullwhip effect. Supply risk for the factory is zero because the factory knows that it will receive its orders for sure. Other echelons' orders, however, may not be filled by their upstream partner, which causes them to place even larger orders in return. This turned our supply chains into "backlog chains" where most echelons' effective inventory levels are negative for many number of periods, and where inventory holding costs are dwarfed by the backlog costs.

In the second study, we tried to determine the behavioral factors affecting the participants' ordering decisions. We tried a number of different regression models and focused on the most promising two that have quite consistent coefficient signs across participants. The results of the second study are as follows.

We observed that the participants are seriously underweighting the supply line, consistent with Sterman (1989a) and Croson and Donohue (2006)'s results. In fact, contrary to expectations, the coefficient of outstanding orders in regressions turned out to be positive for most participants. We believe the explanation is related to supply risk. Orders are not satisfied, which lead to the placement of higher orders. Hence, there exists some level of autocorrelation in order series, although it is not at a level to hinder the regression study.

With the exception of Oliva and Gonçalves (2007), the literature studies effective inventory as a single factor in regression studies. In some of our models, we analyzed the participants' reactions to backlog and inventory separately. Oliva and Gonçalves (2007) pooled all participants' data to conduct a single regression; whereas we studied participant-level regressions. These authors found that the subjects do not react to backlogs different from on-hand inventory. Our results indicate the opposite. The average absolute value of the backlog coefficient is higher than that of the on-hand inventory coefficient (which is negative). This holds true for most of the participants at the participant level as well.

Our study has certain weaknesses, and it can be extended in a number of directions. We discuss these in what follows.

Practicing managers as subjects

We conducted a high number of experiments with human decision makers from 2008 to 2010, using students from different departments. One might question the representativeness of students' results to the real managers' behaviors. Croson and Donohue (2006) report that their experimental results with students and real managers do not exhibit a significant difference. In a newsvendor experiment setting, Bolton et al. (2008) show that managers perform similar ordering behavior to the students. Yet, we are planning to conduct the same experiments with practicing managers for external validity. In particular, we aim to use practicing operations and sales managers to fill these roles experiments.

Other demand patterns and experiment settings

In our experiments, similar to the standard beer game, the customer demand to retailer is 4 cases/ period in periods 1 to 4, and 8 cases/ period afterwards. One might argue that this demand pattern is not realistic. Steckel et al. (2004) conduct experiments under different demand patterns and showed that the value of POS sharing and the impact of time lag reduction depend on the pattern of demand data. We also suspect that our results will depend on the demand pattern. However, we chose to stick to the standard step-up pattern in order to be able to compare our results with the literature. Likewise, a change in other experiment settings (such as the ratio of inventory holding and backorder costs, or the length of time delays) would also affect our results.

Computerized experiments

We conducted board game experiments. The board game environment offers its own advantages, allowing lively discussion between participants which is at the core of our modified experiments. In the future, we may conduct computerized versions of the game, keeping the same discussion environment in place. This would help us greatly in the data collection process and overcome manual data entry errors.

Contamination effect

Our experiments spanned a time period from 2008 to 2010. We could not have conducted all at once for practical reasons. Given this setting, there is the possibility that participants from earlier experiments share their knowledge with later players. This would undermine our assumption of using participants with zero experience. To overcome this, we asked the participants not to discuss their strategy with others. More effectively, there was usually plenty of time between different sets of experiments, which minimized the strength of any such knowledge transfer.

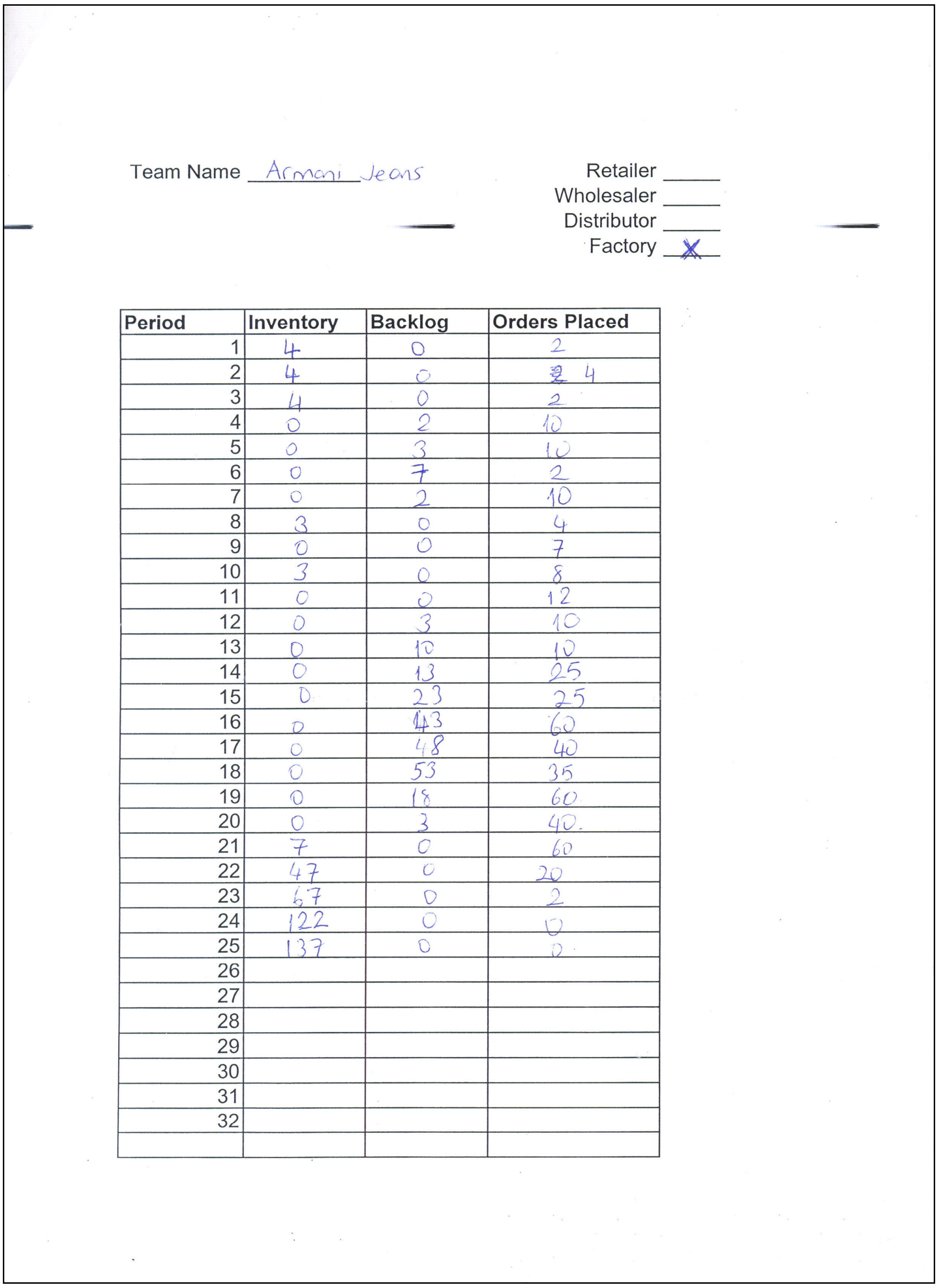
Monetary incentives

We did not provide monetary incentives to participants. We motivated them by announcing the winner team (and the winner supply and sales manager) at the end of the experiments. Although the participants indicated that they did not have a motivation problem, we may offer monetary incentives to formalize the process. This, of course, requires funding.

Future analysis

This study can be extended in many different directions. We provide two such directions as examples. First, we may conduct an "intertemporal" analysis by comparing the decisions among different time windows (such as periods 1-6, 7-12, 13-18, 19-24 in a 24-period experiment). This would give us better results regarding the "time lag" aspect of the bullwhip effect. Second, we can try to establish connections between personality characteristics of the participants and the bullwhip effect. We have already collected this information in post-experiment surveys, yet, we have not had a chance to analyze it as part of this thesis.

Appendix A: Record Sheet



Team Name Armani Jeans

Retailer ____
Wholesaler ____
Distributor ____
Factory X

Period	Inventory	Backlog	Orders Placed
1	4	0	2
2	4	0	~~2~~ 4
3	4	0	2
4	0	2	10
5	0	3	10
6	0	7	2
7	0	2	10
8	3	0	4
9	0	0	7
10	3	0	8
11	0	0	12
12	0	3	10
13	0	10	10
14	0	13	25
15	0	23	25
16	0	43	60
17	0	48	40
18	0	53	35
19	0	18	60
20	0	3	40.
21	7	0	60
22	47	0	20
23	67	0	2
24	122	0	0
25	137	0	0
26			
27			
28			
29			
30			
31			
32			

Figure 5-1: Record Sheet of One of Our Participants

Appendix B: Participants Information

Table 5-1: Participants Information

Total Number of Participants	**208**
Gender	
Female	42%
Male	58%
Age	
Average over All Participants	21.7
University	
Sabanci University	82%
Other	18%
Department	
Industrial Engineering	70%
Other Departments	30%
Motivation to Play the Game	
Liked the Game (0: No - 10: Yes)	7.5

Table 5-2: Attitude towards Risk and Service

Questions \ Average over All Participants	Average
Perception of Service Quality (0: Dissatisfied - 10: Satisfied)	6.6
Attitude towards Stock out (0: Angry - 10: Relaxed)	4.5
Tendency to Hold Inventory (0: No - 10: Yes)	5.0
Willingness to Wait (0: Not wait - 10: Wait)	4.3

Appendix C: Data Acquisition Process

1. Four graduation project groups helped us in conducting the experiments since 2008.
2. We trained the facilitators.
3. We arranged around 200 participants and organized the experiments.
4. Training and pilot experiments before real ones.
5. After the experiments, data is transferred to MS Excel, controlled and filtered. Some data eliminated at this stage.
6. Outlier elimination.
7. Descriptive analysis in MS Excel.
8. Statistical analysis with SPSS and Matlab.

Appendix D: Post-Experiment Survey

Name Surname	
Team Name	
Echelon	
Gender	
Work	
What is your favorite game?	
What is your favorite color?	
What is your favorite football team?	
Do you like the beer game? (0: No - 10: Yes)	
Perception of service quality (0: Dissatisfied - 10: Satisfied)	
Attitude towards stock out (0: Angry - 10: Relaxed)	
Tendency to hold inventory (0: No - 10: Yes)	
Willingness to wait (0: Not wait - 10:Wait)	
What was your ordering strategy ?	

Please draw your prediction of the exogenous customer demand that the retailer echelon faced.

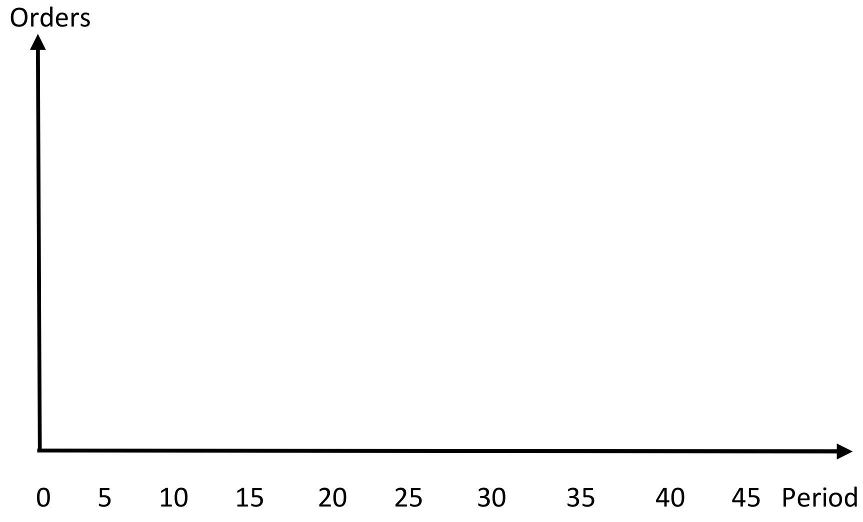
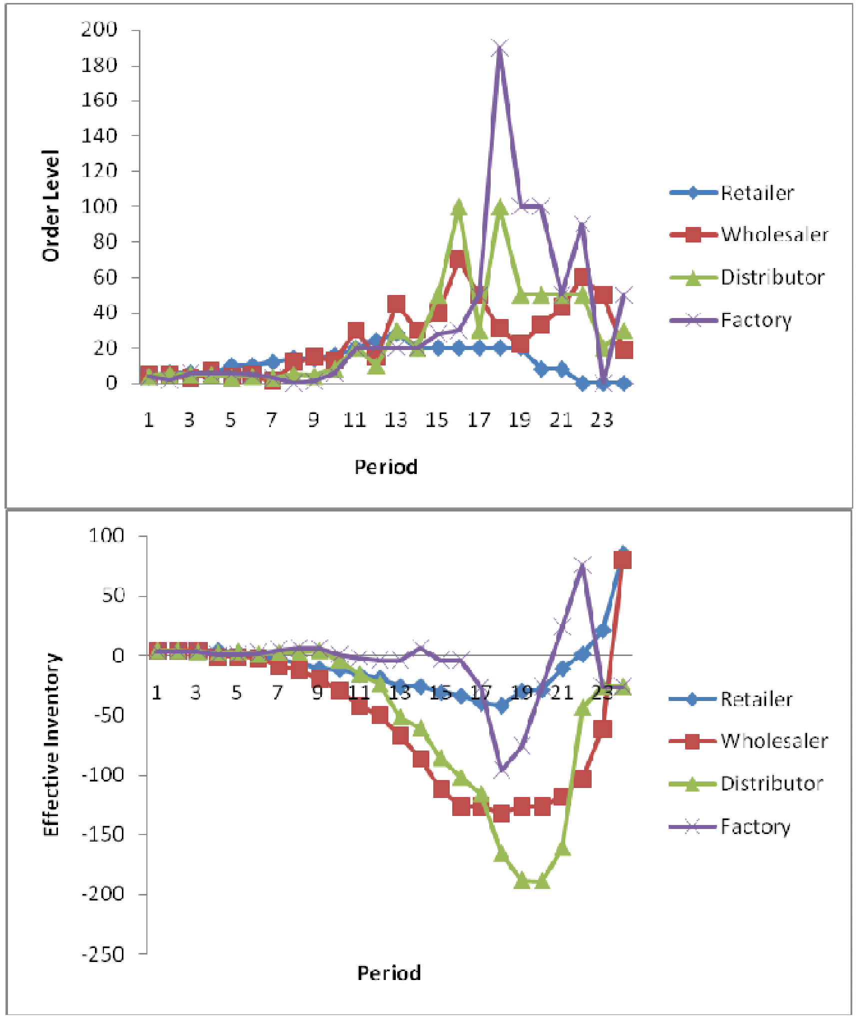
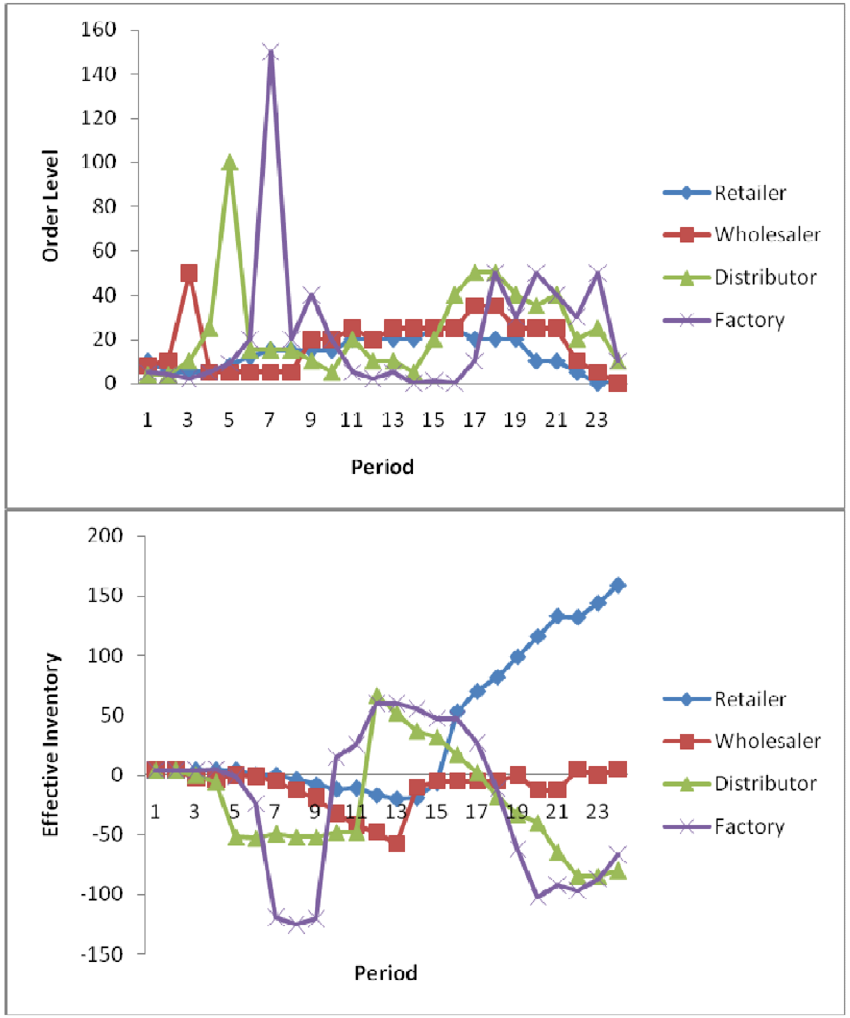


Figure 5-2: Post Experiment Survey

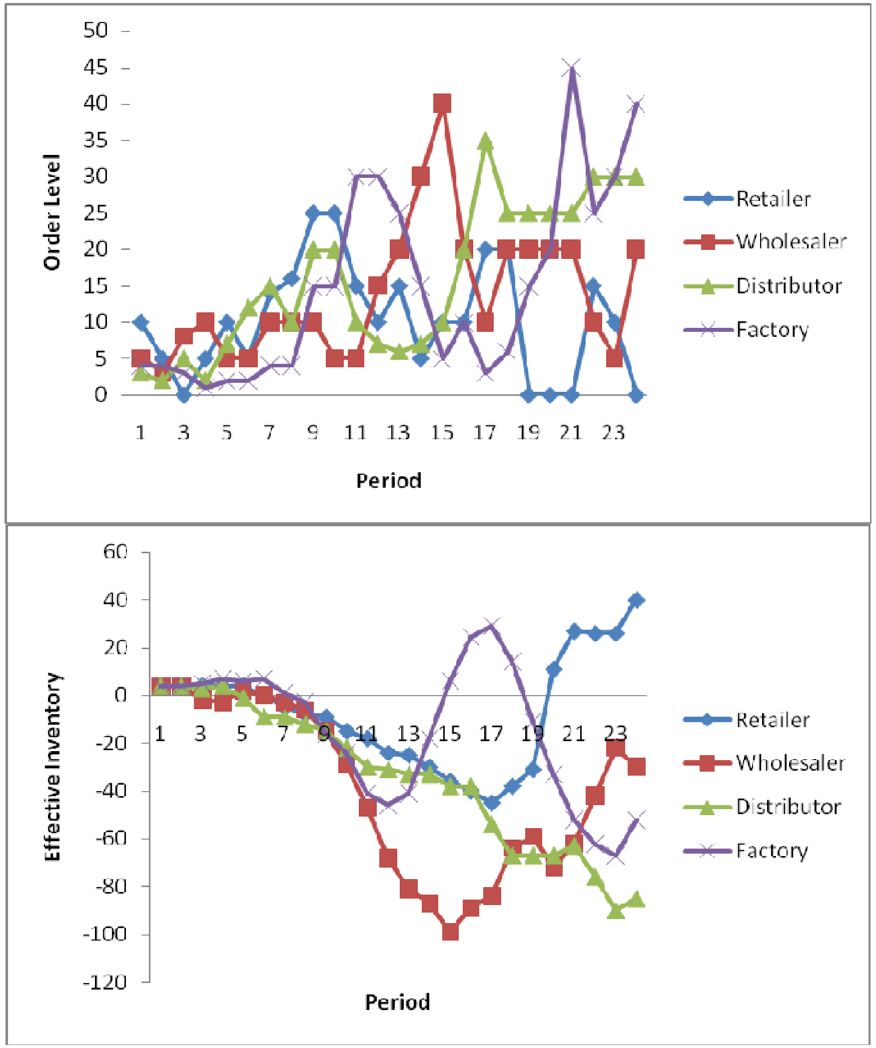
Appendix E: The Graphs of the Standard Experiments



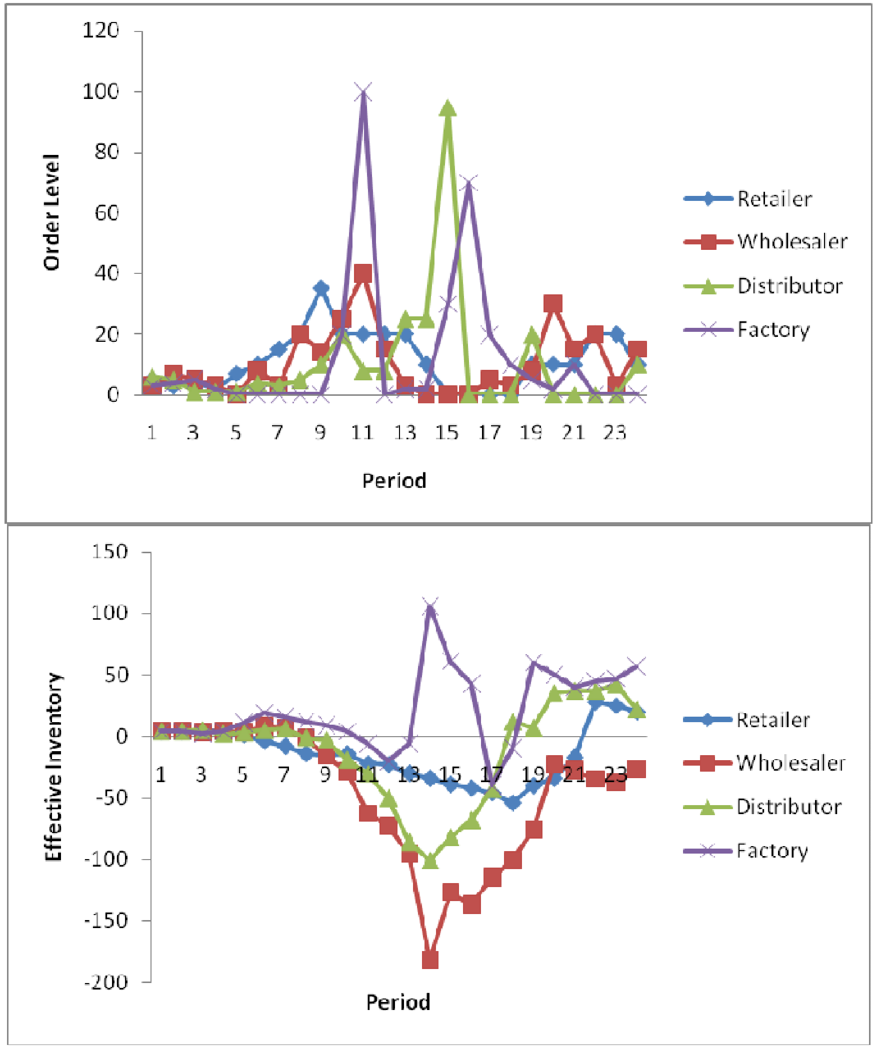
Order Level and Effective Inventory of Team 14



