Exploring Retailer's Ordering Decisions under Delays

Sebastián Villa
University of Lugano, Switzerland
Via Buffi 13, CH-6904 Lugano, Switzerland.
e-mail: villabes@usi.ch

Paulo Gonçalves
University of Lugano, Switzerland
Via Buffi 13, CH-6904 Lugano, Switzerland.
e-mail: paulo.goncalves@usi.ch

Santiago Arango
Universidad Nacional de Colombia
Carrera 80 no. 65 – 223 M8A Medellín
e-mail: saarango@unal.edu.co

ABSTRACT:
When final customer demand exceeds available supply, retailers often hedge against shortages by inflating orders to their suppliers. While the amplification in orders caused by competition for scarce resources has been described in the literature almost a century ago, there is little experimental research examining the factors influencing retailer's order amplification.
This paper analyzes retailer order decisions in response to a surge in demand. In an experimental environment based on a formal mathematical model we test subject's ordering decisions under different ordering and supplier capacity acquisition delays and compared them to an optimal benchmark. Our results from different treatments allow us to characterize subjects' performance in this system and formulate a heuristic that closely replicates subjects' ordering behavior in all treatments.

KEYWORDS:
Order Amplification, Laboratory Experiment, Behavioral Operations, Supply Chain Management, Demand Bubbles, System Dynamics.

1. INTRODUCTION
In the past decades, different approaches have been used to understand the interrelations among suppliers, retailers and customers. These relations are typically approached through the development of mathematical models that mimic real world situations, such as non-

stationary demand, physical delays, backlogs, order amplifications, etc. (Ilkyeong, 2005). The general objective is to produce and distribute the product and services at the right quantities, to the right locations, and at the right time, in order to minimize system-wide costs while satisfying service level requirements (Zhang, 2006). One of the most common and costly problems in supply chains is caused by retailer orders' amplification (Armony and Plambeck 2005). These amplifications have been captured in the literature as early as 1924, when Mitchell described the case of retailers inflating their orders to manufacturers when competing with other retailers for scarce supply. He argued "if [retailers] want 90 units of an article, they order 100, so as to be sure, each, of getting the 90 in the pro rata share delivered." (Mitchell 1924, p. 645). When faced with limited capacity, suppliers typically allocate available supply among multiple retailers. In turn, retailers receiving only a fraction of previous orders, amplify future ones in an attempt to secure more units (Lee et al., 1997a, 1997b). This phenomenon can propagate the supply chain causing orders and inventories to chronically overshoot and undershoot around desired levels. These fluctuations can lead retailers and suppliers alike to overreact, leading to problems such as excessive supplier capital investment, inventory gluts, low capacity utilization, and poor service (Armony and Plambeck, 2005; Gonçalves, 2003; Lee et al., 1997a; Sterman, 2000; Anderson et al., 1999).

Academic interest in the subject has its roots on real and frequent problems faced by businesses in diverse industries. For example, in the 80's, the computer industry faced shortages of DRAM chips in several occasions: orders surged because of retailers anticipation (Li, 1992). Similarly Hewlett-Packard could not distinguish between real and inflated orders place from the resellers for the laserJet printers; which later lead to excess inventory and unnecessary capacity (Lee et al., 1997a). In the summer of 2000, Cisco began to experience shortages of key components and this caused customers' order amplification. Cisco failed to recognize the magnitude of customers' order amplification, and the sales forecast were overestimated. This caused a strong capacity expansion through long term contracts with its OEMs. Once capacity became available and possible delivery delays went back to normal levels, customers canceled duplicated orders, leaving Cisco with significant excess capacity, rigid long-term contracts and a remarkable amount of inventory (Byrne and Elgin, 2002).

Informed by these industry experiences, our research explores the impact that delays may have on subjects' ordering decisions. We hypothesize that longer retailer ordering delays and supplier capacity acquisition delays increase retailer order amplification. Both conditions are consistent with studies by Sterman (1989a, 1989b) and Gonçalves (2003) and Gonçalves and Arango (2010). Our experimental setting is based on a system dynamics model adapted from Gonçalves (2003) and Gonçalves and Arango (2010). The model captures retailers' order amplification when competing for scarce supply. Our results show that subjects systematically deviate from an optimal dynamic control. As expected, longer capacity acquisition and order delays complicate the subjects' ordering task, leading to higher order amplification. While subjects' ordering behavior is not optimal, it can be explained econometrically by a simple decision rule.

This paper proceeds as follows. Section 2 describes and analyzes the proposed mathematical model. Section 3 develops a decision-making laboratory experiment based on the model developed in Section 2. Section 4 shows our main results; particularly that subjects' performance deteriorates under longer ordering and capacity acquisition delays. Section 5 derives an econometric model based on a simple anchoring and adjustment to analyze subjects' decision rules. Finally, we discuss the main finding in Section 6.