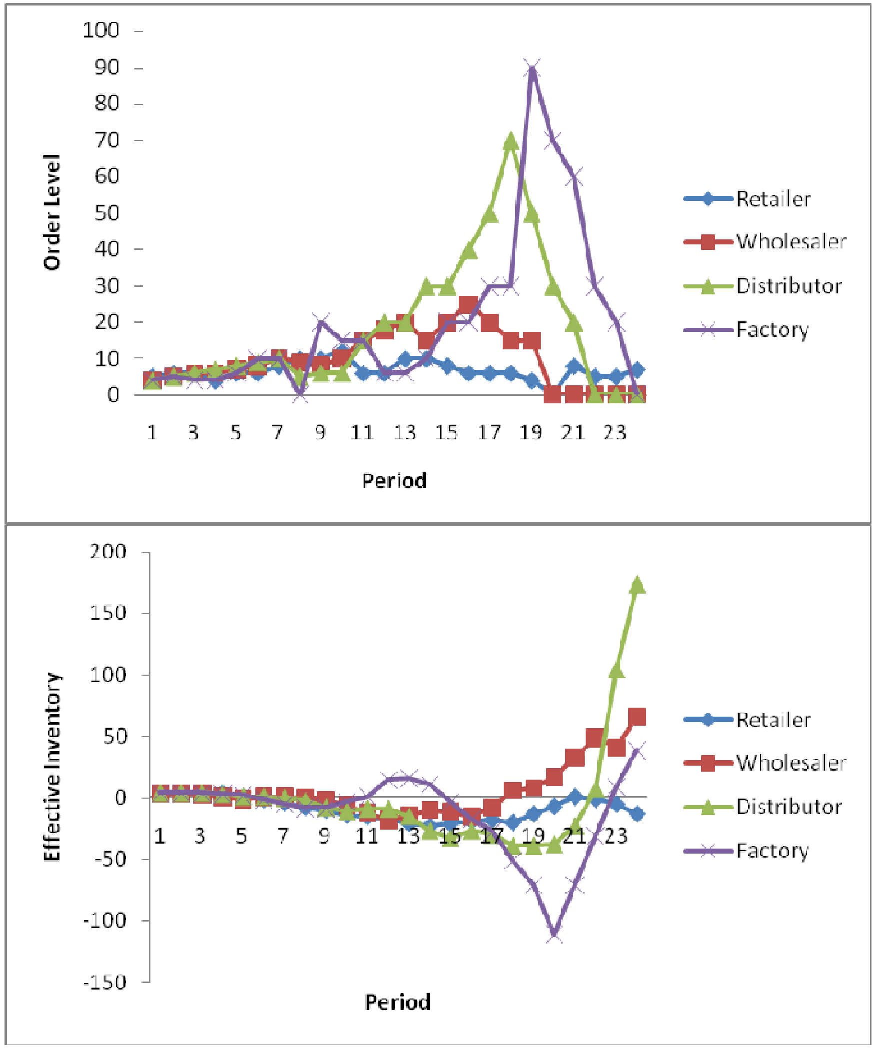
Order Level and Effective Inventory of Team 16



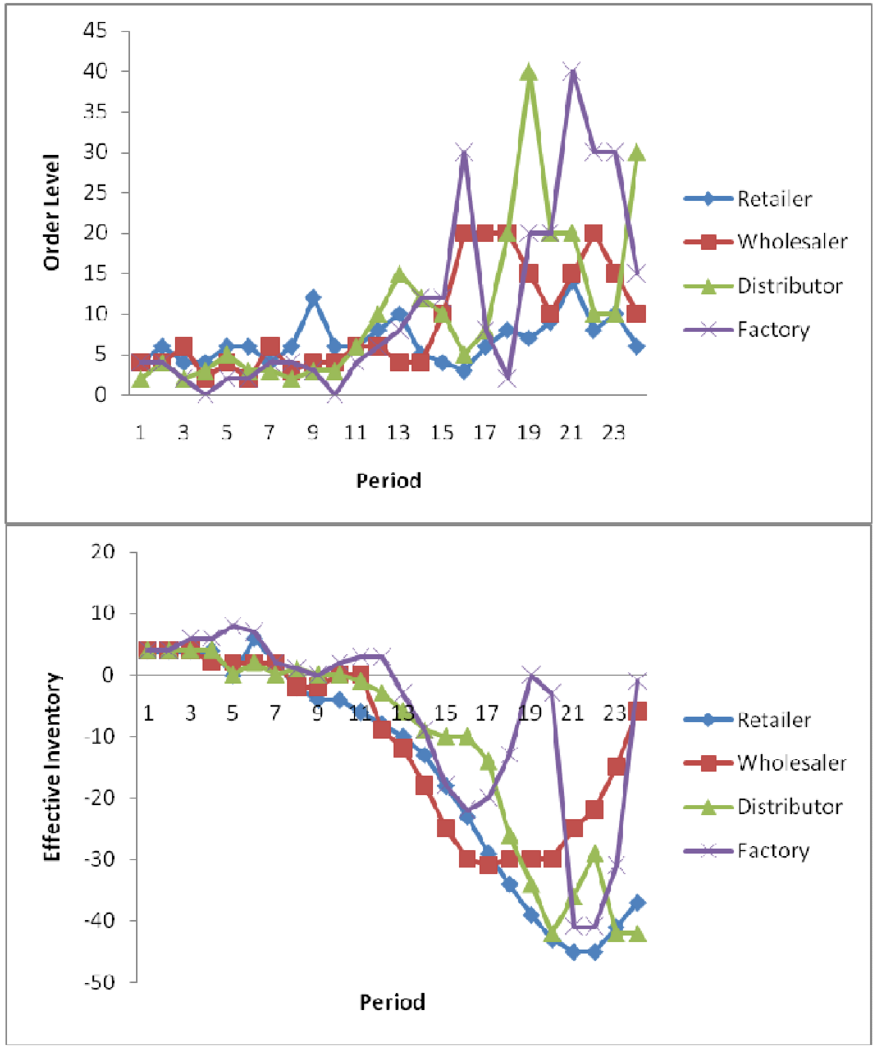
Order Level and Effective Inventory of Team 29



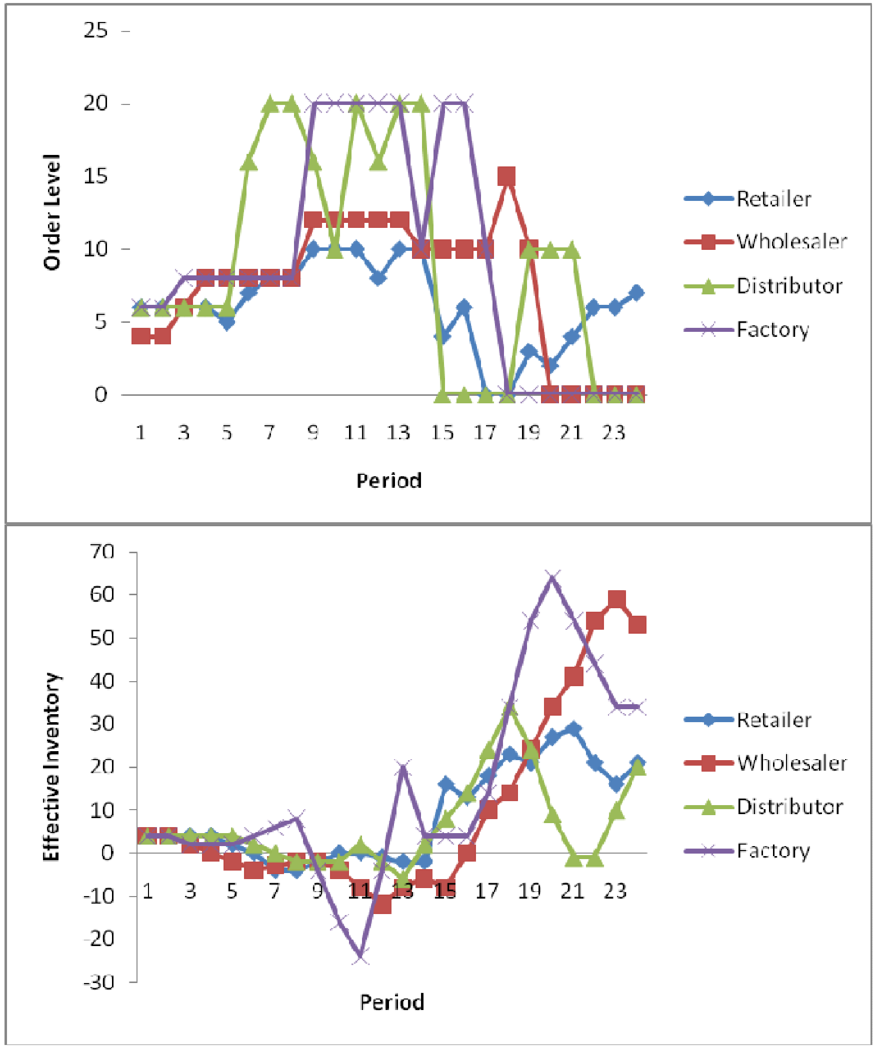
Order Level and Effective Inventory of Team 37



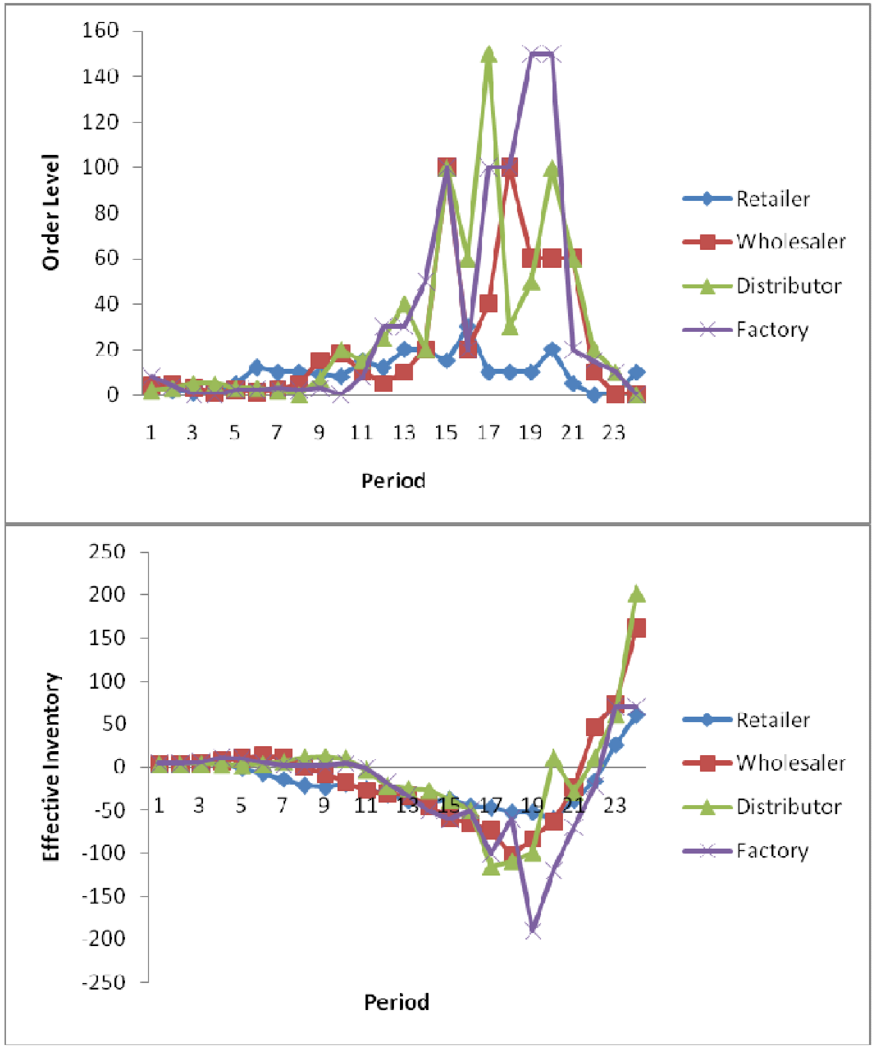
Order Level and Effective Inventory of Team 38



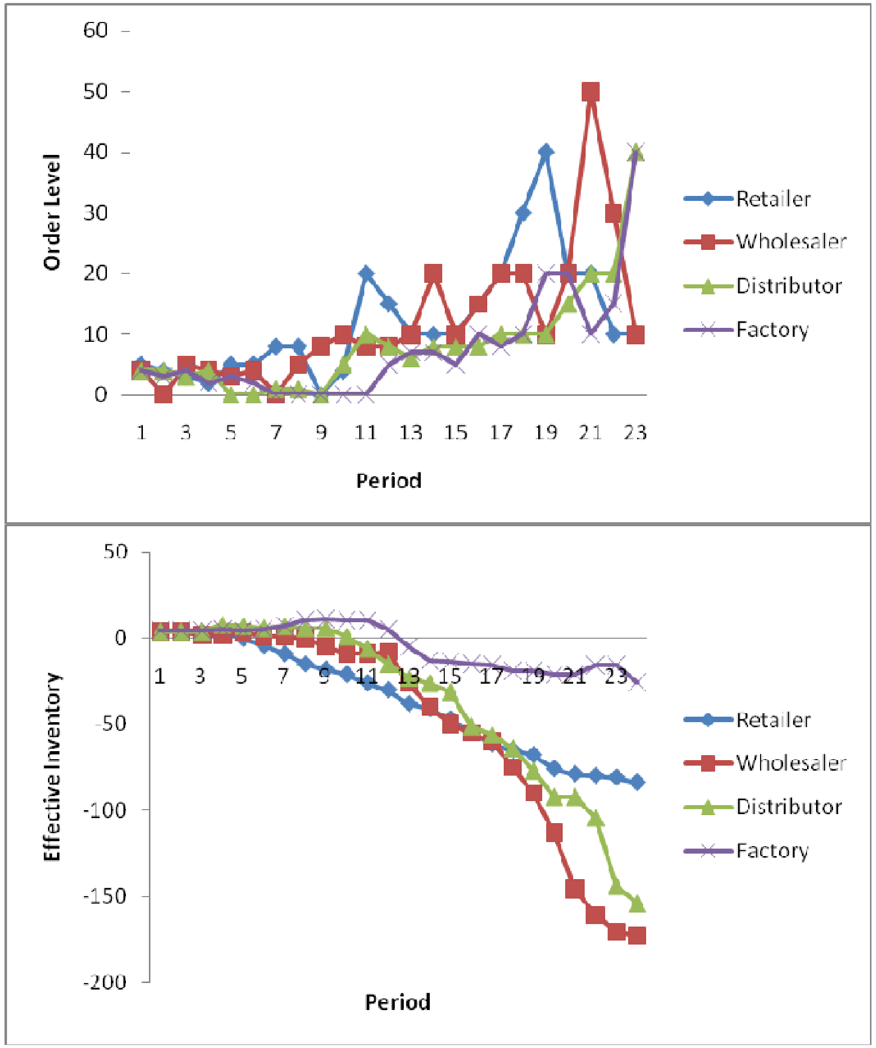
Order Level and Effective Inventory of Team 39



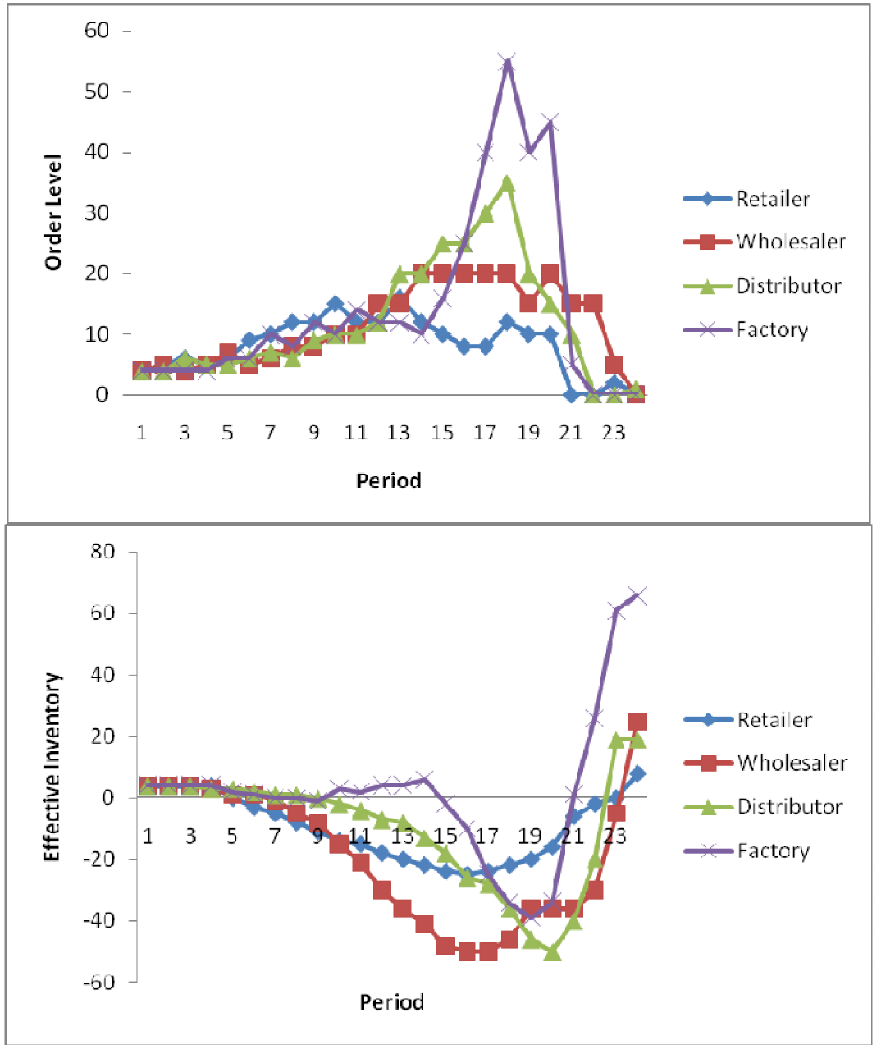
Order Level and Effective Inventory of Team 43



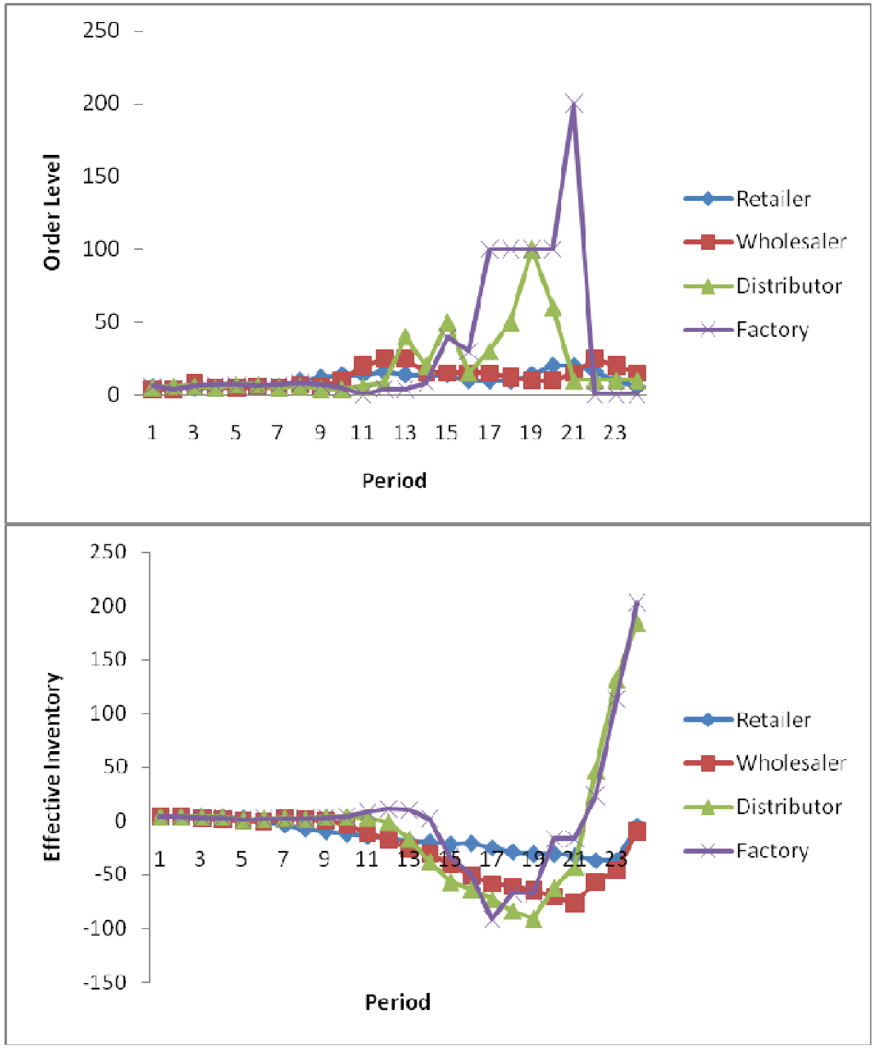
Order Level and Effective Inventory of Team 46



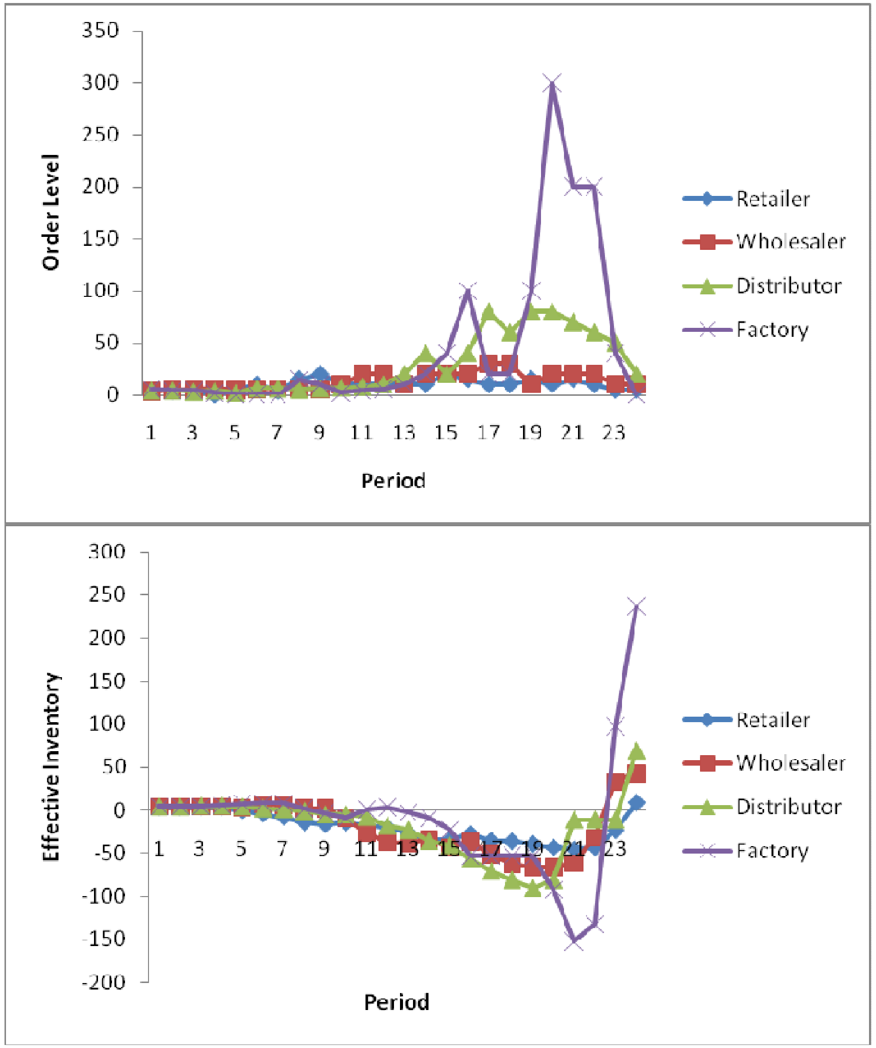
Order Level and Effective Inventory of Team 49



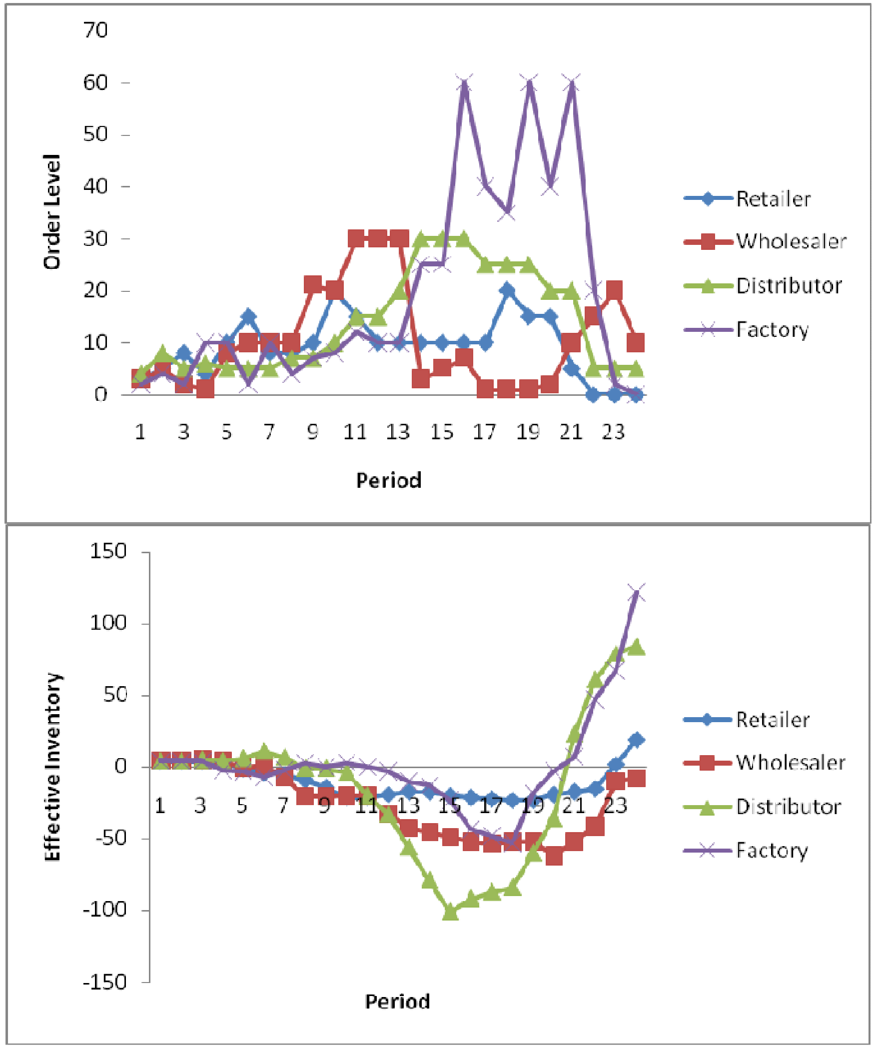
Order Level and Effective Inventory of Team 50



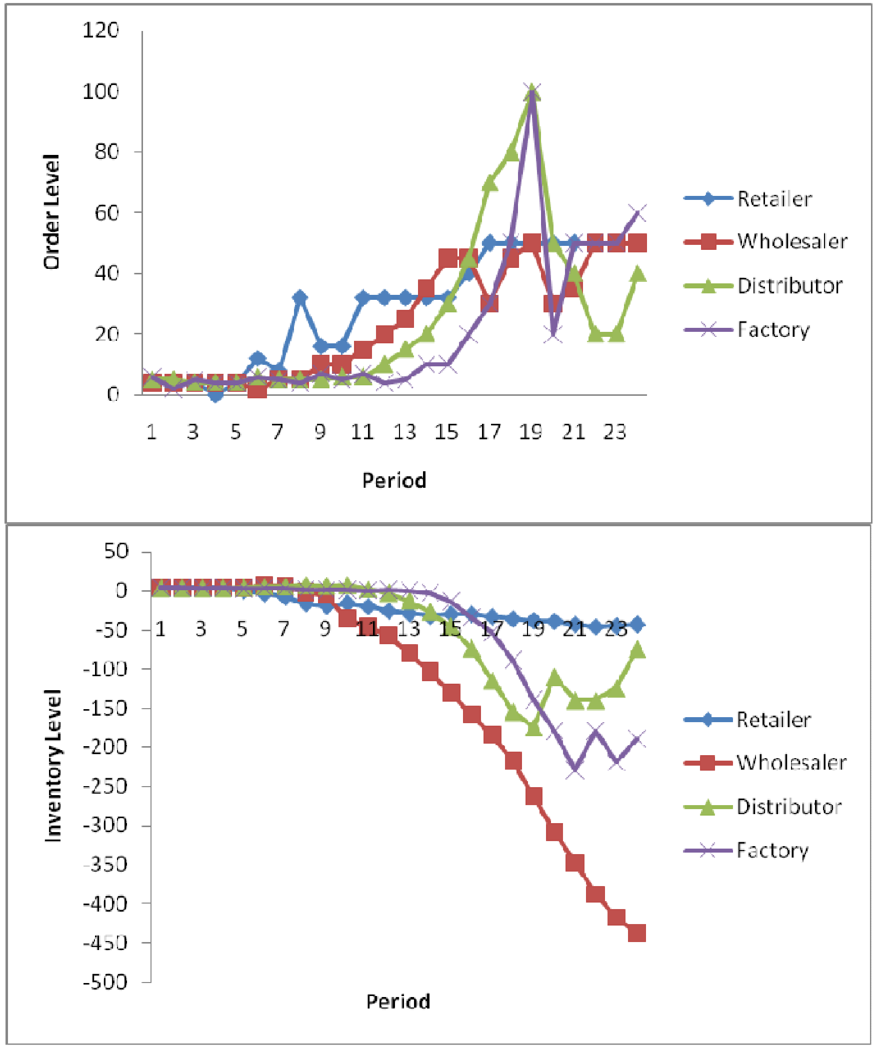
Order Level and Effective Inventory of Team 53



Order Level and Effective Inventory of Team 54

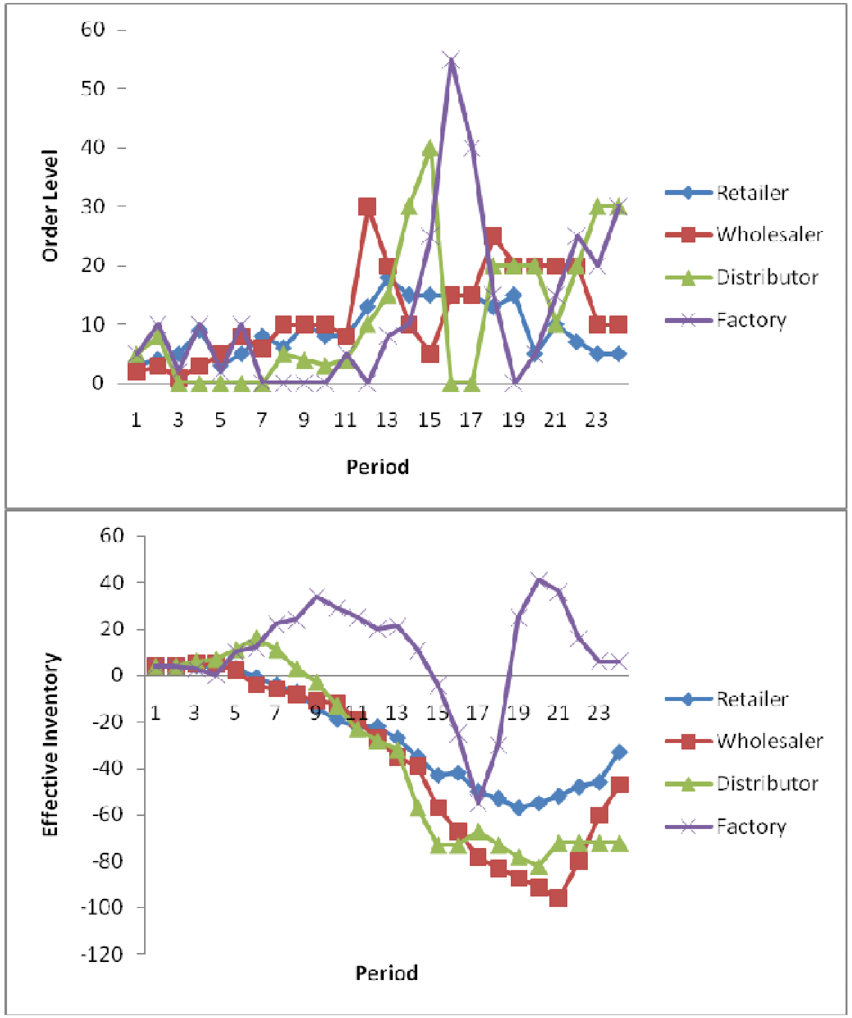


Order Level and Effective Inventory of Team 55

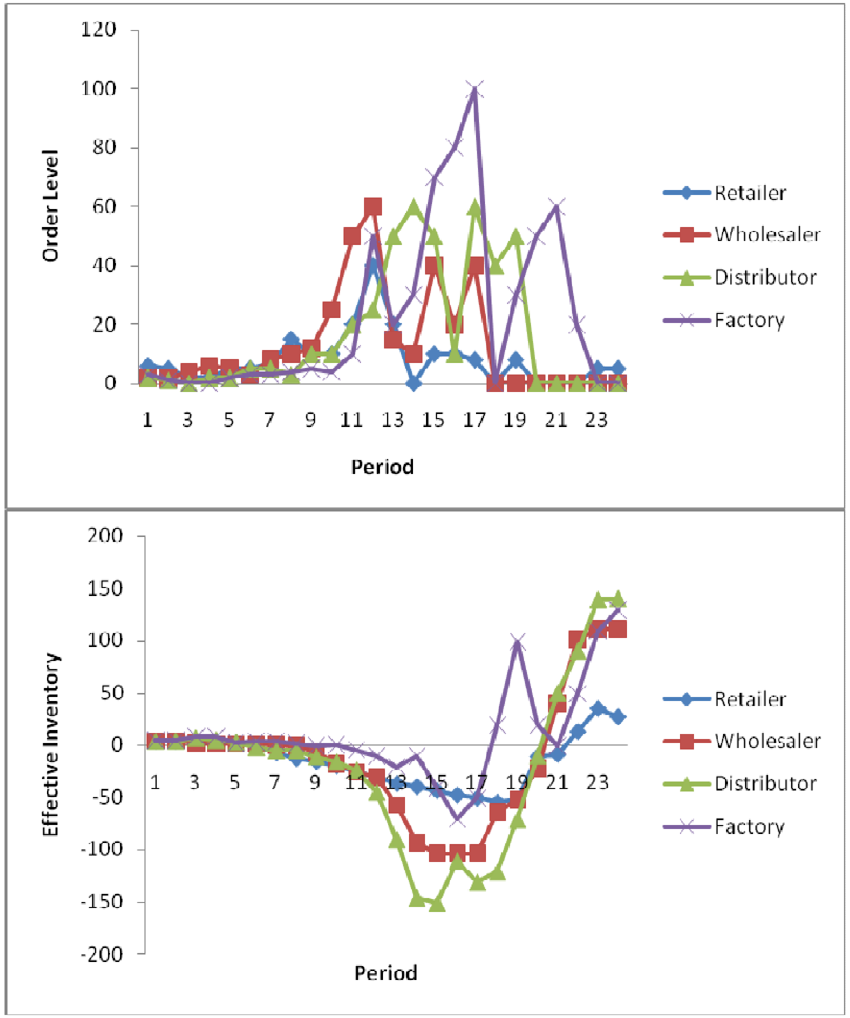


Order Level and Effective Inventory of Team 60

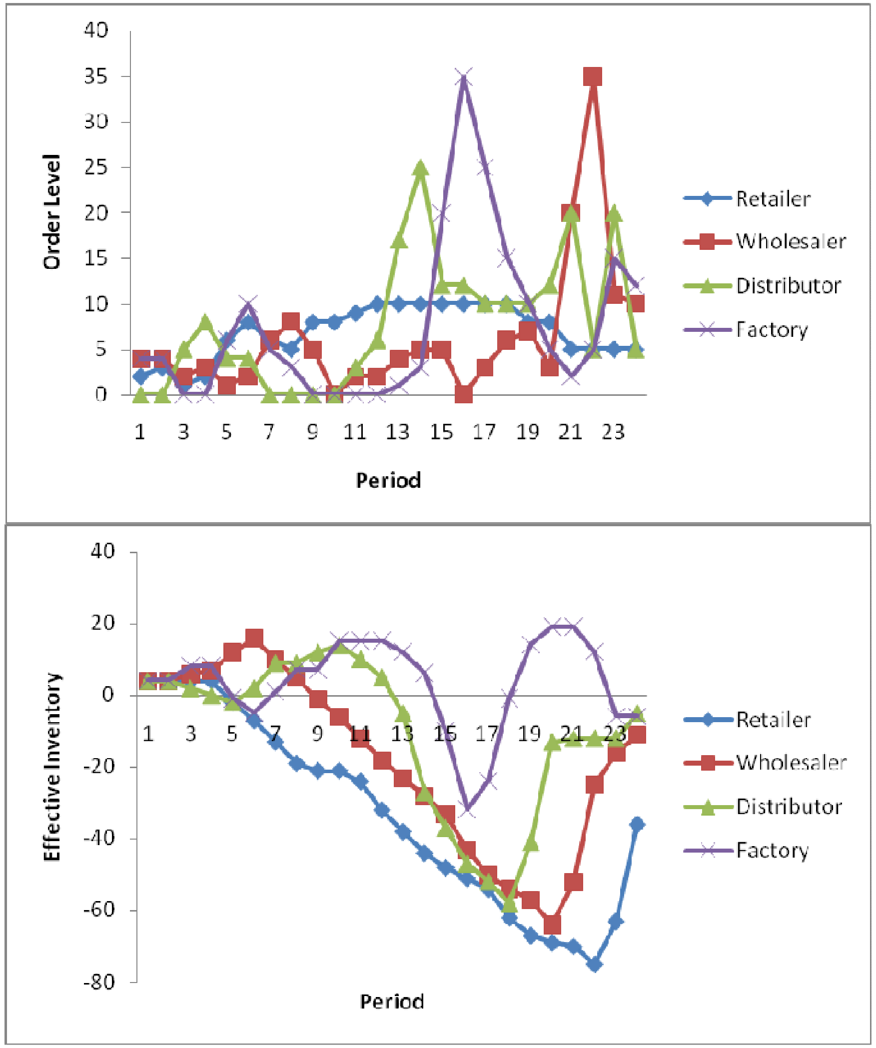
Appendix F: The Graphs of the Modified Experiments



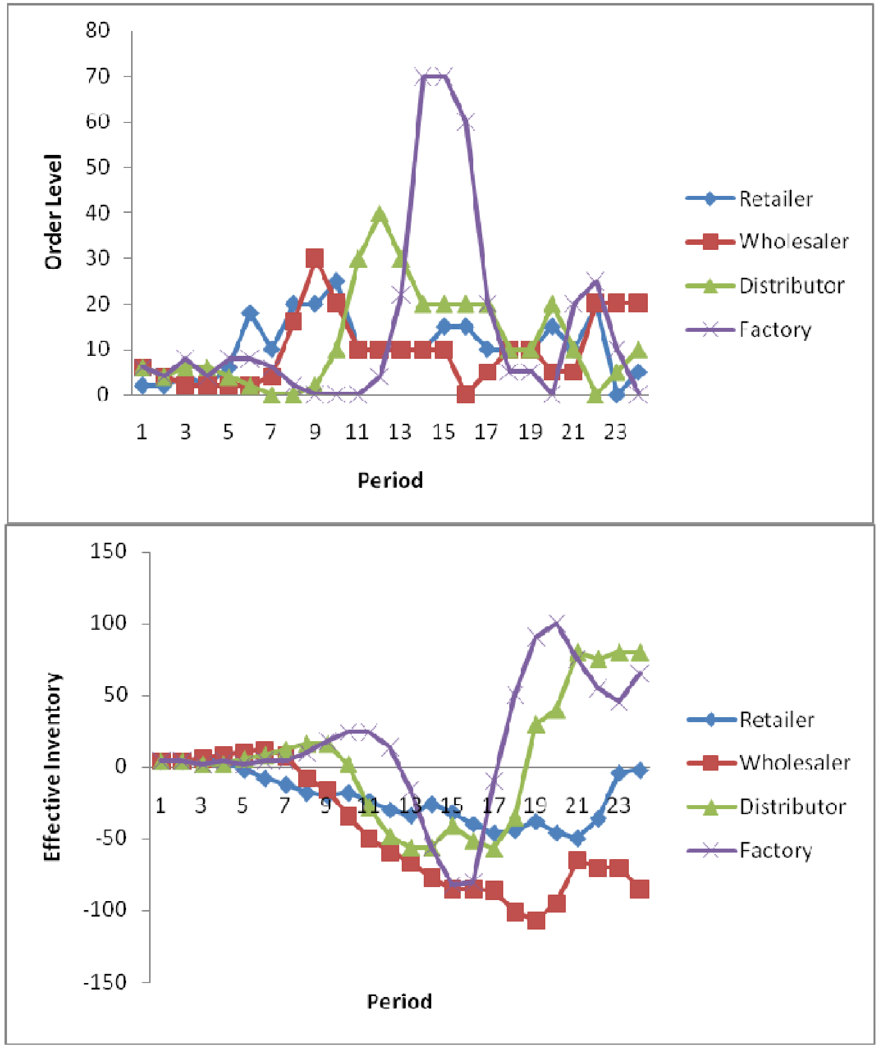
Order Level and Effective Inventory of Team 3



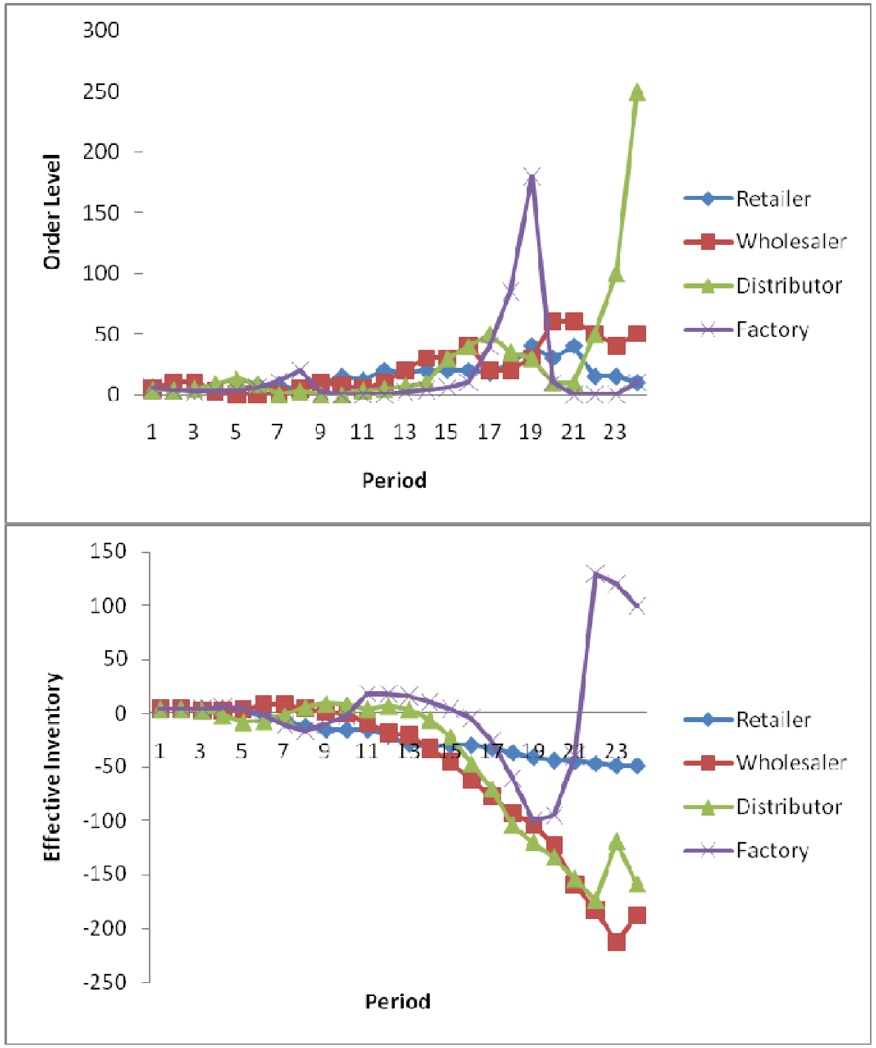
Order Level and Effective Inventory of Team 5



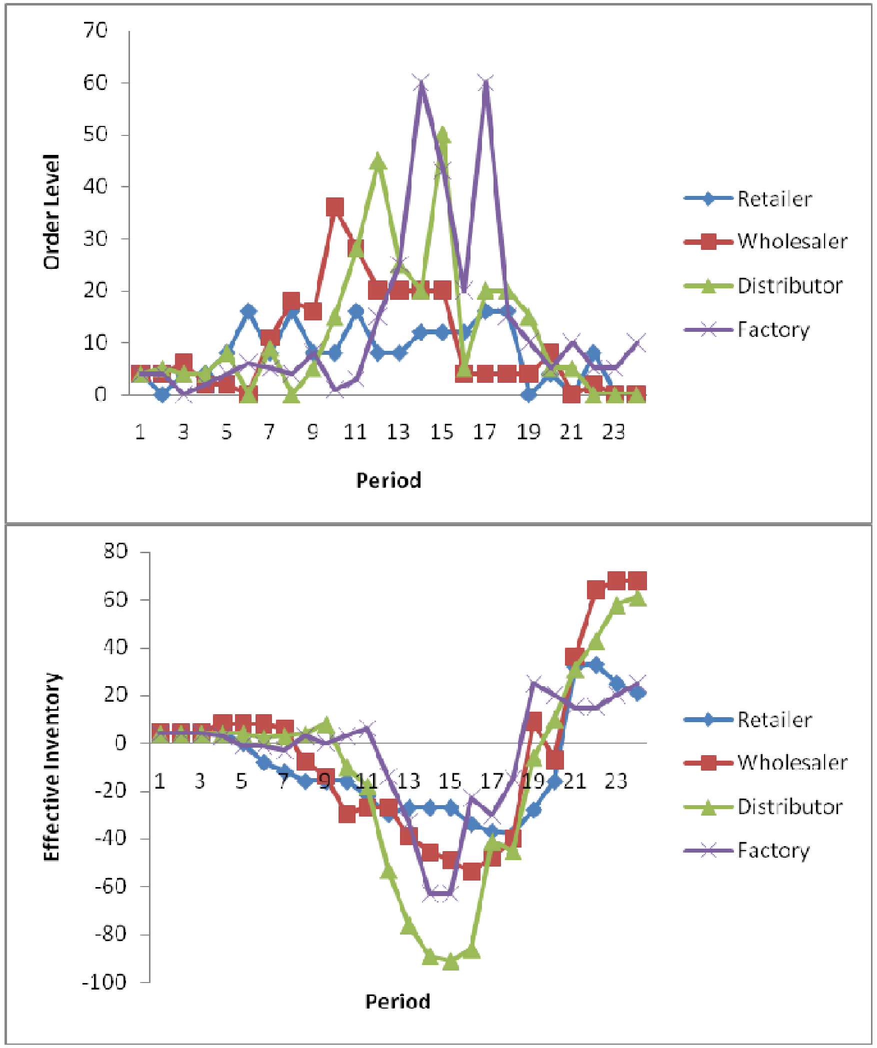
Order Level and Effective Inventory of Team 7