2. MODEL DESCRIPTION

We build upon a model proposed by Gonçalves (2003) capturing a supply chain with a single supplier offering a unique, non-substitutable, product to multiple retailers. The emphasis of our analysis is on retailers' ordering problem trying to match supplier shipments and final customer demand. Figure 1 displays the structure of this supply chain.

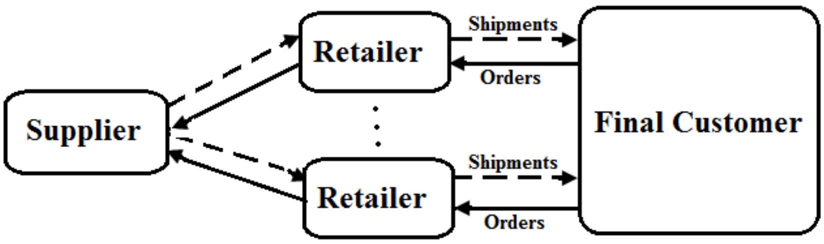


Figure 1. Supply Chain structure

To model the supplier system, we first define the supplier's backlog of orders (*B*) as a function of retailers' orders (*Rd*) and supplier shipments (*S*).

$$\dot{B} = R_D - S \tag{1}$$

Shipments (*S*) are typically given by the minimum between of desired shipments and the available capacity. However, since we are interested in situations characterized by supply shortages, we model shipments as always constrained by available capacity (*K*).

$$S = K \tag{2}$$

The change in supplier's capacity ($\dot{K}$) is given by a first order exponential smooth between desired shipments (S^*) and capacity (K), with an adjustment time given by the time to build capacity (τ_K). This formulation captures a naïve capacity adjustment process, where the supplier tries to maintain sufficient capacity to satisfy customer demand with a target delivery delay. Finally, desired shipments (S^*), given by the ratio of Backlog (B) and the Target Delivery Delay (τ_D), capture the shipment rate required to maintain delivery delays at the target level for the existing level of backlog. This process can be written as:

$$\dot{K} = \frac{B/\tau_D - K}{\tau_K} \tag{3}$$

Modeling supplier capacity as a first-order exponential smooth of desired shipments follows a traditional formulation in system dynamics. In addition, Gonçalves and Arango (2010) find empirical support for this formulation for supplier's capacity investment. Finally, to measure retailers' ability to meet final customer demand, we also capture the supply gap, measuring the difference between ($\dot{D}_R$) a level that accumulates total orders from final customer orders (d), and ($\dot{E}_S$) a level that accumulates the total shipments received from the supplier.

$$\dot{D}_R = d \tag{4}$$

$$\dot{E}_S = S \tag{5}$$

Figure 2 provides an overview of the supplier-retailer model driving the lab experiment.

Cost Objective

To motivate subjects' performance, we measure retailers' total cost (TC) given by two components: (1) a Supply Gap Cost (C_{gap}), given by the summed differences between cumulative customer demand and cumulative shipments received from the supplier; and (2) Ordering Cost (C_o), given by the retailer's order decisions (R_D). In addition, we assume quadratic costs to penalize larger deviations from equilibrium.

$$TC = \sum_{t=1}^{T} (C_{gap} + C_o) \tag{6}$$

Where,

$$C_{gap} = \alpha \cdot (D_R - E_S)^2, \quad \text{where } \alpha = 2 \cdot 10^{-3} \tag{7}$$

$$C_o = \beta \cdot R_D^{\ 2}, \text{ where } \beta = 1 \cdot 10^{-3} \tag{8}$$

The parameters α and β were chosen such as they are comparable given the order of magnitude of the variables R_D, D_R and E_S Appendix 1 presents the general units of measure used for each variable or parameter.