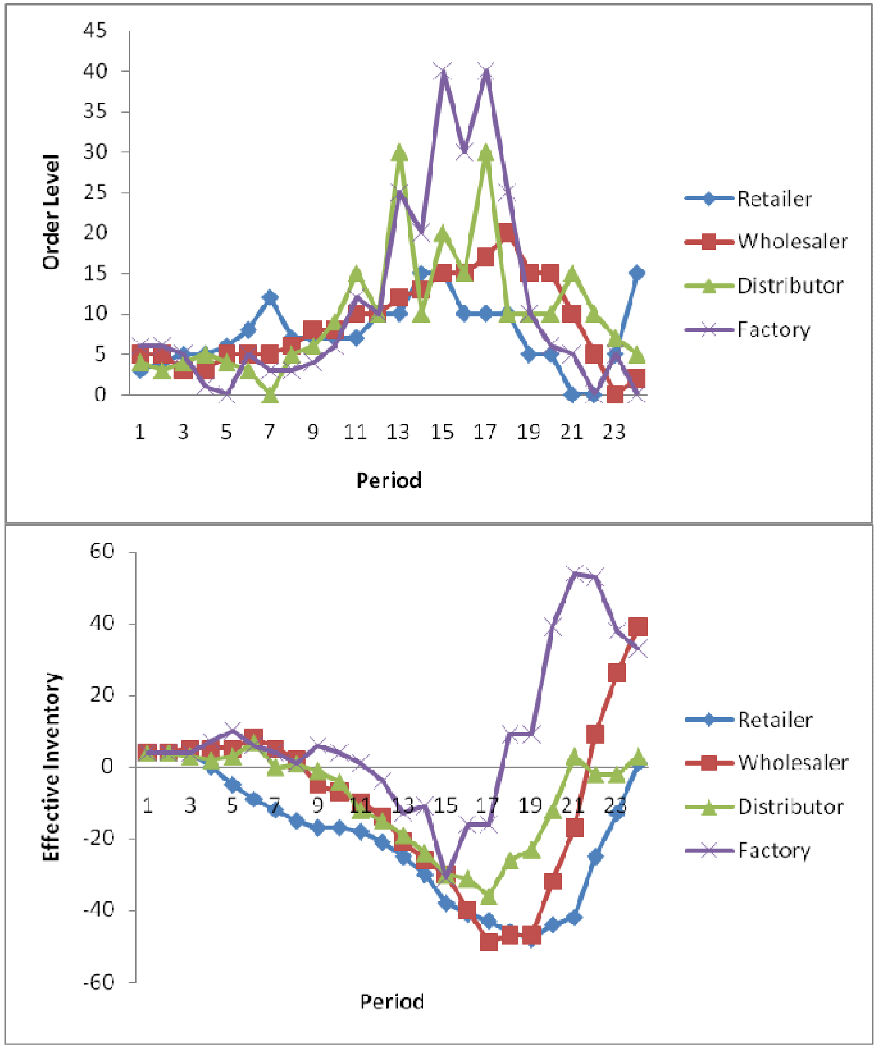
Order Level and Effective Inventory of Team 8



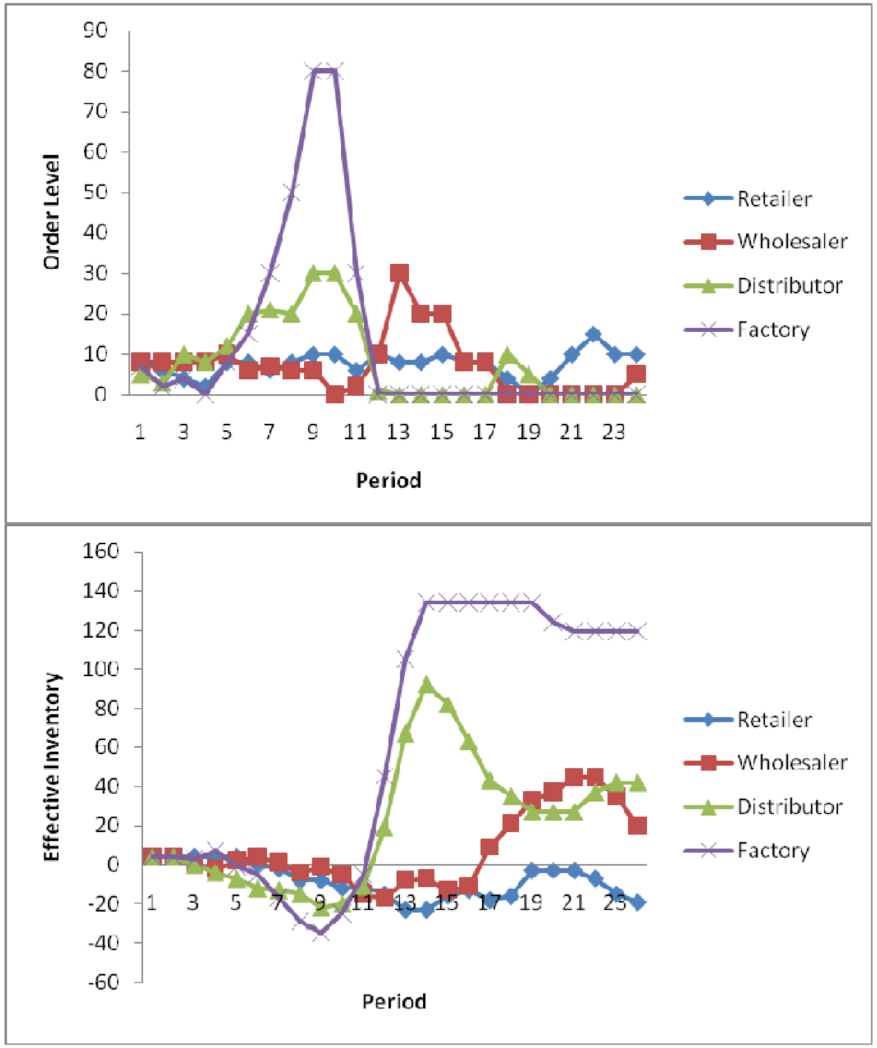
Order Level and Effective Inventory of Team 11



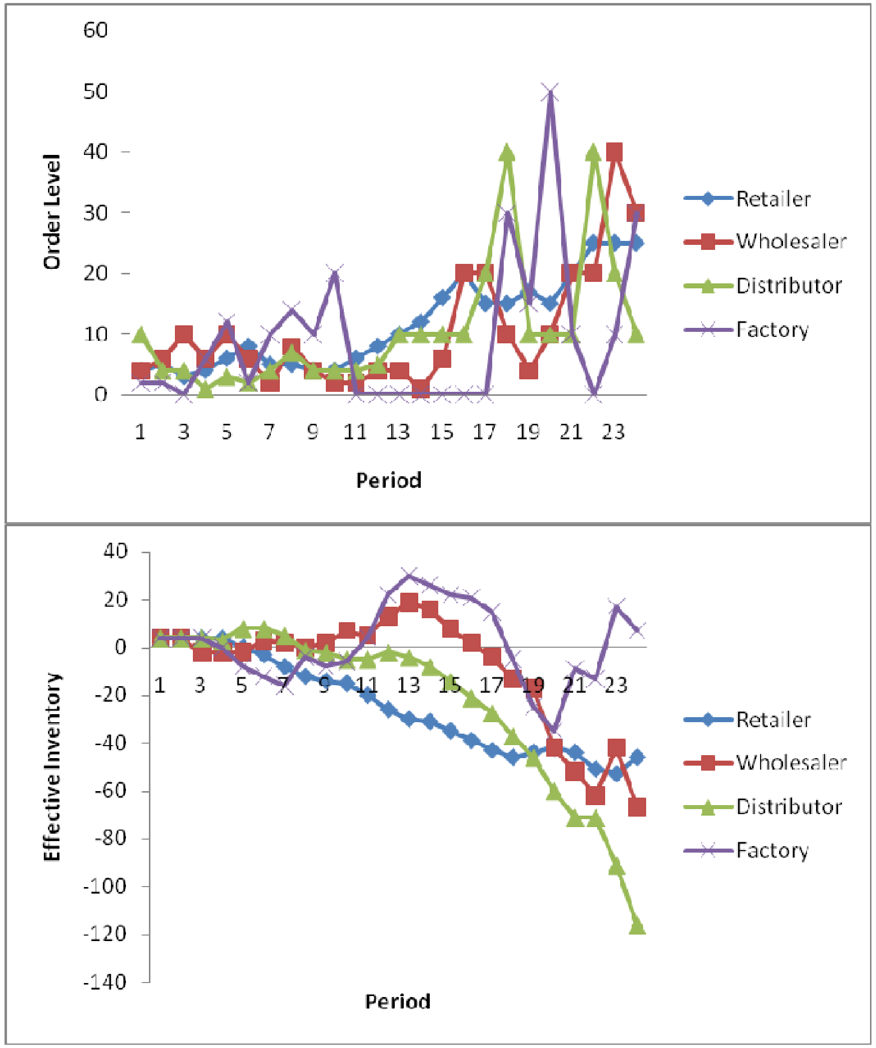
Order Level and Effective Inventory of Team 12



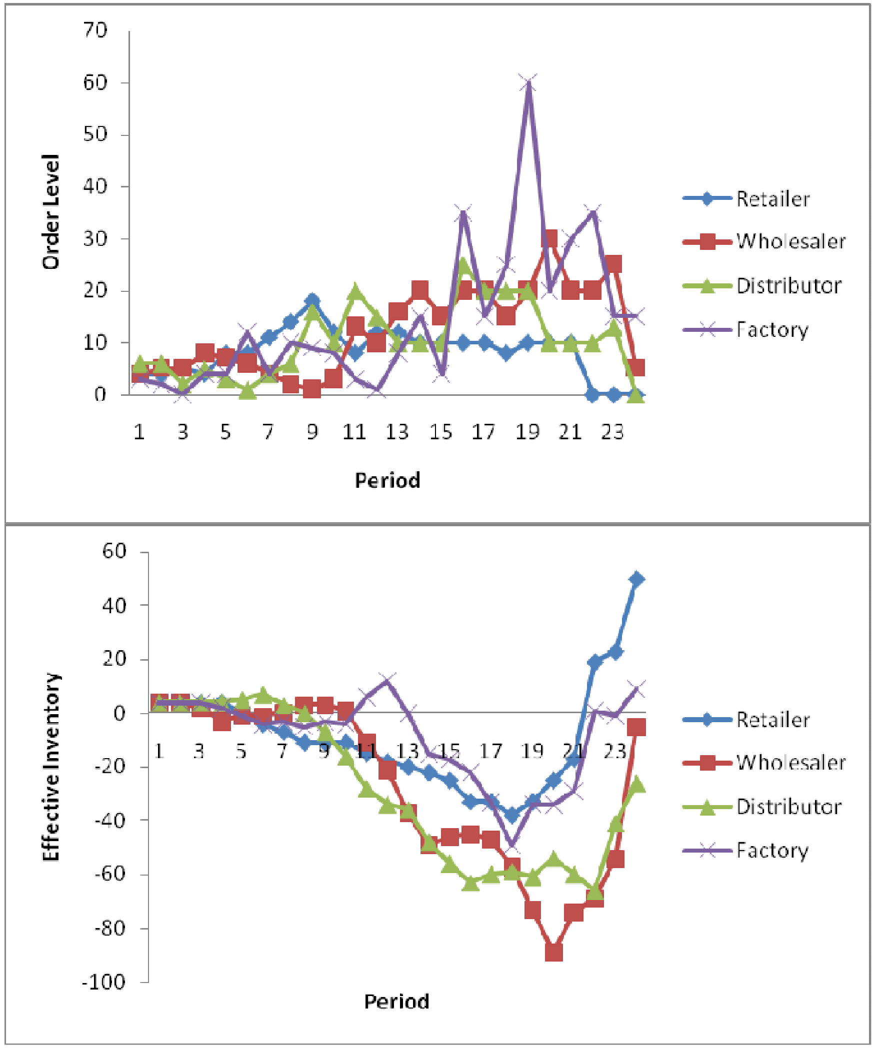
Order Level and Effective Inventory of Team 30



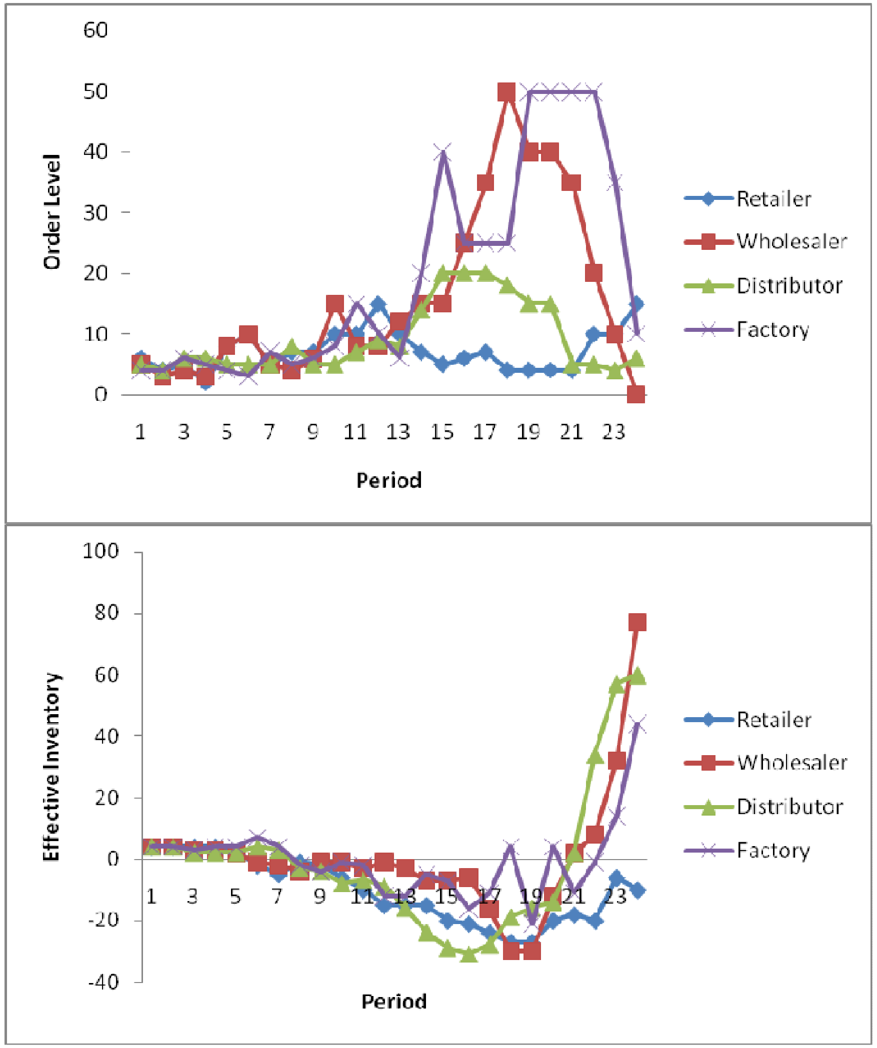
Order Level and Effective Inventory of Team 31



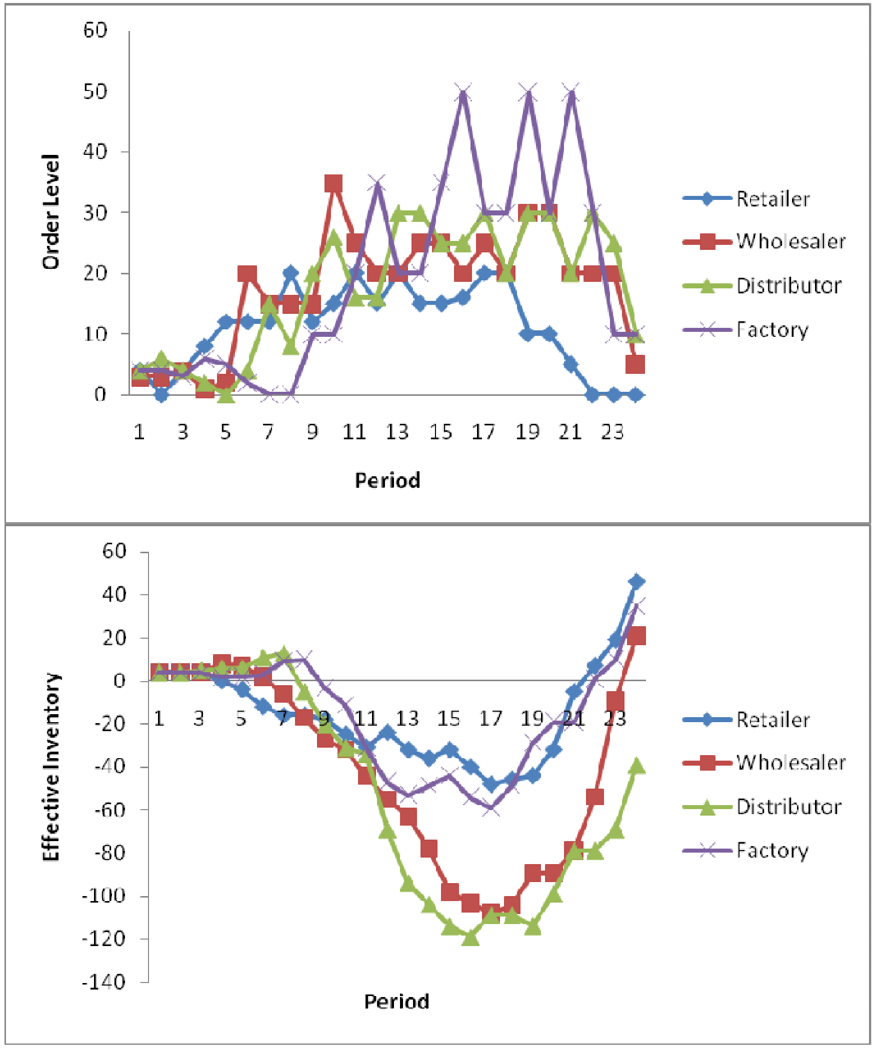
Order Level and Effective Inventory of Team 34



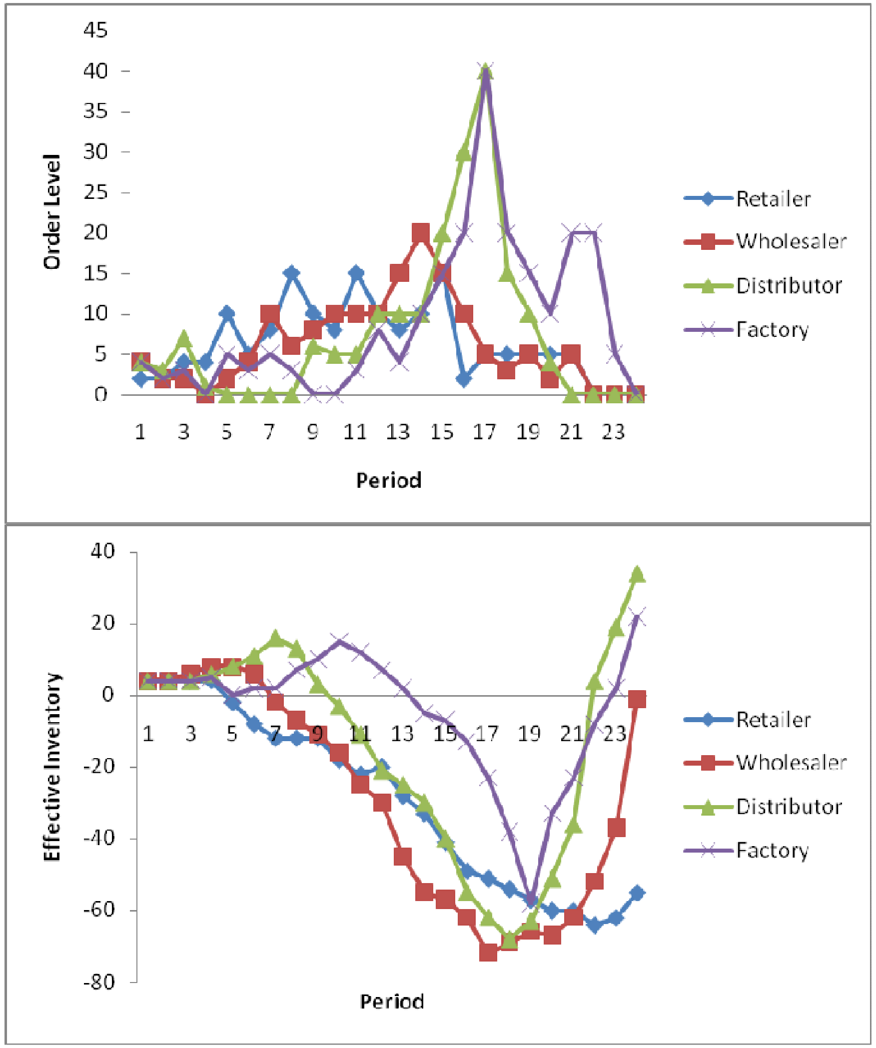
Order Level and Effective Inventory of Team 35



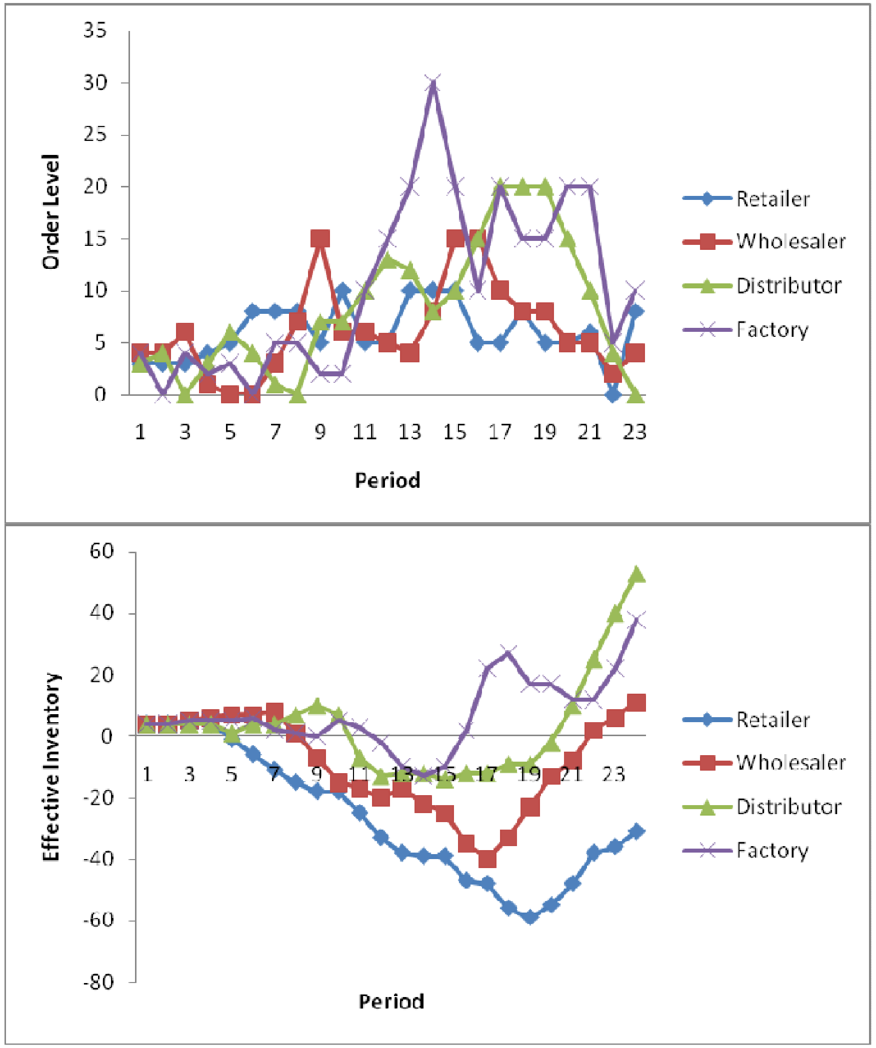
Order Level and Effective Inventory of Team 36



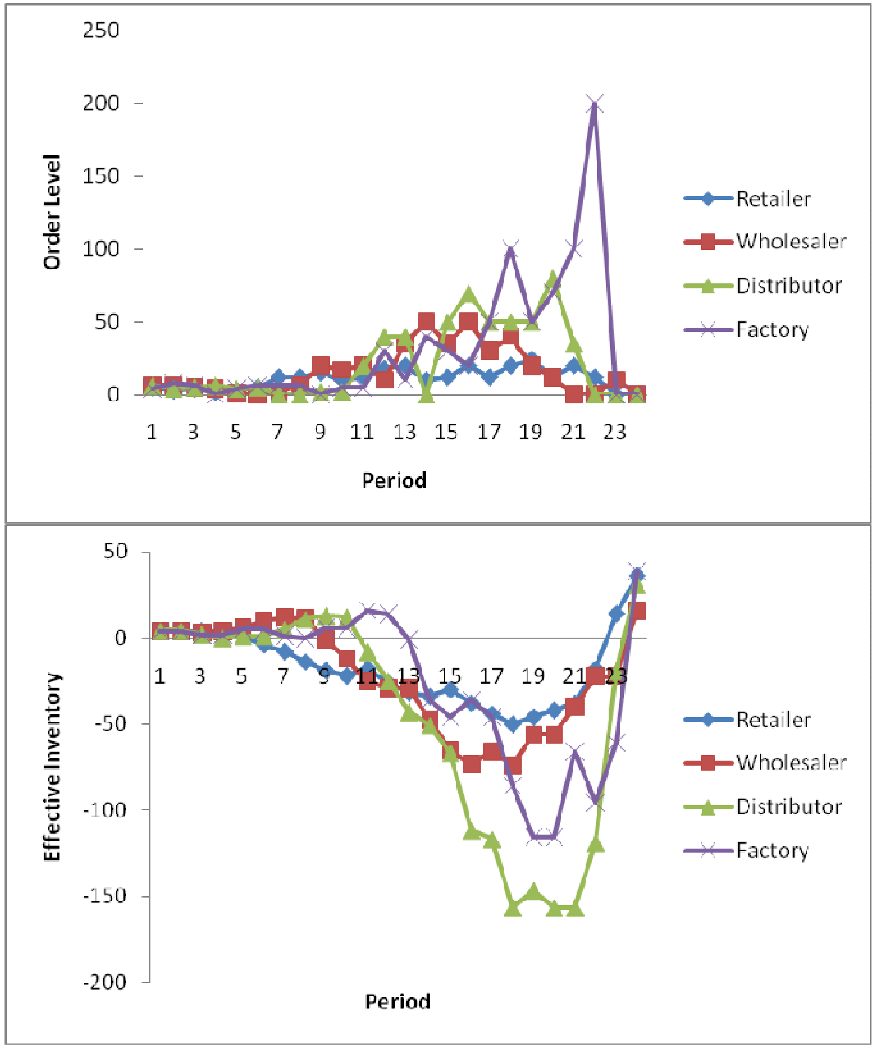
Order Level and Effective Inventory of Team 40



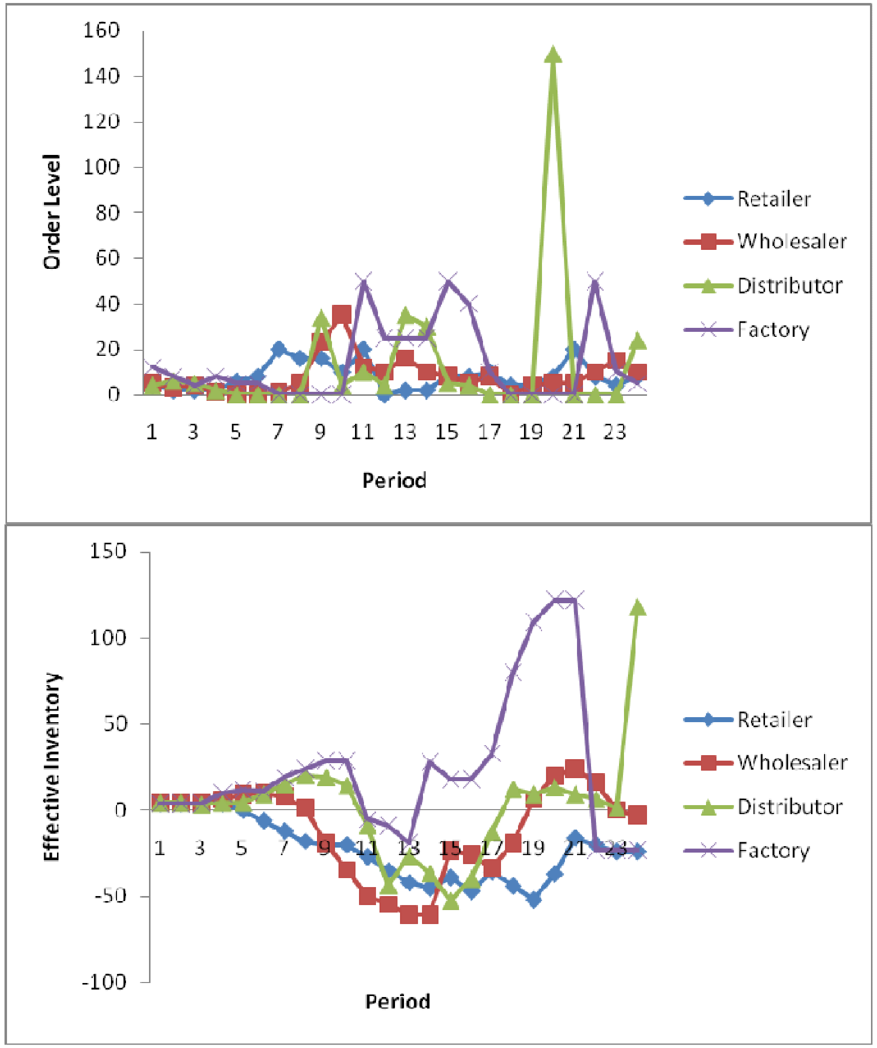
Order Level and Effective Inventory of Team 44



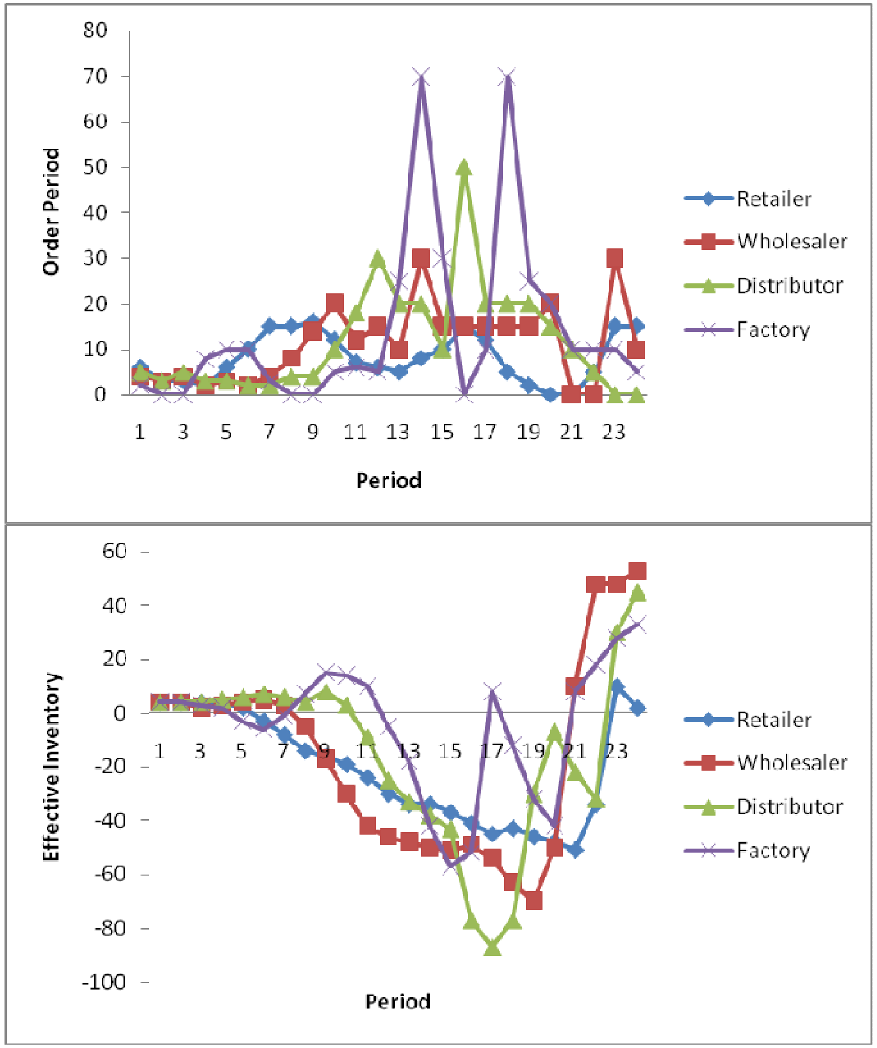
Order Level and Effective Inventory of Team 45



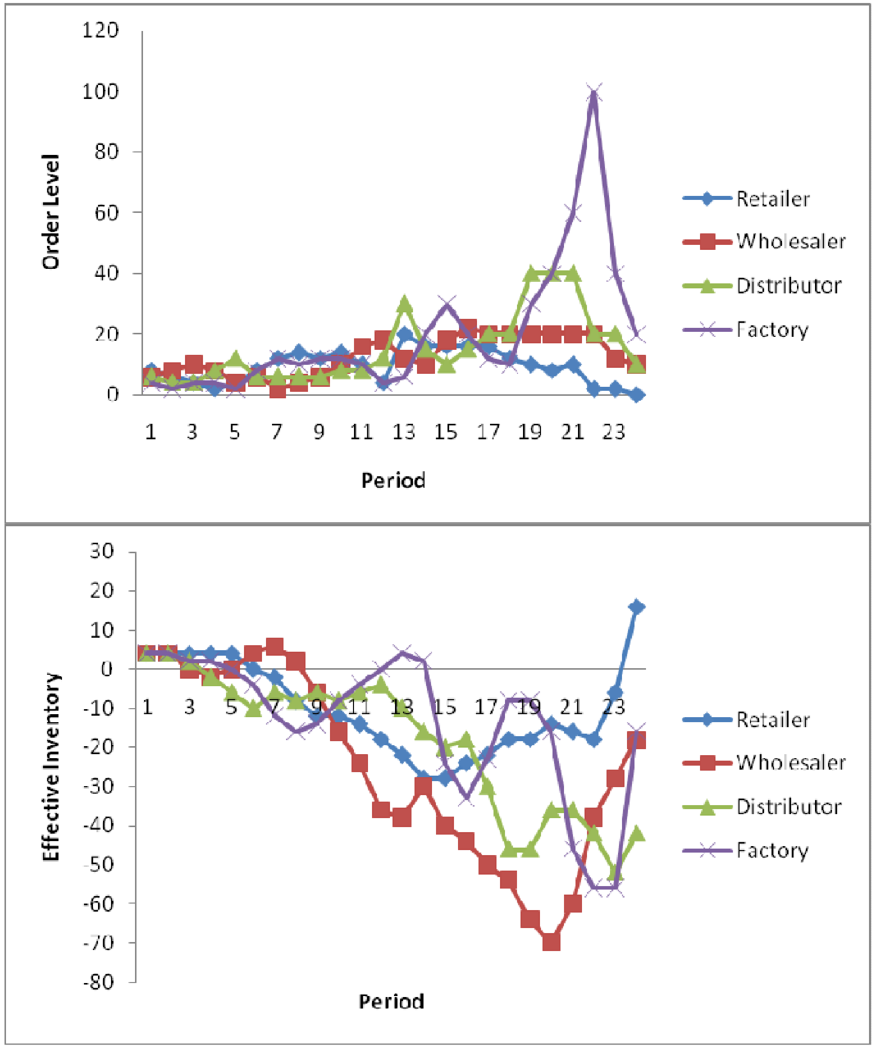
Order Level and Effective Inventory of Team 56



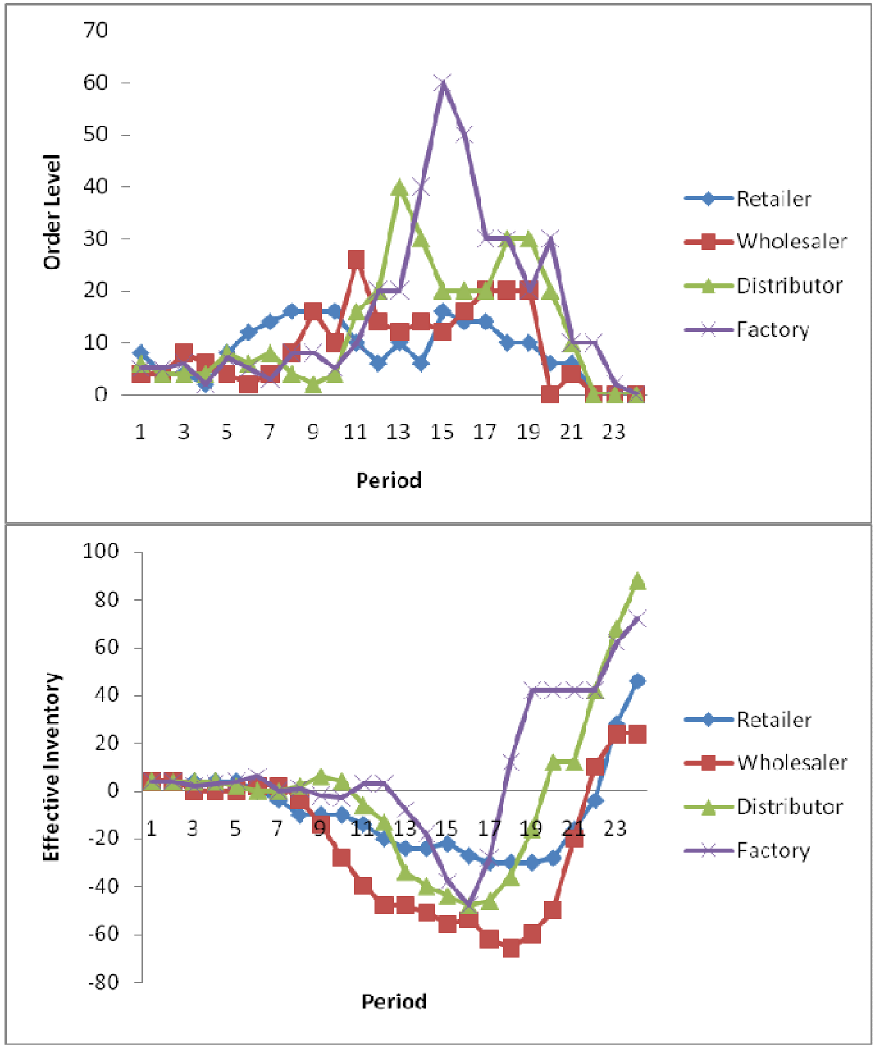
Order Level and Effective Inventory of Team 57



Order Level and Effective Inventory of Team 59



Order Level and Effective Inventory of Team 61



Order Level and Effective Inventory of Team 62

Appendix G: Box plots of Variance of Orders and Amplification Ratios

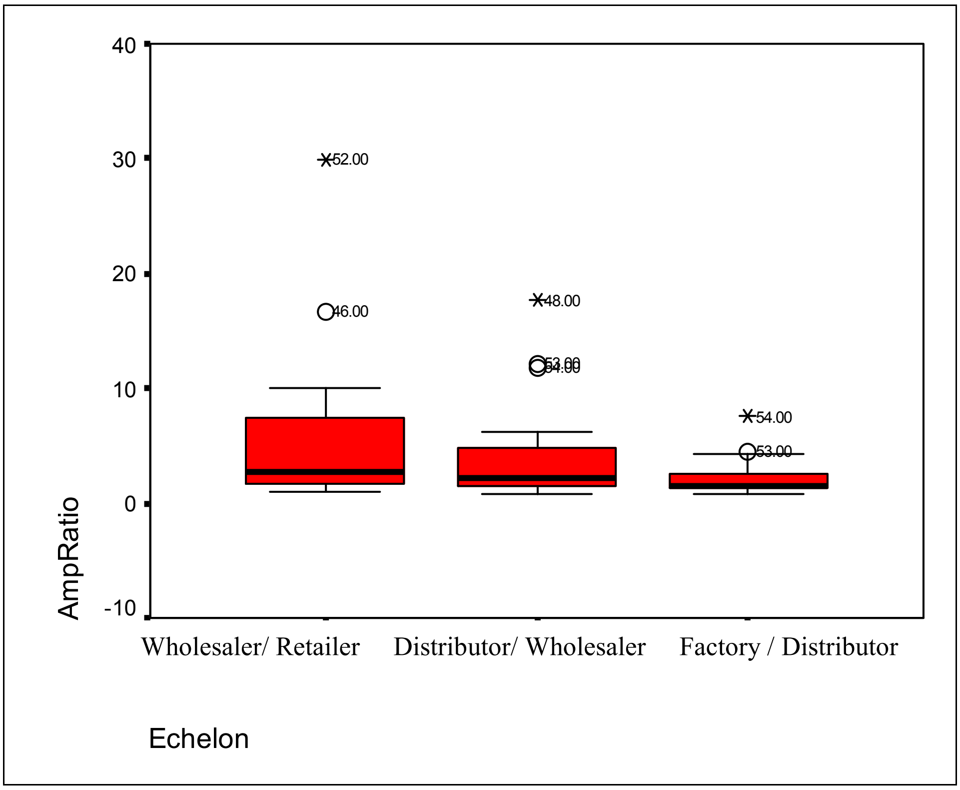


Figure 5-3: Box Plot of Amplification Ratios for Standard Experiments

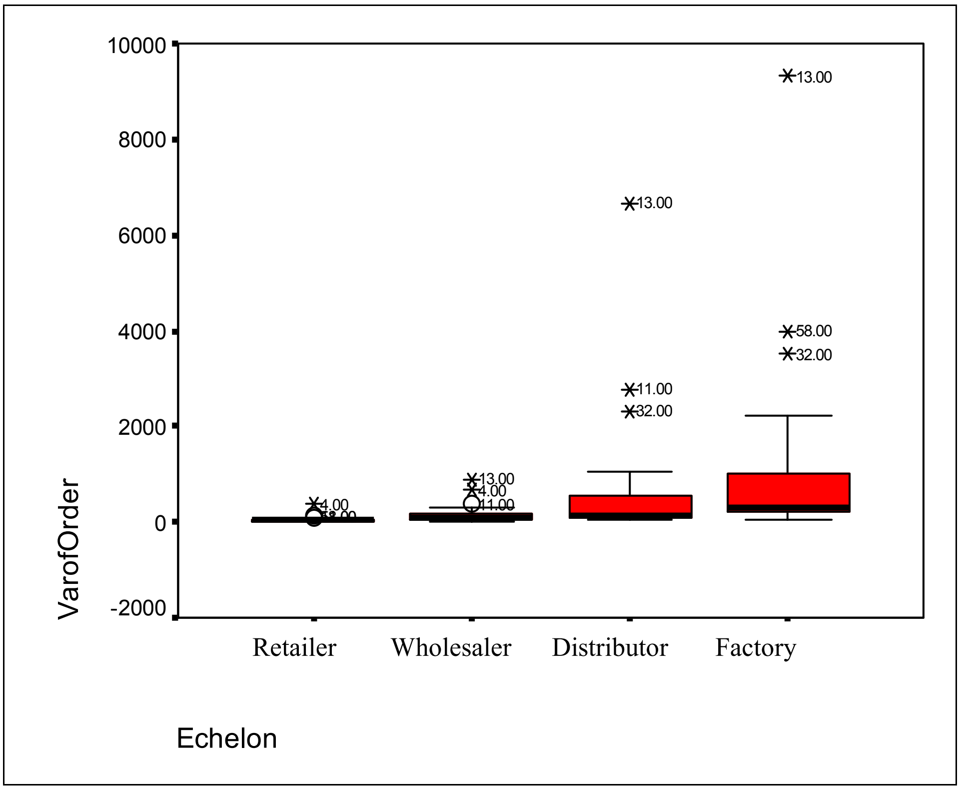


Figure 5-4: Box Plot of Order Variances for Modified Experiments

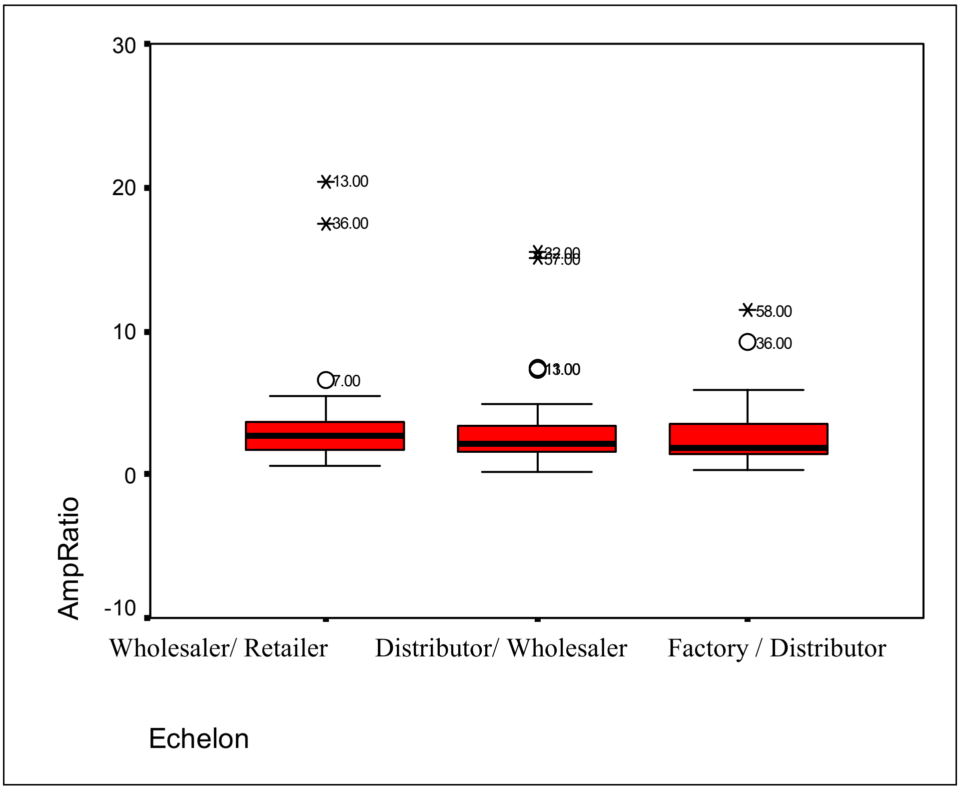


Figure 5-5: Box Plot of Amplification Ratios for Standard Experiments

Appendix H: Descriptive Comparison Detail Tables

Table 5-3: Order Variances

Standard Experiments					Modified Experiments				
Team	**R**	**W**	**D**	**F**	**Team**	**R**	**W**	**D**	**F**
14	65.3	399.1	817.4	2089.0	**3**	21.2	61.0	145.7	200.8
16	56.1	155.2	470.1	1027.5	**5**	79.1	304.2	472.3	883.2
29	59.8	83.5	107.4	171.1	**7**	8.5	55.9	51.7	80.9
37	82.3	112.0	390.2	587.2	**8**	44.3	58.8	116.9	455.0
38	6.5	56.5	347.4	523.0	**11**	113.3	380.6	2772.1	1537.4
39	7.4	40.9	94.4	131.4	**12**	32.7	101.2	189.1	292.5
43	8.7	21.7	59.5	62.6	**30**	17.4	28.9	60.1	147.9
46	56.0	935.2	1512.2	2324.5	**31**	10.1	55.0	101.1	589.4
49	88.6	120.1	95.3	101.9	**34**	56.1	96.4	106.0	158.3
50	21.5	42.9	94.8	235.2	**35**	20.4	68.2	47.5	203.6
53	22.4	47.2	572.9	2592.6	**36**	11.5	201.4	32.7	300.9
54	26.4	70.3	828.5	6272.2	**40**	49.0	94.9	112.1	273.3
55	31.8	93.5	92.2	388.7	**44**	21.1	29.1	102.3	92.8
60	352.7	349.4	738.3	644.8	**45**	7.6	18.3	43.1	67.9
					56	49.5	258.8	654.9	2202.3
					57	38.3	64.8	975.0	302.0
					59	27.8	73.0	136.9	369.1
					61	30.8	41.2	133.1	502.8
					62	28.8	57.6	132.1	261.8
Average over Teams	**63.3**	**180.5**	**444.3**	**1225.1**		**35.1**	**107.9**	**336.0**	**469.6**