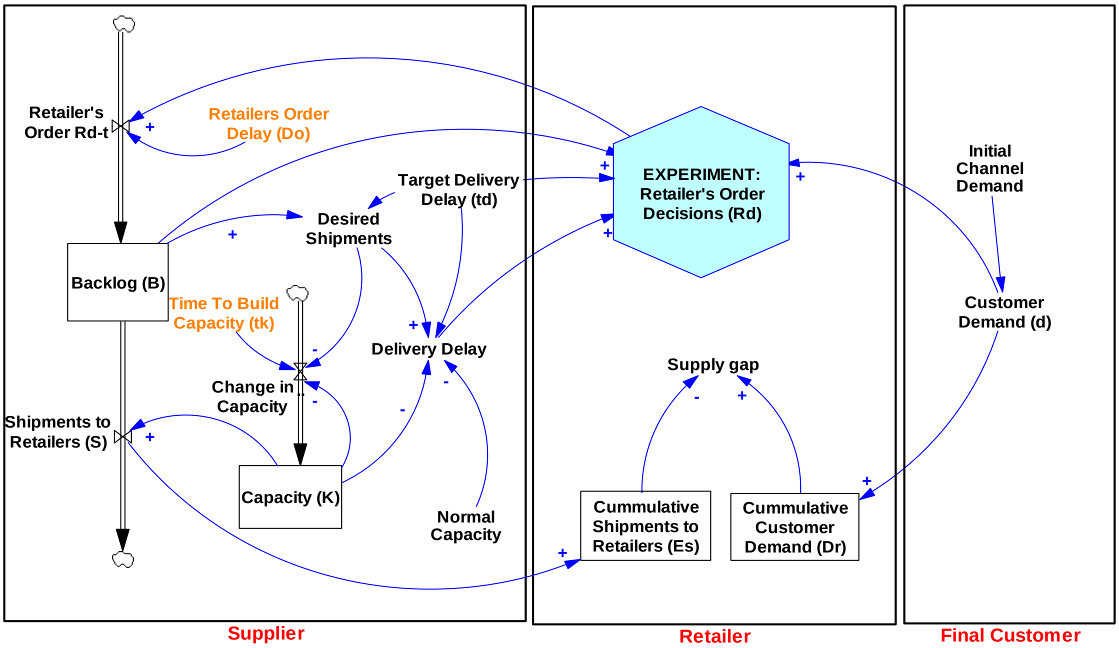


Figure 2. System dynamics diagram for supplier-retailer system.

3. THE EXPERIMENT

We use the model described above as a basis for a "management flight simulator" (Sterman 1989, Senge and Sterman 1992). Subjects play the role of a single retailer, placing orders to a supplier and trying to minimize total costs. The experiment starts in dynamic equilibrium. Initially the supplier has sufficient production capacity to meet total retailer's demand according to the target delivery delay. After the third period (week), the supplier faces a sudden increase in retailer's orders. Subjects are informed that customer demand will increase in 20% and that the supplier faces a delay to build additional capacity. Subjects must decide how many units to order from the supplier each week through 35 simulated weeks. Subjects are asked to minimize the total accumulated cost (*TC*), during 35 weeks

3.1. Experimental Treatments

Our experiment explores two characteristics previously identified by Gonçalves (2003) and Gonçalves and Arango (2010) affecting the performance of retailers' decisions: retailer ordering delays and supplier capacity acquisition delays. We use a full experimental design, with four experimental treatments. Table 1 specifies all treatments conducted in the experiment and the number of participants (N) in each treatment. We model the retailer ordering delay (Δ_O) as a pipeline delay

Table 1. Experimental treatments.

		Supplier's Capacity Investment Delay (τ_K)	
		1	3
Retailer's Order Decision Delay (Δ_O)	2	**T1** (*N=18*)	**T2** (*N=18*)
	3	**T3** (*N=18*)	**T4** (*N=18*)

3.2. Protocol

We follow standard experimental economics protocol (see Friedman & Sunder, 1994 and Friedman & Cassar, 2004). Subjects were fourth and fifth year Industrial and Management Engineering students at the National University of Colombia, in the autumn of 2010. The subjects did not have previous experience in a related experiment. Participants were told they would earn a show-up fee of Col$10.000 (approximately US$5) and a variable amount contingent on their performance, between Col$0 and Col$30.000 (US$0 - US$15) for an overall average payoff of Col$24.000 (US$12). The experiment ran for around 1 hour and students were informed about the duration of the experiment. The payoff was more than two times larger than their opportunity cost in Colombia. The students were also given a set of instructions describing the production system, the decisions and the goals of the game (Appendix 2)

We ran the experiment with 18 subjects per treatment. Upon arrival, subjects were seated behind computers and a treatment was assigned randomly (see Appendix 3). Participants were allowed to ask questions and test out the computer interface (Appendix 4). All the experiment parameters were common knowledge to all participants. The experiment was run in the computer simulation software *Powersim-Constructor-2.51®*. The software ran automatically and kept record of all variables, including subjects' decisions. Subjects were also asked to write their decisions in a sheet of paper, which served as a physical backup of the data.

3.3 Optimal Simulated Trajectory

In order to have a framework for comparison, we find an optimal simulated trajectory for each treatment. These optimal retailers' order decision trajectories were estimated using the Solver in *Powersim Studio 8* (Appendix 5 shows the optimization specifications used) and minimizing the total cost over all periods. Figure 3 shows the behavior of these optimal trajectories, considering the ordering decisions are made under deterministic demand. For optimization purposes, this Powersim Studio 8 Solver uses a method called the evolutionary search method (for more details related with this method see Appendix 6).