Table 5-4: Amplification Ratios

Standard Experiments				Modified Experiments			
Team	**W/ R**	**D/ W**	**F/ D**	**Team**	**W/ R**	**D/ W**	**F/ D**
14	6.1	2.0	2.6	**3**	2.9	2.4	1.4
16	2.8	3.0	2.2	**5**	3.8	1.6	1.9
29	1.4	1.3	1.6	**7**	6.6	0.9	1.6
37	1.4	3.5	1.5	**8**	1.3	2.0	3.9
38	8.7	6.1	1.5	**11**	3.4	7.3	0.6
39	5.5	2.3	1.4	**12**	3.1	1.9	1.5
43	2.5	2.7	1.1	**30**	1.7	2.1	2.5
46	16.7	1.6	1.5	**31**	5.5	1.8	5.8
49	1.4	0.8	1.1	**34**	1.7	1.1	1.5
50	2.0	2.2	2.5	**35**	3.3	0.7	4.3
53	2.1	12.1	4.5	**36**	17.5	0.2	9.2
54	2.7	11.8	7.6	**40**	1.9	1.2	2.4
55	2.9	1.0	4.2	**44**	1.4	3.5	0.9
60	1.0	2.1	0.9	**45**	2.4	2.4	1.6
				56	5.2	2.5	3.4
				57	1.7	15.0	0.3
				59	2.6	1.9	2.7
				61	1.3	3.2	3.8
				62	2.0	2.3	2.0
Average over Teams	**4.1**	**3.8**	**2.4**		**3.7**	**2.8**	**2.7**

Table 5-5: The Period of Peak Order Levels

Standard Experiments					Modified Experiments				
Team	**R**	**W**	**D**	**F**	**Team**	**R**	**W**	**D**	**F**
14	13	16	16	18	**3**	13	12	15	16
16	15	3	5	7	**5**	12	12	14	17
29	9	15	17	21	**7**	12	22	14	16
37	9	11	15	11	**8**	10	9	12	14
38	10	16	18	19	**11**	19	20	24	19
39	21	16	19	21	**12**	6	10	15	14
43	9	18	7	9	**30**	14	18	13	15
46	16	15	17	19	**31**	22	13	9	9
49	19	21	23	23	**34**	22	23	18	20
50	13	14	18	18	**35**	9	20	16	19
53	20	12	19	21	**36**	12	18	15	19
54	9	17	17	20	**40**	8	10	13	16
55	10	11	14	16	**44**	8	14	17	17
60	17	19	19	19	**45**	10	9	17	14
					56	19	14	20	22
					57	7	10	20	11
					59	9	14	16	18
					61	13	16	19	22
					62	8	11	13	15
Average over Teams	13.6	14.6	16.0	17.3		12.3	14.5	15.8	16.5
Median over Teams	13.0	15.5	17.0	19.0		12.0	14.0	15.0	16.0

Table 5-6: Peak Order Magnitudes

Standard Experiments					Modified Experiments				
Team	**R**	**W**	**D**	**F**	**Team**	**R**	**W**	**D**	**F**
14	28	70	100	190	**3**	18	30	40	55
16	25	50	100	150	**5**	40	60	60	100
29	25	40	35	45	**7**	10	35	25	35
37	35	40	95	100	**8**	25	30	40	70
38	12	25	70	90	**11**	40	60	250	180
39	14	20	40	40	**12**	16	36	50	60
43	10	15	20	20	**30**	15	20	30	40
46	30	100	150	150	**31**	15	30	30	80
49	40	50	40	40	**34**	25	40	40	50
50	16	20	35	55	**35**	18	30	25	60
53	20	25	100	200	**36**	15	50	20	50
54	20	30	80	300	**40**	20	35	30	50
55	20	30	30	60	**44**	15	20	40	40
60	50	50	100	100	**45**	10	15	20	30
					56	24	50	80	200
					57	20	35	150	50
					59	16	30	50	70
					61	20	22	40	100
					62	16	26	40	60
Average over Teams	24.6	40.4	71.1	110.0		19.9	34.4	55.8	72.6
Median over Teams	22.5	35.0	75.0	95.0		18.0	30.0	40.0	60.0

Table 5-7: The Period of the Peak Backlog Level

Standard Experiments				
Team	**R**	**W**	**D**	**F**
14	18	18	20	18
16	13	13	22	8
29	17	15	23	23
37	18	14	14	17
38	14	12	18	20
39	21	17	20	21
43	7	12	13	11
46	20	18	17	19
49	24	24	24	24
50	16	16	20	19
53	22	21	19	17
54	21	19	19	21
55	18	20	15	18
60	22	24	19	21
Average over Teams	17.9	17.4	18.8	18.4
Median over Teams	18.0	17.5	19.0	19.0

Modified Experiments				
Team	**R**	**W**	**D**	**F**
3	19	21	20	17
5	18	15	15	16
7	22	20	18	16
8	21	19	17	15
11	23	23	22	19
12	18	16	15	14
30	19	17	17	15
31	13	12	9	9
34	23	24	24	20
35	18	20	22	18
36	18	18	16	19
40	17	17	16	17
44	22	17	18	19
45	19	17	15	14
56	18	18	18	19
57	19	13	15	22
59	21	19	17	15
61	14	20	23	22
62	17	18	16	16
Average over Teams	18.9	18.1	17.5	16.9
Median over Teams	19.0	18.0	17.0	17.0

Table 5-8: Peak Backlog Magnitudes

Standard Experiments				
Team	**R**	**W**	**D**	**F**
14	-42	-132	-189	-96
16	-20	-58	-85	-125
29	-45	-99	-90	-67
37	-54	-182	-101	-40
38	-23	-18	-39	-111
39	-45	-31	-42	-41
43	-4	-12	-6	-24
46	-60	-103	-115	-190
49	-84	-173	-154	-26
50	-25	-50	-50	-39
53	-37	-77	-91	-92
54	-45	-67	-91	-153
55	-23	-62	-101	-53
60	-46	-438	-174	-229
Average over Teams	-39.5	-107.3	-94.9	-91.9
Median over Teams	-43.5	-72.0	-91.0	-79.5

Modified Experiments				
Team	**R**	**W**	**D**	**F**
3	-57	-96	-82	-55
5	-55	-104	-151	-71
7	-75	-64	-58	-32
8	-50	-107	-57	-82
11	-49	-213	-174	-100
12	-38	-54	-91	-63
30	-48	-49	-36	-31
31	-23	-17	-22	-35
34	-53	-67	-116	-35
35	-38	-89	-66	-49
36	-27	-30	-31	-21
40	-48	-108	-119	-59
44	-64	-72	-68	-58
45	-59	-40	-14	-13
56	-50	-74	-157	-116
57	-52	-61	-53	-23
59	-51	-70	-87	-57
61	-28	-70	-52	-56
62	-30	-66	-48	-48
Average over Teams	-47.1	-76.4	-78.0	-52.8
Median over Teams	-50.0	-70.0	-66.0	-55.0

Table 5-9: First Backlog Periods

Standard Experiments				
Team	**R**	**W**	**D**	**F**
14	6	4	10	11
16	8	3	4	5
29	7	3	5	8
37	6	8	8	11
38	6	5	8	6
39	8	8	11	13
43	7	5	8	9
46	5	9	11	11
49	6	9	11	13
50	6	7	10	9
53	6	6	12	15
54	6	10	8	9
55	5	5	8	4
60	6	8	12	14
Average over Teams	6.3	6.4	9.0	9.9
Median over Teams	6.0	6.5	9.0	10.0

Modified Experiments				
Team	**R**	**W**	**D**	**F**
3	6	6	9	15
5	6	8	6	9
7	5	9	5	5
8	5	8	11	13
11	6	10	4	6
12	6	8	10	12
30	5	9	9	12
31	7	4	4	5
34	6	3	8	5
35	6	4	9	5
36	6	6	8	8
40	6	6	8	8
44	5	7	10	14
45	5	9	11	12
56	6	9	11	13
57	6	9	11	11
59	6	8	11	12
61	7	4	4	6
62	7	8	11	9
Average over Teams	5.9	7.1	8.4	9.5
Median over Teams	6.0	8.0	9.0	9.0

Table 5-10: Mean Orders

Standard Experiments					Modified Experiments				
Team	R	W	D	F	Team	R	W	D	F
14	12.8	25.3	27.3	33.6	3	9.2	11.9	11.4	12.2
16	13.3	18.3	24.1	23.3	5	8.0	13.0	16.9	22.7
29	10.2	13.6	15.9	14.7	7	6.8	6.2	7.8	7.5
37	11.3	10.2	10.3	11.9	8	10.8	9.7	11.9	14.9
38	6.7	9.8	18.4	20.2	11	15.3	21.5	28.3	16.9
39	6.8	8.9	10.3	10.9	12	7.8	9.7	12.2	13.5
43	6.2	7.5	9.1	9.2	30	7.5	8.8	10.0	11.1
46	9.9	23.0	30.4	33.6	31	7.5	7.1	8.1	12.8
49	11.8	11.8	9.4	8.5	34	11.5	10.4	10.5	9.3
50	8.2	11.3	11.9	14.3	35	8.3	12.3	10.5	14.0
53	10.9	12.4	19.8	31.2	36	7.3	15.7	9.2	19.3
54	9.7	13.1	28.5	46.0	40	11.0	17.4	17.8	19.3
55	9.2	10.6	13.8	19.1	44	6.4	6.2	7.5	9.0
60	29.2	24.0	24.8	21.4	45	6.2	6.0	8.0	10.3
					56	11.0	15.8	21.6	31.3
					57	8.0	8.1	13.2	13.8
					59	8.0	11.1	11.6	13.9
					61	9.4	12.6	15.3	19.7
					62	8.67	9.33	12.75	16.08
Average over Teams	**11.1**	**14.3**	**18.1**	**21.3**		**8.9**	**11.2**	**12.9**	**15.1**

Table 5-11: Mean Inventory Costs

Standard Experiments					Modified Experiments				
Team	**R**	**W**	**D**	**F**	**Team**	**R**	**W**	**D**	**F**
14	5.2	3.9	1.0	5.8	**3**	0.8	0.8	2.6	14.5
16	42.0	0.8	9.0	14.7	**5**	3.9	15.7	18.4	19.0
29	6.3	0.4	0.6	4.5	**7**	0.7	2.7	3.0	6.9
37	3.8	1.3	9.2	24.8	**8**	0.7	2.2	19.0	24.8
38	0.8	9.8	12.6	4.5	**11**	0.8	1.5	2.0	18.3
39	1.0	0.8	0.8	1.9	**12**	5.3	12.0	10.0	6.1
43	9.3	12.5	7.0	16.5	**30**	0.5	4.7	1.3	11.9
46	4.3	14.0	14.5	7.8	**31**	0.8	10.8	25.5	65.5
49	0.7	0.7	2.2	3.3	**34**	0.7	3.5	1.5	7.3
50	1.0	1.8	2.5	8.0	**35**	4.5	0.7	1.3	1.8
53	0.8	0.6	16.6	16.6	**36**	0.8	5.6	7.3	4.0
54	1.0	4.7	3.9	15.8	**40**	3.5	2.1	2.0	3.5
55	1.5	0.8	12.0	10.9	**44**	0.7	1.5	5.3	4.1
60	0.7	1.5	2.5	1.1	**45**	0.7	2.5	7.4	8.7
					56	2.8	2.9	3.5	4.3
					57	0.7	4.7	11.0	28.2
					59	1.3	7.7	5.3	6.4
					61	1.5	0.8	0.4	0.8
					62	3.9	2.9	10.5	14.3
Average over Teams	**5.6**	**3.8**	**6.7**	**9.7**		**1.8**	**4.5**	**7.2**	**13.2**

Table 5-12: Mean Backlog Costs

Standard Experiments				
Team	**R**	**W**	**D**	**F**
14	28.1	112.4	105.1	24.5
16	8.2	23.3	63.8	75.7
29	27.0	80.3	70.0	38.8
37	36.3	97.2	40.3	6.8
38	19.2	8.1	25.8	33.9
39	33.4	23.9	25.3	16.8
43	1.3	4.9	1.3	4.0
46	47.3	52.5	42.8	64.5
49	74.8	99.3	77.9	16.8
50	21.3	41.2	24.8	12.1
53	31.3	52.5	44.3	28.7
54	39.9	47.5	45.9	53.3
55	24.0	53.8	54.5	19.0
60	46.1	264.8	98.8	110.7
Average over Teams	**31.3**	**68.7**	**51.5**	**36.1**

Modified Experiments				
Team	**R**	**W**	**D**	**F**
3	52.5	75.6	74.2	9.5
5	38.5	56.8	78.3	17.8
7	68.0	41.1	26.9	7.1
8	44.2	96.7	31.0	20.3
11	46.3	111.0	94.0	30.4
12	29.5	32.4	42.9	20.5
30	42.4	28.8	19.8	7.6
31	17.9	6.9	8.7	9.8
34	50.2	25.4	48.4	11.8
35	26.9	56.8	59.6	21.2
36	22.0	10.3	17.3	8.8
40	38.4	87.9	107.3	38.9
44	60.0	61.3	38.8	17.3
45	55.1	22.9	8.5	2.9
56	40.3	51.4	98.3	58.8
57	47.0	32.3	18.6	8.5
59	43.9	47.9	40.0	22.5
61	23.3	51.5	37.5	28.7
62	25.3	50.1	23.6	12.1
	40.6	**49.9**	**46.0**	**18.6**

Table 5-13: Mean Total Costs

Standard Experiments					Modified Experiments				
Team	**R**	**W**	**D**	**F**	**Team**	**R**	**W**	**D**	**F**
14	33.3	116.3	106.1	30.3	**3**	53.3	76.4	76.8	24.0
16	50.2	24.0	72.8	90.4	**5**	42.4	72.5	96.7	36.8
29	33.3	80.8	70.6	43.3	**7**	68.7	43.8	29.9	14.0
37	40.1	98.5	49.5	31.6	**8**	44.8	98.8	50.0	45.1
38	19.9	17.8	38.4	38.4	**11**	47.0	112.5	96.0	48.8
39	34.4	24.8	26.1	18.8	**12**	34.8	44.4	52.9	26.6
43	10.5	17.4	8.4	20.5	**30**	43.0	33.4	21.0	19.5
46	51.6	66.5	57.3	72.3	**31**	18.8	17.8	34.1	75.3
49	75.5	100.0	80.1	20.0	**34**	50.8	29.0	49.9	19.1
50	22.3	42.9	27.3	20.1	**35**	31.4	57.5	60.9	22.9
53	32.0	53.1	60.8	45.3	**36**	22.8	16.0	24.6	12.8
54	41.0	52.2	49.8	69.1	**40**	41.9	90.0	109.3	42.4
55	25.5	54.5	66.5	29.9	**44**	60.7	62.8	44.0	21.4
60	46.8	266.3	101.2	111.8	**45**	55.8	25.5	15.9	11.6
					56	43.0	54.3	101.8	63.2
					57	47.7	36.9	29.6	36.7
					59	45.2	55.6	45.3	28.9
					61	24.8	52.3	37.9	29.4
					62	29.2	53.0	34.1	26.4
Average over Teams	**36.9**	**72.5**	**58.2**	**45.8**		**42.4**	**54.3**	**53.2**	**31.8**

Table 5-14: Inventory Cost Variances

Standard Experiments					Modified Experiments				
Team	**R**	**W**	**D**	**F**	**Team**	**R**	**W**	**D**	**F**
14	308.0	271.7	2.3	248.8	**3**	2.5	3.0	20.2	164.5
16	3328.9	3.0	347.2	482.7	**5**	79.6	1330.1	1804.2	1431.6
29	128.0	1.4	2.1	59.3	**7**	2.3	20.5	19.9	44.7
37	66.8	5.4	195.7	796.5	**8**	2.3	14.2	844.5	985.2
38	2.3	333.8	1626.1	77.4	**11**	2.4	6.1	8.8	1494.9
39	3.5	2.1	2.3	6.8	**12**	111.8	502.8	337.7	74.4
43	105.3	401.1	89.2	417.5	**30**	1.8	85.0	3.7	297.2
46	174.7	1285.8	1755.0	376.7	**31**	2.8	256.1	810.5	3910.0
49	2.3	1.7	8.8	15.2	**34**	2.3	29.4	6.8	100.9
50	4.5	26.5	28.0	320.3	**35**	129.0	1.9	4.7	10.5
53	2.4	1.5	2044.0	2099.3	**36**	2.4	274.6	296.4	83.4
54	5.2	112.2	195.5	2601.5	**40**	99.1	21.8	13.9	55.6
55	16.2	2.6	624.8	819.9	**44**	2.3	7.7	68.5	32.2
60	2.3	6.2	8.8	2.3	**45**	2.3	11.4	179.3	103.2
					56	59.6	22.2	50.6	73.3
					57	2.3	46.8	559.5	1502.7
					59	5.8	269.6	110.7	84.9
					61	12.3	3.1	1.4	2.0
					62	113.6	47.3	517.3	496.2
Average over Teams	**296.5**	**175.4**	**495.0**	**594.6**		**33.5**	**155.5**	**297.8**	**576.2**

Table 5-15: Backlog Cost Variances

Standard Experiments					Modified Experiments				
Team	**R**	**W**	**D**	**F**	**Team**	**R**	**W**	**D**	**F**
14	850.6	11047.3	17409.4	2463.0	**3**	1811.6	4856.7	4657.4	699.9
16	181.5	1099.1	3757.5	9518.8	**5**	1646.0	5956.1	11786.9	1416.1
29	1003.8	4851.7	3477.6	2181.9	**7**	2679.3	1929.7	1454.6	260.5
37	1253.4	11273.7	4320.8	329.0	**8**	1202.8	6187.7	2147.8	2458.3
38	277.5	143.1	875.4	3522.3	**11**	1275.6	19855.3	15885.2	3451.8
39	1221.7	640.2	1031.9	713.0	**12**	753.8	1603.5	4399.6	1455.0
43	6.2	48.7	7.5	133.6	**30**	1119.3	1245.2	608.1	249.4
46	1603.4	4125.3	5272.0	9484.6	**31**	247.1	117.6	201.3	427.9
49	3833.0	14860.8	9384.2	344.3	**34**	1451.8	1943.5	4581.6	326.4
50	364.5	1512.7	1095.3	641.2	**35**	654.6	3650.4	2726.9	874.8
53	668.5	3095.4	4141.7	2912.2	**36**	361.0	301.8	460.4	150.9
54	1026.6	2633.5	3713.2	7771.4	**40**	1156.9	6693.7	8768.6	2038.3
55	337.7	1975.1	5434.5	1043.8	**44**	2259.8	3171.4	2442.7	949.4
60	1068.0	91774.6	15309.8	27520.2	**45**	1597.6	675.8	125.7	56.0
					56	1218.9	2960.9	15307.9	6549.2
					57	1154.1	1928.8	1151.5	281.8
					59	1426.6	2679.8	3021.2	1392.4
					61	379.4	2208.4	1215.7	1214.5
					62	557.8	2696.9	1309.9	705.6
Average over Teams	**978.3**	**10648.6**	**5373.6**	**4898.5**		**1208.1**	**3719.1**	**4329.1**	**1313.6**

Table 5-16: Total Cost Variances

Standard Experiments					Modified Experiments				
Team	**R**	**W**	**D**	**F**	**Team**	**R**	**W**	**D**	**F**
14	853.3	10400.1	17183.2	2413.6	**3**	1727.3	4728.3	4277.7	576.1
16	2794.6	1065.7	2911.3	7678.9	**5**	1414.2	5428.0	10583.5	2142.3
29	779.7	4783.2	3388.3	1879.9	**7**	2587.0	1721.6	1308.3	203.0
37	1035.9	11008.7	3743.0	772.5	**8**	1143.6	5764.8	1760.4	2393.2
38	249.8	312.5	1823.1	3281.1	**11**	1205.6	19513.9	15493.5	3783.0
39	1155.5	600.6	992.4	652.5	**12**	539.8	1297.3	3841.7	1267.4
43	87.3	322.0	77.1	413.3	**30**	1073.2	1050.2	560.3	358.0
46	1354.2	3872.6	5734.5	8806.8	**31**	218.7	217.3	551.3	3005.2
49	3731.2	14715.8	9033.9	244.4	**34**	1384.3	1785.0	4436.8	247.5
50	324.6	1388.9	993.7	759.8	**35**	530.9	3568.3	2571.0	807.9
53	621.9	3032.9	4654.2	4019.4	**36**	329.0	455.1	494.5	161.2
54	945.0	2283.1	3533.4	8613.4	**40**	975.4	6333.3	8325.5	1809.6
55	276.7	1893.6	4694.4	1432.5	**44**	2178.7	2987.1	2086.6	833.9
60	1006.2	90974.7	14812.0	27262.7	**45**	1523.2	565.7	174.1	106.2
					56	1044.0	2670.1	14640.9	6090.4
					57	1091.0	1661.5	1284.3	1284.8
					59	1317.9	2182.8	2693.7	1176.0
					61	318.6	2122.0	1184.5	1171.6
					62	465.0	2439.3	1310.4	840.3
Average over Teams	**1086.9**	**10475.3**	**5255.3**	**4873.6**		**1108.8**	**3499.5**	**4083.1**	**1487.2**

Appendix I: Hypothesis Test Results

Table 5-17: Results for the Supply Chain (R, W, D, F)

- *Order level comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgOrderLev	.00	56	77.34	4331.00
	1.00	76	58.51	4447.00
	Total	132		
VarOrderLev	.00	56	73.04	4090.00
	1.00	76	61.68	4688.00
	Total	132		

Test Statistics[a]

	AvgOrderLev	VarOrderLev
Mann-Whitney U	1521.000	1762.000
Wilcoxon W	4447.000	4688.000
Z	-2.795	-1.685
Asymp. Sig. (2-tailed)	.005	.092
Exact Sig. (2-tailed)	.005	.092
Exact Sig. (1-tailed)	.002	.046
Point Probability	.000	.000

a. Grouping Variable: GameType

- *Total cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgTotalCost	.00	56	69.56	3895.50
	1.00	76	64.24	4882.50
	Total	132		
VarTotalCost	.00	56	71.63	4011.00
	1.00	76	62.72	4767.00
	Total	132		

Test Statistics[a]

	AvgTotalCost	VarTotalCost
Mann-Whitney U	1956.500	1841.000
Wilcoxon W	4882.500	4767.000
Z	-.790	-1.321
Asymp. Sig. (2-tailed)	.430	.186
Exact Sig. (2-tailed)	.432	.188
Exact Sig. (1-tailed)	.216	.094
Point Probability	.001	.001

a. Grouping Variable: GameType

- *Inventory cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgInventoryCost	.00	56	66.49	3723.50
	1.00	76	66.51	5054.50
	Total	132		
VarInventoryCost	.00	56	67.74	3793.50
	1.00	76	65.59	4984.50
	Total	132		

Test Statistics[a]

	AvgInventory Cost	VarInventory Cost
Mann-Whitney U	2127.500	2058.500
Wilcoxon W	3723.500	4984.500
Z	-.002	-.320
Asymp. Sig. (2-tailed)	.998	.749
Exact Sig. (2-tailed)	.999	.751
Exact Sig. (1-tailed)	.500	.375
Point Probability	.001	.001

a. Grouping Variable: GameType

- *Backlog cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgBacklogCost	.00	56	69.72	3904.50
	1.00	76	64.13	4873.50
	Total	132		
VarBacklogCost	.00	56	70.07	3924.00
	1.00	76	63.87	4854.00
	Total	132		

Test Statistics[a]

	AvgBacklog Cost	VarBacklog Cost
Mann-Whitney U	1947.500	1928.000
Wilcoxon W	4873.500	4854.000
Z	-.831	-.921
Asymp. Sig. (2-tailed)	.406	.357
Exact Sig. (2-tailed)	.408	.360
Exact Sig. (1-tailed)	.204	.180
Point Probability	.001	.001

a. Grouping Variable: GameType

Table 5-18: Results for Downstream Echelons (R, W)

- *Order level comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgOrderLev	.00	28	39.46	1105.00
	1.00	38	29.11	1106.00
	Total	66		
VarOrderLev	.00	28	36.71	1028.00
	1.00	38	31.13	1183.00
	Total	66		

Test Statistics[a]

	AvgOrderLev	VarOrderLev
Mann-Whitney U	365.000	442.000
Wilcoxon W	1106.000	1183.000
Z	-2.167	-1.168
Asymp. Sig. (2-tailed)	.030	.243
Exact Sig. (2-tailed)	.030	.247
Exact Sig. (1-tailed)	.015	.124
Point Probability	.000	.003

a. Grouping Variable: GameType

- *Total cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgTotalCost	.00	28	31.93	894.00
	1.00	38	34.66	1317.00
	Total	66		
VarTotalCost	.00	28	32.57	912.00
	1.00	38	34.18	1299.00
	Total	66		

Test Statistics[a]

	AvgTotalCost	VarTotalCost
Mann-Whitney U	488.000	506.000
Wilcoxon W	894.000	912.000
Z	-.571	-.337
Asymp. Sig. (2-tailed)	.568	.736
Exact Sig. (2-tailed)	.575	.742
Exact Sig. (1-tailed)	.287	.371
Point Probability	.004	.005

a. Grouping Variable: GameType

- *Inventory cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgInventoryCost	.00	28	33.66	942.50
	1.00	38	33.38	1268.50
	Total	66		
VarInventoryCost	.00	28	33.50	938.00
	1.00	38	33.50	1273.00
	Total	66		

Test Statistics[a]

	AvgInventory Cost	VarInventory Cost
Mann-Whitney U	527.500	532.000
Wilcoxon W	1268.500	1273.000
Z	-.058	.000
Asymp. Sig. (2-tailed)	.953	1.000
Exact Sig. (2-tailed)	.956	1.000
Exact Sig. (1-tailed)	.478	.501
Point Probability	.003	.003

a. Grouping Variable: GameType

- *Backlog cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgBacklogCost	.00	28	31.57	884.00
	1.00	38	34.92	1327.00
	Total	66		
VarBacklogCost	.00	28	31.14	872.00
	1.00	38	35.24	1339.00
	Total	66		

Test Statistics[a]

	AvgBacklog Cost	VarBacklog Cost
Mann-Whitney U	478.000	466.000
Wilcoxon W	884.000	872.000
Z	-.701	-.856
Asymp. Sig. (2-tailed)	.484	.392
Exact Sig. (2-tailed)	.488	.398
Exact Sig. (1-tailed)	.244	.199
Point Probability	.002	.004

a. Grouping Variable: GameType

Table 5-19: Results for Upstream Echelons (D, F)

- *Order level comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgOrderLev	.00	28	40.20	1125.50
	1.00	38	28.57	1085.50
	Total	66		
VarOrderLev	.00	28	38.36	1074.00
	1.00	38	29.92	1137.00
	Total	66		

Test Statistics[a]

	AvgOrderLev	VarOrderLev
Mann-Whitney U	344.500	396.000
Wilcoxon W	1085.500	1137.000
Z	-2.433	-1.764
Asymp. Sig. (2-tailed)	.015	.078
Exact Sig. (2-tailed)	.014	.079
Exact Sig. (1-tailed)	.007	.039
Point Probability	.000	.001

a. Grouping Variable: GameType

- *Total cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgTotalCost	.00	28	37.43	1048.00
	1.00	38	30.61	1163.00
	Total	66		
VarTotalCost	.00	28	38.96	1091.00
	1.00	38	29.47	1120.00
	Total	66		

Test Statistics[a]

	AvgTotalCost	VarTotalCost
Mann-Whitney U	422.000	379.000
Wilcoxon W	1163.000	1120.000
Z	-1.427	-1.985
Asymp. Sig. (2-tailed)	.154	.047
Exact Sig. (2-tailed)	.155	.047
Exact Sig. (1-tailed)	.078	.024
Point Probability	.001	.001

a. Grouping Variable: GameType

- *Inventory cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgInventoryCost	.00	28	32.91	921.50
	1.00	38	33.93	1289.50
	Total	66		
VarInventoryCost	.00	28	34.86	976.00
	1.00	38	32.50	1235.00
	Total	66		

Test Statistics[a]

	AvgInventoryCost	VarInventoryCost	CVInventoryCost
Mann-Whitney U	515.500	494.000	413.000
Wilcoxon W	921.500	1235.000	1154.000
Z	-.214	-.493	-1.544
Asymp. Sig. (2-tailed)	.830	.622	.123
Exact Sig. (2-tailed)	.834	.627	.125
Exact Sig. (1-tailed)	.417	.313	.062
Point Probability	.003	.002	.002

a. Grouping Variable: GameType

- *Backlog cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgBacklogCost	.00	28	38.32	1073.00
	1.00	38	29.95	1138.00
	Total	66		
VarBacklogCost	.00	28	38.75	1085.00
	1.00	38	29.63	1126.00
	Total	66		

Test Statistics[a]

	AvgBacklogCost	VarBacklogCost
Mann-Whitney U	397.000	385.000
Wilcoxon W	1138.000	1126.000
Z	-1.752	-1.907
Asymp. Sig. (2-tailed)	.080	.056
Exact Sig. (2-tailed)	.080	.057
Exact Sig. (1-tailed)	.040	.028
Point Probability	.001	.001

a. Grouping Variable: GameType

Table 5-20: Results for Retailer Echelons

- *Order level comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgOrderLev	.00	14	19.93	279.00
	1.00	19	14.84	282.00
	Total	33		
VarOrderLev	.00	14	18.86	264.00
	1.00	19	15.63	297.00
	Total	33		

Test Statistics[b]

	AvgOrderLev	VarOrderLev
Mann-Whitney U	92.000	107.000
Wilcoxon W	282.000	297.000
Z	-1.494	-.947
Asymp. Sig. (2-tailed)	.135	.344
Exact Sig. [2*(1-tailed Sig.)]	.142	.358
Exact Sig. (2-tailed)	.139	.358
Exact Sig. (1-tailed)	.070	.179
Point Probability	.002	.009

b. Grouping Variable: GameType

- *Total cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgTotalCost	.00	14	14.43	202.00
	1.00	19	18.89	359.00
	Total	33		
VarTotalCost	.00	14	14.93	209.00
	1.00	19	18.53	352.00
	Total	33		

Test Statistics[b]

	AvgTotalCost	VarTotalCost
Mann-Whitney U	97.000	104.000
Wilcoxon W	202.000	209.000
Z	-1.311	-1.056
Asymp. Sig. (2-tailed)	.190	.291
Exact Sig. [2*(1-tailed Sig.)]	.199	.304
Exact Sig. (2-tailed)	.199	.304
Exact Sig. (1-tailed)	.099	.152
Point Probability	.006	.008

b. Grouping Variable: GameType

- *Inventory cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgInventoryCost	.00	14	20.14	282.00
	1.00	19	14.68	279.00
	Total	33		
VarInventoryCost	.00	14	19.50	273.00
	1.00	19	15.16	288.00
	Total	33		

Test Statistics[b]

	AvgInventory Cost	VarInventory Cost
Mann-Whitney U	89.000	98.000
Wilcoxon W	279.000	288.000
Z	-1.616	-1.284
Asymp. Sig. (2-tailed)	.106	.199
Exact Sig. [2*(1-tailed Sig.)]	.114	.212
Exact Sig. (2-tailed)	.109	.205
Exact Sig. (1-tailed)	.055	.103
Point Probability	.002	.003

b. Grouping Variable: GameType

- *Backlog cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgBacklogCost	.00	14	13.64	191.00
	1.00	19	19.47	370.00
	Total	33		
VarBacklogCost	.00	14	13.86	194.00
	1.00	19	19.32	367.00
	Total	33		

Test Statistics[b]

	AvgBacklog Cost	VarBacklog Cost
Mann-Whitney U	86.000	89.000
Wilcoxon W	191.000	194.000
Z	-1.712	-1.603
Asymp. Sig. (2-tailed)	.087	.109
Exact Sig. [2*(1-tailed Sig.)]	.091	.114
Exact Sig. (2-tailed)	.091	.114
Exact Sig. (1-tailed)	.045	.057
Point Probability	.003	.004

b. Grouping Variable: GameType

Table 5-21: Results for Wholesaler Echelons

- *Order level comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgOrderLev	.00	14	20.21	283.00
	1.00	19	14.63	278.00
	Total	33		
VarOrderLev	.00	14	18.50	259.00
	1.00	19	15.89	302.00
	Total	33		

Test Statistics[b]

	AvgOrderLev	VarOrderLev
Mann-Whitney U	88.000	112.000
Wilcoxon W	278.000	302.000
Z	-1.639	-.765
Asymp. Sig. (2-tailed)	.101	.444
Exact Sig. [2*(1-tailed Sig.)]	.106	.461
Exact Sig. (2-tailed)	.104	.461
Exact Sig. (1-tailed)	.052	.230
Point Probability	.002	.011

b. Grouping Variable: GameType

- *Total cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgTotalCost	.00	14	17.86	250.00
	1.00	19	16.37	311.00
	Total	33		
VarTotalCost	.00	14	18.00	252.00
	1.00	19	16.26	309.00
	Total	33		

Test Statistics[b]

	AvgTotalCost	VarTotalCost
Mann-Whitney U	121.000	119.000
Wilcoxon W	311.000	309.000
Z	-.437	-.510
Asymp. Sig. (2-tailed)	.662	.610
Exact Sig. [2*(1-tailed Sig.)]	.679	.627
Exact Sig. (2-tailed)	.679	.627
Exact Sig. (1-tailed)	.340	.314
Point Probability	.013	.013

b. Grouping Variable: GameType

- *Inventory cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgInventoryCost	.00	14	14.11	197.50
	1.00	19	19.13	363.50
	Total	33		
VarInventoryCost	.00	14	14.79	207.00
	1.00	19	18.63	354.00
	Total	33		

Test Statistics[b]

	AvgInventory Cost	VarInventory Cost
Mann-Whitney U	92.500	102.000
Wilcoxon W	197.500	207.000
Z	-1.477	-1.129
Asymp. Sig. (2-tailed)	.140	.259
Exact Sig. [2*(1-tailed Sig.)]	.142	.271
Exact Sig. (2-tailed)	.144	.271
Exact Sig. (1-tailed)	.072	.135
Point Probability	.003	.008

b. Grouping Variable: GameType

- *Backlog cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgBacklogCost	.00	14	18.14	254.00
	1.00	19	16.16	307.00
	Total	33		
VarBacklogCost	.00	14	17.50	245.00
	1.00	19	16.63	316.00
	Total	33		

Test Statistics[b]

	AvgBacklog Cost	VarBacklog Cost
Mann-Whitney U	117.000	126.000
Wilcoxon W	307.000	316.000
Z	-.583	-.255
Asymp. Sig. (2-tailed)	.560	.799
Exact Sig. [2*(1-tailed Sig.)]	.577	.815
Exact Sig. (2-tailed)	.571	.815
Exact Sig. (1-tailed)	.286	.408
Point Probability	.006	.014

b. Grouping Variable: GameType

Table 5-22: Results for Distributor Echelons

- *Order level comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgOrderLev	.00	14	20.82	291.50
	1.00	19	14.18	269.50
	Total	33		
VarOrderLev	.00	14	19.07	267.00
	1.00	19	15.47	294.00
	Total	33		

Test Statistics[b]

	AvgOrderLev	VarOrderLev
Mann-Whitney U	79.500	104.000
Wilcoxon W	269.500	294.000
Z	-1.949	-1.056
Asymp. Sig. (2-tailed)	.051	.291
Exact Sig. [2*(1-tailed Sig.)]	.050	.304
Exact Sig. (2-tailed)	.051	.304
Exact Sig. (1-tailed)	.026	.152
Point Probability	.001	.008

b. Grouping Variable: GameType

- *Total cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgTotalCost	.00	14	18.36	257.00
	1.00	19	16.00	304.00
	Total	33		
VarTotalCost	.00	14	19.14	268.00
	1.00	19	15.42	293.00
	Total	33		

Test Statistics[b]

	AvgTotalCost	VarTotalCost
Mann-Whitney U	114.000	103.000
Wilcoxon W	304.000	293.000
Z	-.692	-1.093
Asymp. Sig. (2-tailed)	.489	.274
Exact Sig. [2*(1-tailed Sig.)]	.506	.287
Exact Sig. (2-tailed)	.506	.287
Exact Sig. (1-tailed)	.253	.143
Point Probability	.011	.008

b. Grouping Variable: GameType

- *Inventory cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgInventoryCost	.00	14	16.86	236.00
	1.00	19	17.11	325.00
	Total	33		
VarInventoryCost	.00	14	17.36	243.00
	1.00	19	16.74	318.00
	Total	33		

Test Statistics[b]

	AvgInventory Cost	VarInventory Cost
Mann-Whitney U	131.000	128.000
Wilcoxon W	236.000	318.000
Z	-.073	-.182
Asymp. Sig. (2-tailed)	.942	.855
Exact Sig. [2*(1-tailed Sig.)]	.957	.872
Exact Sig. (2-tailed)	.950	.864
Exact Sig. (1-tailed)	.475	.432
Point Probability	.007	.007

b. Grouping Variable: GameType

- *Backlog cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgBacklogCost	.00	14	18.64	261.00
	1.00	19	15.79	300.00
	Total	33		
VarBacklogCost	.00	14	18.57	260.00
	1.00	19	15.84	301.00
	Total	33		

Test Statistics[b]

	AvgBacklog Cost	VarBacklog Cost
Mann-Whitney U	110.000	111.000
Wilcoxon W	300.000	301.000
Z	-.838	-.801
Asymp. Sig. (2-tailed)	.402	.423
Exact Sig. [2*(1-tailed Sig.)]	.418	.439
Exact Sig. (2-tailed)	.418	.439
Exact Sig. (1-tailed)	.209	.219
Point Probability	.010	.011

b. Grouping Variable: GameType

Table 5-23: Results for Factory Echelons

- *Order level comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgOrderLev	.00	14	20.00	280.00
	1.00	19	14.79	281.00
	Total	33		
VarOrderLev	.00	14	19.57	274.00
	1.00	19	15.11	287.00
	Total	33		

Test Statistics[b]

	AvgOrderLev	VarOrderLev
Mann-Whitney U	91.000	97.000
Wilcoxon W	281.000	287.000
Z	-1.530	-1.311
Asymp. Sig. (2-tailed)	.126	.190
Exact Sig. [2*(1-tailed Sig.)]	.132	.199
Exact Sig. (2-tailed)	.130	.199
Exact Sig. (1-tailed)	.065	.099
Point Probability	.002	.006

b. Grouping Variable: GameType

- *Total cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgTotalCost	.00	14	19.86	278.00
	1.00	19	14.89	283.00
	Total	33		
VarTotalCost	.00	14	20.36	285.00
	1.00	19	14.53	276.00
	Total	33		

Test Statistics[b]

	AvgTotalCost	VarTotalCost
Mann-Whitney U	93.000	86.000
Wilcoxon W	283.000	276.000
Z	-1.457	-1.712
Asymp. Sig. (2-tailed)	.145	.087
Exact Sig. [2*(1-tailed Sig.)]	.152	.091
Exact Sig. (2-tailed)	.152	.091
Exact Sig. (1-tailed)	.076	.045
Point Probability	.005	.003

b. Grouping Variable: GameType

- *Inventory cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgInventoryCost	.00	14	16.54	231.50
	1.00	19	17.34	329.50
	Total	33		
VarInventoryCost	.00	14	17.79	249.00
	1.00	19	16.42	312.00
	Total	33		

Test Statistics[b]

	AvgInventory Cost	VarInventory Cost
Mann-Whitney U	126.500	122.000
Wilcoxon W	231.500	312.000
Z	-.237	-.401
Asymp. Sig. (2-tailed)	.813	.689
Exact Sig. [2*(1-tailed Sig.)]	.815	.706
Exact Sig. (2-tailed)	.823	.706
Exact Sig. (1-tailed)	.411	.353
Point Probability	.007	.013

b. Grouping Variable: GameType

- *Backlog cost comparison*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AvgBacklogCost	.00	14	20.29	284.00
	1.00	19	14.58	277.00
	Total	33		
VarBacklogCost	.00	14	20.64	289.00
	1.00	19	14.32	272.00
	Total	33		

Test Statistics[b]

	AvgBacklog Cost	VarBacklog Cost
Mann-Whitney U	87.000	82.000
Wilcoxon W	277.000	272.000
Z	-1.676	-1.858
Asymp. Sig. (2-tailed)	.094	.063
Exact Sig. [2*(1-tailed Sig.)]	.098	.065
Exact Sig. (2-tailed)	.096	.065
Exact Sig. (1-tailed)	.048	.033
Point Probability	.002	.003

b. Grouping Variable: GameType

Table 5-24: Results for Amplification Ratio Comparisons

- *The supply chain (W/ R, D/ W, F/ D)*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AmpRatio	.00	42	50.67	2128.00
	1.00	57	49.51	2822.00
	Total	99		

Test Statistics[a]

	AmpRatio
Mann-Whitney U	1169.000
Wilcoxon W	2822.000
Z	-.198
Asymp. Sig. (2-tailed)	.843
Exact Sig. (2-tailed)	.846
Exact Sig. (1-tailed)	.423
Point Probability	.003

a. Grouping Variable: GameType

- *Amplification ratio between wholesaler and retailer echelons (W/ R)*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AmpRatio	.00	14	16.93	237.00
	1.00	19	17.05	324.00
	Total	33		

Test Statistics[b]

	AmpRatio
Mann-Whitney U	132.000
Wilcoxon W	237.000
Z	-.036
Asymp. Sig. (2-tailed)	.971
Exact Sig. [2*(1-tailed Sig.)]	.986
Exact Sig. (2-tailed)	.986
Exact Sig. (1-tailed)	.493
Point Probability	.014

b. Grouping Variable: GameType

- *Amplification ratio between distributor and wholesaler echelons (D/ W)*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AmpRatio	.00	14	18.86	264.00
	1.00	19	15.63	297.00
	Total	33		

Test Statistics[b]

	AmpRatio
Mann-Whitney U	107.000
Wilcoxon W	297.000
Z	-.947
Asymp. Sig. (2-tailed)	.344
Exact Sig. [2*(1-tailed Sig.)]	.358
Exact Sig. (2-tailed)	.358
Exact Sig. (1-tailed)	.179
Point Probability	.009

b. Grouping Variable: GameType

- *Amplification ratio between factory and distributor echelons (F/ D)*

Ranks

	GameType	N	Mean Rank	Sum of Ranks
AmpRatio	.00	14	16.07	225.00
	1.00	19	17.68	336.00
	Total	33		

Test Statistics[b]

	AmpRatio
Mann-Whitney U	120.000
Wilcoxon W	225.000
Z	-.474
Asymp. Sig. (2-tailed)	.636
Exact Sig. [2*(1-tailed Sig.)]	.653
Exact Sig. (2-tailed)	.653
Exact Sig. (1-tailed)	.327
Point Probability	.013

b. Grouping Variable: GameType

Appendix J: Regression Results

Coefficient of Determination: *Adjusted* R^2

Coefficient of determination refers to the proportionate reduction of the total variation in the response data (dependent variable) that can be obtained by the use of independent variables (Neter et al. 1996).

In our study, we obtain 59.7% average *adjusted* R^2 in Model 3, and 64.6% in Model 11. The values exhibit variation among participants. The range for *adjusted* R^2 is 18.17%-86.90% in Model 3, and 22.21% - 87.71% in Model 11.

Next, we outline the stages of our regression study.

F Test: *P* value

Analysis of variance approach to regression analysis is based on dividing the sums of squares and degrees of freedoms associated to each dependent variable. This approach requires conducting F tests for regression models (Neter et al. 1996).

The overall significance of the regression can be checked with F test. In the F test, null hypothesis states that coefficient of each independent variable is equal to zero. This means that there is no relationship between dependent variable and the independent variables. If the *p* value (smallest level of significance that would lead to the rejection of the null hypothesis) is smaller than the selected significance level (α= 0.05), one can reject the null hypothesis.

Durbin Watson Test: *D* statistics

Residuals from a linear regression should be independent. Durbin Watson test is used to detect the existence of autocorrelation in the residuals. The null hypothesis suggests that there is no autocorrelation in the data. Generally, the residuals tends to show positive autocorrelation, therefore alternative the hypothesis supports positive autocorrelation in the data. Upper and lower critical values are found from the Durbin Watson critical values table according to the significance level, the number of observations and the number of independent variables in the regression model (Montgomery et al. 2001). The test statistic D is compared to lower and upper critical values. At significance level α, decision rule is as follows:

- If $D < D_{l},\alpha$, the residuals are autocorrelated.
- If $D > D_{U},\alpha$, the residuals are not autocorrelated.
- If $D_{l},\alpha < D < D_{U},\alpha$, the test is inconclusive.

In Model 3, 20 out of 28 regressions' residuals are not statistically autocorrelated. For the 8 out of 28, the test is inconclusive. Since there is not enough evidence for rejecting or not rejecting the null hypothesis, we could assume that inconclusive tests refer to no autocorrelation. In Model 11, 18 out of 28 regressions' residuals are not statistically autocorrelated. For the 10 out of 28, the test is inconclusive.

Variance Inflation Factor: *VIF*

Multicollinearity is the dependency (correlation) of independent variables to each other. Detecting multicollinearity in regression analysis is crucial to obtaining correct and reliable beta coefficients for the model. "The *VIF* value shows that how much the variances of the estimated regression coefficients are inflated in comparison to the case that the independent variables are not linearly related" (Neter et al. 1996). When the *VIF* value is greater than 10, this refers to excessive multicollinearity.

In Model 3, we observe that all *VIF* values are smaller than 10. In Model 11, the *VIF* value of the only one participant's backlog independent variable is greater than 10. We exclude this participant from our further analysis.

Normal Probability Plot: *P-P* plot

Normal probability plot is a graphical technique which can be used to evaluate whether the data set is normally distributed or not (Chambers et al. 1983). Residuals should be normally distributed in the linear regression analysis. Therefore, each residual is plotted against its expected value under normality (Neter et al. 1996). If the data set is normally distributed, the line in the graph should be approximately linear.

We graph probability plots of residuals for each regression. Figure 5-6 shows one of our regressions' probability plot. *P-P* plot column in Table 5-25 and Table 5-26 show the results of our normality checks. "Nor" means that residuals are normally distributed, "Not" means that residuals are not normally distributed.

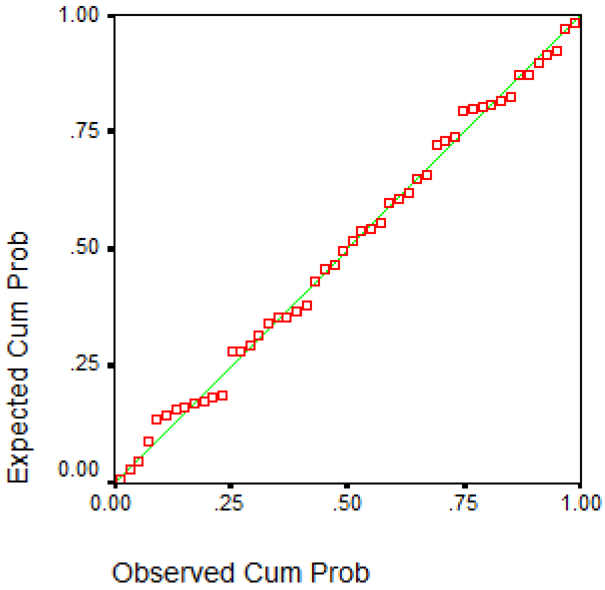


Figure 5-6: Normal *P-P* Plot of Residuals of One of Our Regressions

Bold numbers refer to the regressions that cannot pass the test of the related column.

Table 5-25: Regression Results for Model 3

Participants	**Echelon**	***Adj. R^2***	**Standardized Coefficients**			***D***	***P***	**Variance Inflation Factor *(VIF)***			***P-P* Plot**
			EI(t)	***D(t)***	***∑O(t-i)***			***EI(t)***	***D(t)***	***∑O(t-i)***	
1	**Factory**	60.71%	-0.39	0.40	0.13	2.01	0.00^{**}	1.55	2.09	1.98	Nor
2	**Distributor**	60.93%	-0.51	0.12	0.23	1.81	0.00^{**}	4.23	1.78	3.24	Nor
3	**Wholesaler**	72.61%	-0.23	0.39	0.46	1.16	0.00^{**}	2.17	1.29	1.84	Nor
4	**Retailer**	23.90%	0.07	0.06	0.55	1.19	0.00^{**}	1.72	1.31	1.40	Nor
5	**Factory**	68.37%	-0.44	0.57	-0.17	2.02	0.00^{**}	5.23	3.23	3.02	Nor
6	**Distributor**	86.90%	-0.58	0.48	-0.05	2.32	0.00^{**}	9.38	2.22	6.73	Nor
7	**Wholesaler**	42.04%	-0.12	0.42	0.30	1.73	0.00^{**}	2.37	1.99	1.30	Nor
8	**Retailer**	55.85%	-0.51	0.12	0.33	1.15	0.00^{**}	1.63	1.28	1.84	Nor
9	**Factory**	41.87%	-0.38	0.22	0.24	2.20	0.00^{**}	1.76	1.34	1.40	**Not**
10	**Distributor**	65.50%	0.50	1.21	0.04	1.53	0.00^{**}	6.34	4.91	2.39	**Not**
11	**Wholesaler**	83.67%	-0.41	0.21	0.39	1.20	0.00^{**}	3.65	2.07	2.81	Nor
12	**Retailer**	82.76%	-0.61	0.12	0.37	1.36	0.00^{**}	2.60	1.57	2.88	Nor
13	**Factory**	80.50%	-0.16	0.75	0.11	2.30	0.00^{**}	2.54	1.58	1.83	**Not**
14	**Distributor**	42.83%	-0.06	0.72	-0.14	2.06	0.00^{**}	2.16	2.21	2.18	**Not**
15	**Wholesaler**	59.61%	-0.56	0.33	0.20	1.74	0.00^{**}	1.74	1.08	1.80	Nor

16	**Retailer**	23.17%	-0.31	0.13	0.27	2.13	0.00**	1.24	1.30	1.53	Nor
17	**Factory**	83.40%	-0.24	0.26	0.70	2.34	0.00**	1.09	1.28	1.19	**Not**
18	**Distributor**	24.15%	-0.63	-0.36	0.13	2.47	0.00**	3.70	2.61	1.77	Nor
19	**Wholesaler**	68.60%	-0.24	0.13	0.55	1.36	0.00**	2.80	1.65	3.03	Nor
20	**Retailer**	40.50%	-0.66	0.27	-0.06	1.42	0.00**	2.03	1.54	2.55	**Not**
21	**Factory**	75.85%	-0.26	0.44	0.32	1.19	0.00**	1.39	2.64	2.39	Nor
22	**Distributor**	62.52%	-0.30	0.45	0.35	1.46	0.00**	1.32	1.07	1.33	Nor
23	**Wholesaler**	0.02%	-0.19	-0.13	0.13	1.51	**0.40**	1.16	1.15	1.13	**Not**
24	**Retailer**	-1.11%	-0.18	0.02	0.10	1.42	**0.49**	1.07	1.03	1.08	**Not**
25	**Factory**	82.18%	-0.20	0.52	0.32	1.75	0.00**	2.44	2.40	1.61	Nor
26	**Distributor**	81.88%	-0.25	0.50	0.29	1.31	0.00**	3.02	1.77	2.42	Nor
27	**Wholesaler**	61.41%	-0.70	0.13	0.05	1.45	0.00**	2.42	1.22	2.11	Nor
28	**Retailer**	18.17%	-0.73	0.01	-0.43	1.60	0.01	2.75	1.68	2.67	Nor

Table 5-26: Regression Results for Model 11

Participants	Echelon	*Adj.* R^2	Standardized Coefficients				*D*	*P*	Variance Inflation Factor (*VIF*)				*P-P* Plot
			B(t)	*I(t)*	*D(t)*	*∑O(t-i)*			*B(t)*	*I(t)*	*D(t)*	*∑O(t-i)*	
1	**Factory**	61.78%	0.14	-0.33	0.42	0.11	2.08	0.00**	1.57	1.45	2.11	2.01	**Not**
2	**Distributor**	60.26%	0.44	-0.12	0.12	0.23	1.81	0.00**	3.34	1.76	1.81	3.25	Nor
3	**Wholesaler**	73.61%	-0.02	-0.22	0.40	0.55	1.24	0.00**	2.85	1.28	1.31	2.37	Nor
4	**Retailer**	22.21%	-0.05	0.03	0.06	0.55	1.19	0.00**	1.72	2.62	1.76	1.45	Nor
5	**Factory**	67.66%	0.33	-0.18	0.57	-0.17	2.02	0.00**	**12.13**	1.28	5.54	4.73	Nor
6	**Distributor**	86.63%	0.55	-0.07	0.48	-0.05	2.32	0.00**	8.37	1.45	2.22	6.73	Nor
7	**Wholesaler**	41.88%	0.19	-0.03	0.38	0.26	1.73	0.00**	2.15	1.42	2.14	1.42	Nor
8	**Retailer**	59.80%	0.24	-0.49	0.12	0.27	1.11	0.00**	1.13	1.62	1.28	1.92	Nor
9	**Factory**	47.87%	0.38	-0.27	0.13	0.17	2.43	0.00**	1.48	1.59	1.46	1.48	**Not**
10	**Distributor**	65.74%	-0.44	0.27	1.34	0.08	1.61	0.00**	6.32	2.78	6.74	2.55	**Not**
11	**Wholesaler**	85.40%	0.50	-0.39	0.00	0.18	1.27	0.00**	7.47	2.56	4.45	5.04	Nor
12	**Retailer**	83.06%	0.16	-0.56	0.12	0.33	1.46	0.00**	1.43	2.38	1.57	3.16	Nor
13	**Factory**	82.79%	0.34	-0.03	0.63	0.01	2.19	0.00**	3.41	1.37	2.09	2.28	Nor
14	**Distributor**	43.62%	0.30	0.03	0.67	-0.30	2.00	0.00**	4.36	1.50	2.38	3.53	**Not**
15	**Wholesaler**	61.98%	0.40	-0.34	0.25	0.11	1.77	0.00**	2.26	1.73	1.30	2.07	**Not**

16	**Retailer**	23.54%	-0.06	-0.37	0.18	0.28	2.17	0.00**	1.82	1.80	1.41	1.53	Nor
17	**Factory**	85.30%	0.34	0.06	0.22	0.60	2.23	0.00**	2.06	1.73	1.35	1.69	**Not**
18	**Distributor**	22.60%	0.60	-0.08	-0.37	0.14	2.49	0.00**	3.72	1.20	2.63	1.78	**Not**
19	**Wholesaler**	75.51%	0.75	-0.10	0.11	0.03	1.47	0.00**	6.28	1.31	1.66	6.93	Nor
20	**Retailer**	41.61%	0.39	-0.48	0.26	-0.15	1.44	0.00**	1.91	1.66	1.54	2.94	Nor
21	**Factory**	82.22%	0.39	-0.12	0.29	0.28	1.67	0.00**	1.92	1.33	2.97	2.42	Nor
22	**Distributor**	79.45%	0.76	-0.04	0.33	-0.07	2.27	0.00**	2.90	1.23	1.16	2.42	Nor
23	**Wholesaler**	1.90%	0.33	0.13	-0.08	0.07	1.51	**0.31**	1.58	1.55	1.21	1.22	**Not**
24	**Retailer**	-2.71%	-0.06	-0.22	0.05	0.11	1.43	**0.61**	1.56	1.50	1.14	1.10	**Not**
25	**Factory**	85.04%	0.33	-0.20	0.35	0.22	1.97	0.00**	2.87	1.96	3.31	1.96	Nor
26	**Distributor**	87.71%	0.43	-0.29	0.39	0.05	1.59	0.00**	2.54	2.52	1.96	3.44	Nor
27	**Wholesaler**	60.56%	0.18	-0.59	0.13	0.05	1.45	0.00**	1.61	2.21	1.24	2.18	Nor
28	**Retailer**	19.09%	0.07	-0.84	-0.06	-0.54	1.54	0.01	1.23	4.15	1.85	3.18	**Not**

Table 5-27: Results for SRM1

Participants	Echelon	Adj R2	Standardized Coefficients								
			EI(t)	**D(t)**	**∑O(t-i)**	**O(t-1)**	**O(t-2)**	**O(t-3)**	**D(t)-D(t-1)**	**In backlog (1,0)**	**D(t)-D(t-1) (1, 0)**
1	**Factory**	65.10%	-0.34	0.62					-0.25		
2	**Distributor**	60.20%	-0.78								
3	**Wholesaler**	80.40%		0.35		0.70					
4	**Retailer**	39.90%				0.64					
5	**Factory**	68.00%	-0.30	0.57							
6	**Distributor**	87.10%	-0.53	0.49							
7	**Wholesaler**	39.60%								0.64	
8	**Retailer**	67.80%	-0.29			0.51				0.21	
9	**Factory**	65.10%		0.59					-0.58	0.37	
10	**Distributor**	66.20%	0.46	1.21							
11	**Wholesaler**	84.80%	-0.41			0.55					
12	**Retailer**	83.80%	-0.42			0.56					
13	**Factory**	81.40%	-0.25	0.82							-0.15
14	**Distributor**	47.70%		0.58							0.23
15	**Wholesaler**	63.10%	-0.48	0.32		0.32					

16	**Retailer**	45.40%	-1.28	0.49						-1.02	
17	**Factory**	89.50%	-0.23	0.53	0.26			0.34	-0.32		
18	**Distributor**	24.80%	-0.75	-0.43							
19	**Wholesaler**	66.20%			0.82						
20	**Retailer**	43.50%	-0.41			0.34					
21	**Factory**	82.30%		0.23		0.52				0.25	
22	**Distributor**	67.20%		0.45		0.36				0.31	
23	**Wholesaler**	8.20%				0.32					
24	**Retailer**	10.60%				0.35					
25	**Factory**	84.60%		0.63		0.38			-0.14		
26	**Distributor**	84.30%		0.48		0.54					
27	**Wholesaler**	61.70%	-0.79								
28	**Retailer**	22.70%	-0.61					-0.38			

Table 5-28: Results for SRM2

Participants	Echelon	Adj R2	Standardized Coefficients						
			B(t)	**D(t)**	**I(t)**	**∑O(t-i)**	**O(t-1)**	**O(t-2)**	**O(t-3)**
1	**Factory**	61.40%	-0.40	0.54					
2	**Distributor**	59.60%	0.65		-0.21				
3	**Wholesaler**	80.40%		0.35			0.70		
4	**Retailer**	39.90%					0.64		
5	**Factory**	68.10%	-0.19	0.74					
6	**Distributor**	86.80%	0.52	0.50					
7	**Wholesaler**	46.70%		0.48			0.39		
8	**Retailer**	67.00%	0.19		-0.29		0.52		
9	**Factory**	47.20%	0.47		-0.38				
10	**Distributor**	62.40%		0.80					
11	**Wholesaler**	87.10%	0.40		-0.32		0.34		
12	**Retailer**	83.00%			-0.38		0.60		
13	**Factory**	83.50%	0.36	0.64					
14	**Distributor**	44.30%		0.67					
15	**Wholesaler**	58.50%	0.51				0.34		

16	**Retailer**	24.50%			-0.29	0.36			
17	**Factory**	89.30%	0.28	0.28		1.21		-0.62	
18	**Distributor**	24.40%	0.74	-0.41					
19	**Wholesaler**	75.00%	0.87						
20	**Retailer**	40.60%			-0.33		0.42		
21	**Factory**	85.00%	0.34	0.20			0.48		
22	**Distributor**	79.90%	0.72	0.34					
23	**Wholesaler**	8.20%					0.32		
24	**Retailer**	10.60%					0.35		
25	**Factory**	83.40%		0.55			0.46		
26	**Distributor**	88.60%	0.35	0.38	-0.22		0.19		
27	**Wholesaler**	61.50%			-0.56		0.28		
28	**Retailer**	24.40%			-0.66				-0.45

Bibliography

Ackoff, R. L. 1967. Management misinformation systems. *Management Science* **14**(4) 147-156.

Ambrus, A., B. Greiner, P. Pathak. 2009. Group versus individual decision-making: Is there a shift? Working Paper, Princeton University.

Bendoly, E., K. Donohue, K. L. Schultz. 2006. Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management* **24** 737–752.

Blanchard, O. J. 1983. The production and inventory behavior of the American Automobile Industry. *The Journal of Political Economy* **91**(3) 365-400.

Blinder A.S., J. Morgan. 2010. Are two heads better than one? An experimental analysis of group vs. Individual decision making. Working Paper, Princeton University.

Bolton, G. E., A. Ockenfels, U. Thonemann. 2008. Managers and students as newsvendors- How out-of-task experience matters. Working Paper, University of Cologne.

Bourland, K. E., S. G. Powell, D. F. Pyke. 1996. Exploiting timely demand information to reduce inventories. *European Journal of Operational Research* **92**(2) 239-253.

Brehmer, B. 1992. Dynamic decision making: Human control of complex systems. *Acta Psychologica* **81**(3) 211-241.

Cachon, G. P., M. Fisher. 1997. Campbell Soup's continuous replenishment program: evaluation and enhanced inventory decision rules. *Production and Operations Management* **6**(3) 266-276.

Cachon, G. P., M. Fisher. 2000. Supply chain inventory management and the value of shared information. *Management Science* **46**(8) 1032-1048.

Cantor, D. E., E. Katok. 2008. The bullwhip effect and order smoothing in a laboratory beer game. Working Paper, Iowa State and Penn State University.

Cantor, D. E., J. R. Macdonald. 2009. Decision making in the supply chain: Examining problem solving approaches and information availability. *Journal of Operations Management* **27** 220-232.

Chambers, J., W. Cleveland, B. Kleiner, P. Tukey. 1983. *Graphical Methods for Data Analysis*. Duxbury Press, Boston.

Chen, F. 1998. Echelon reorder points, installation reorder points, and the value of centralized demand information. *Management Science* **44**(12) 221-234.

Chen, F. 1999. Decentralized supply chains subject to information delays. *Management Science* **45**(8) 1076-1090.

Chen, F., Z. Drezner, J. K. Ryan, D. S. Levi, 2000a. Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science* **46**(3) 436-443.

Chen, F., J. K. Ryan, D. S. Levi, 2000b. The impact of exponential smoothing forecasts on the bullwhip effect. *Management Science* **47** 269-286.

Chen L., H. L. Lee. 2010. Bullwhip effect measurement and its implications. Working Paper, Duke and Standford University.

Clark, T., H. Hammond. 1997. Reengineering channel reordering processes to improve total supply-chain performance. *Production and Operations Management* **6**(3) 248-265.

Cooper, D. J., J. H. Kagel. 2005. Are two heads better than one? Team versus individual play in signaling games. *American Economic Review* **95**(3) 477-509.

Croson, R., K. Donohue. 2003. Impact of point of sale (POS) data sharing on supply chain management: An experimental study. *Production and Operations Management* **12**(1) 1-11.

Croson, R., K. Donohue. 2005. Upstream versus downstream information and its impact on the bullwhip effect. *System Dynamics Review* **21**(3), 249-260.

Croson, R., K. Donohue, E. Katok, J. D. Sterman. 2005. Order stability in supply chains: Coordination risk and the role of coordination stock. MIT Sloan School of Management.

Croson, R., K. Donohue. 2006. Behavioral causes of the bullwhip effect and the observed value of inventory information. *Management Science* **52**(3) 323-336.

De Kok, T., F. Jannsen, J. Doremalen, E. Wachem, M. Clerkx, W. Peeters. 2005. Philips Electronics synchronizes its supply chain to end the bullwhip effect. *Interfaces* **35**(1) 37-48.

Delhoum, S., B. S. Reiter. 2009. The influence of decision patterns of inventory control on the bullwhip effect based on a simulation game of a production network. *Production Planning & Control* **20**(8) 666-677.

Devore J. L. 1995. *Probability and Statistics for Engineering and the Sciences*. Duxbury Press.

Diehl, E., J. D. Sterman. 1995. Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes* **62**(2) 198-215.

Disney, M., M. Naim, A. Potter. 2004. Assessing the impact of e-business on supply chain dynamics. *International Journal of Production Economics* **89** 109-118.

Disney, M. S., M. R. Lambrecht. 2007. On replenishment rules, forecasting, and the bullwhip effect in supply chains. *Foundations and Trends in Technology, Information and Operations Management* **2**(1) 1-80.

Dumain, B. 1994. The trouble with teams. *Fortune* **130**(5) 86-92.

Fair, R. C. 1989. The production smoothing model is alive and well. *Journal of Monetary Economics* **24**(3) 353-370.

Forrester, J. 1958. Industrial dynamics: a major breakthrough for decision makers. *Harvard Business Review* **36** 37–66.

Fransoo, J. C., M. J. F. Wouters. 2000. Measuring the bullwhip effect in the supply chain. *Supply Chain Management: An International Journal* **5**(2) 78-89.

Gaur, V., A. Giloni, S. Seshadri. 2005. Information sharing in a supply chain under arma demand. *Management Science* **51**(6) 961-969.

Gavirneni, S., R. Kapuscinski, S. Tayur. 1999. Value of information in capacitated supply chains. *Management Science* **45**(1) 16-24.

Ghali, M. A. 1987. Seasonality, aggregation and the testing of the production smoothing hypothesis. *The American Economic Review* **77**(3) 464-469.

Grubbs, F. E. 1969. Procedures for detecting outlying observations in samples. *Technometrics* **11**(1) 1–21.

Hammond, J. H. 2008. Barilla SpA. Case 9-694-046. Harvard Business School Publishing, Boston, MA.

Harps L. H. 2002. Making Dollars & Sense Out of Logistics.

Holland, W., M. S. Sodhi. 2004. Quantifying the effect of batch size and order errors on the bullwhip effect using simulation. *International Journal of Logistics Research and Applications* **7**(3) 251-261.

Jacobs, F. R. 2000. Playing the beer distribution game over the internet. *Production and Operations Management* **9**(1) 31-39.

Jerath, K., S. Netessine, J. Zhang. 1997. Can we all get along? Incentive contracts to bridge the marketing and operations divide. Working Paper, The Wharton School University of Pennsylvania.

Kaminsky, P., D. Simchi-Levi. 1998. A new computerized beer game: A tool for teaching the value of integrated supply chain management. *Global Supply Chain and Technology Management.*

Kampmann, C., J. D. Sterman. 1998. Do markets mitigate misperceptions of feedback in dynamic tasks?. Working Paper, Sloan School of Management, MIT.

Kocher, M. G., M. Sutter. 2005. The decision maker matters. Individual versus team behavior in experimental beauty-contest games. *Economic Journal* **115** 200-223.

Kocher, M. G., S. Strauss, S. Matthias. 2006. Individual or team decision-making--Causes and consequences of self-selection. *Games and Economic Behavior* **56**(2) 259-270.

Kulp, S. C., H. L. Lee, E. Ofek. 2004. Manufacturer benefits from information integration with retail customers. *Management Science* **50**(4) 431–444.

Kurt Salmon Associates 1993. Efficient consumer response: Enhancing consumer value in the grocery industry. Food Marketing Institute, Washington, DC.

Laughlin, P. R. 1980. Social combination process of cooperative problem solving groups at verbal intellective tasks. *Social Psychology* **1** 127-155.

Lee, H. L., P. Padmanabhan, S. Whang. 1997a. Information distortion in a supply chain: the bullwhip effect. *Management Science* **43** 546–558.

Lee, H. L., P. Padmanabhan, S. Whang. 1997b. The bullwhip effect in supply chains. *Sloan Management Review* **38**(3) 93-102.

Lee, H. L., K. C. So, C. S. Tang. 2000. The value of information sharing in a two-level supply chain. *Management Science* **46**(5) 626–643.

Machuca, J. A. D., R. P. Barajas. 2004. The impact of electronic data interchange on reducing the bullwhip effect and supply chain inventory costs. *Transportation Research Part E* **40** 209–228.

Manyem, P., D. L. Santos. 1999. Impact of multiple retailers on the bullwhip effect: A supply chain simulation study. *Proceedings of the 4th International Conference on Industrial Engineering Theory Applications and Practice*.

Manz, C. C., H. P. Sims. 1993. Business without bosses. Wiley, New York.

Massart, D. L., J. S. Verbeke, X. Capron, K. Schlesier. 2005. Visual presentation of data by means of box plots. *Lc-gc-Europe* **18**(4) 215-218.

McGill, R, J. W. Tukey, W. A. Larsen. 1978. Variations of Box plots. *The American Statistician* **32**(1) 12-16.

Metters, R. 1997. Quantifying the bullwhip effect in supply chains. *Journal of Operations Management* **15** 89-100.

Montgomery, D. C., E. A. Peck, G. G. Vining. 2001. *Introduction to Linear Regression Analysis.* 3rd Edition, John Wiley & Sons, New York.

Munson, C. L., J. Hu, M. J. Rosenblatt. 2003. Teaching the costs of uncoordinated supply chains. *Interfaces* **33**(3) 24-39.

Neter, J., M. H. Kutner, C. J. Nachtsheim, W. Wasserman. 1996. *Applied Linear Statistical Models*. 4th Edition, WCB/McGraw-Hill, New York.

Nienhaus, J., A. Ziegenbein, C. Duijts. 2003. How human behaviour amplifies the bullwhip effect - A study based on the beer distribution game online. *Proceedings of the 7th Workshop on Experimental Interactive Learning in Industrial Management.*

Nienhaus, J., A. Ziegenbein, P. Schoensleben. 2006. How human behavior amplifies the bullwhip effect. A study based on the beer distribution game online. *Production Planning & Control* **17**(6) 547-557.

Oliva, R., P. Gonçalves. 2007. Behavioral causes of the bullwhip effect: Satisficing policies with limited information cues. *Under revision for resubmission to Journal of Operations Management*.

Oliva, R., N. H. Watson. 2007. Cross-functional alignment in supply chain planning: A case study of sales and operations planning. Working Paper, Harvard Business School.

Osterman, P. 1995. How common is workplace transformation and who adapts it? *Industrial and Labor Relations Review* **47**(2) 173-87.

Paich, M., J. D. Sterman. 1993. Boom, bust, and failures to learn in experimental markets. *Management Science* **39**(12) 1439-1458.

Potter, A., S. M. Disney. 2006. Bullwhip and batching: An exploration. *International Journal on Production Economics* **104** 408-418.

Raghunathan, S. 2001. Information sharing in a supply chain: A note on its value when demand is nonstationary. *Management Science* **47**(4) 605-610

Ruël, G. C., D. P. van Donk, J. T. van der Vaart. 2006, The beer game revisited: Relating risk-taking behaviour and bullwhip effect. *International EurOMA Conference*, **1** 403-412.

Shapiro, B. P. 1977. Can marketing and manufacturing coexist? *Harvard Business Review* **55**(5) 104–114.

Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences*. McGraw Hill, New York.

Simchi-Levi D., P. Kaminsky, E. Simchi-Levi, 2007. *Designing and Managing the Supply Chain: Concepts, Strategies and Case Studies*. 3rd edition, McGraw-Hill, New York.

Steckel J. H., S. Gupta, A. Banerji. 2004. Supply chain decision making: Will shorter cycle times and shared point-of-sale information necessarily help? *Management Science* **50**(4) 458-464.

Sterman, J. 1989a. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science* **35**(3) 321-339.

Sterman, J. 1989b. Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes* **45**(3) 301-335.

Sterman, J. D. 2006. Operational and behavioral causes of supply chain instability, in: O. Carranza, F. Villegas (Eds.), *The Bullwhip Effect in Supply Chain*, Palgrave McMillan.

Wu, D. Y., E. Katok. 2006. Learning, communication and the bullwhip effect. *Journal of Operations Management* **24** 839-85.