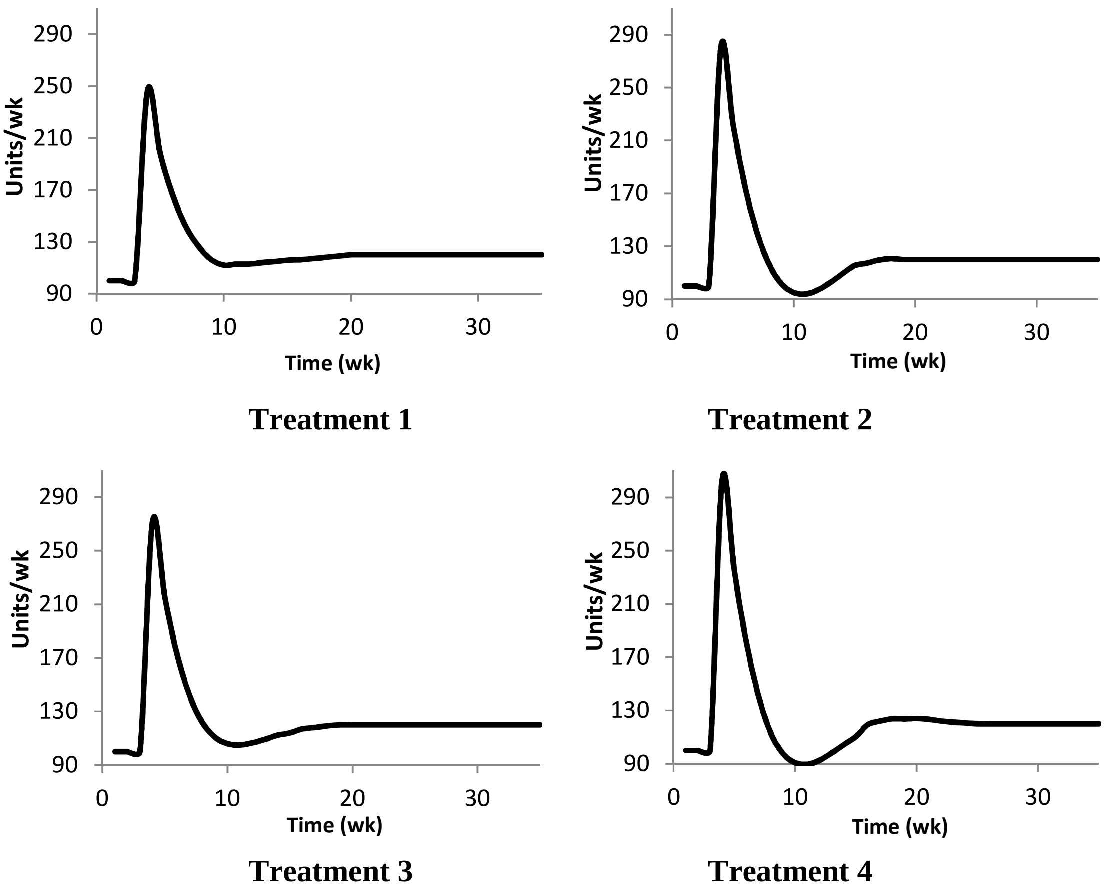


Figure 3. Optimal Retailer's order decisions

Figure 3 shows that the optimal ordering trajectories are characterized by a large initial order at the moment the demand surges. The magnitude of this optimal initial order increases with the complexity (longer delays) of the system. Then, orders exponentially decrease with a final damped oscillation until settling on 120 units per week. The magnitude of the damped oscillation also increases with the complexity of the system. Finally, optimal orders settle at 120 units per week for the rest of the trajectory.

4. RESULTS

In this section we present the overall results of the experiments. We report the four experimental treatments with 15 subjects per treatment, which are chosen among the 18 subjects base on one variance method (see appendix 7 for criteria selection).

4.1. Subjects' Order decisions Behavior

Subjects received information on system structure, delays and costs and then were asked to place orders that would minimize total simulated long-run costs. Figure 4 shows ordering

behavior for four selected subjects (one in each treatment) capturing typical subjects behavior. The results suggest common pattern: subjects’ orders initially over-shoot, then under-shoot until settling around equilibrium close to 120 units (the final demand).
Figure 4 also shows that subjects in treatments T1 and T2 (with shorter ordering delays) over-order for shorter periods of time (around 10 week). In such treatments the shorter ordering delays allows subjects to more quickly adjust their orders.

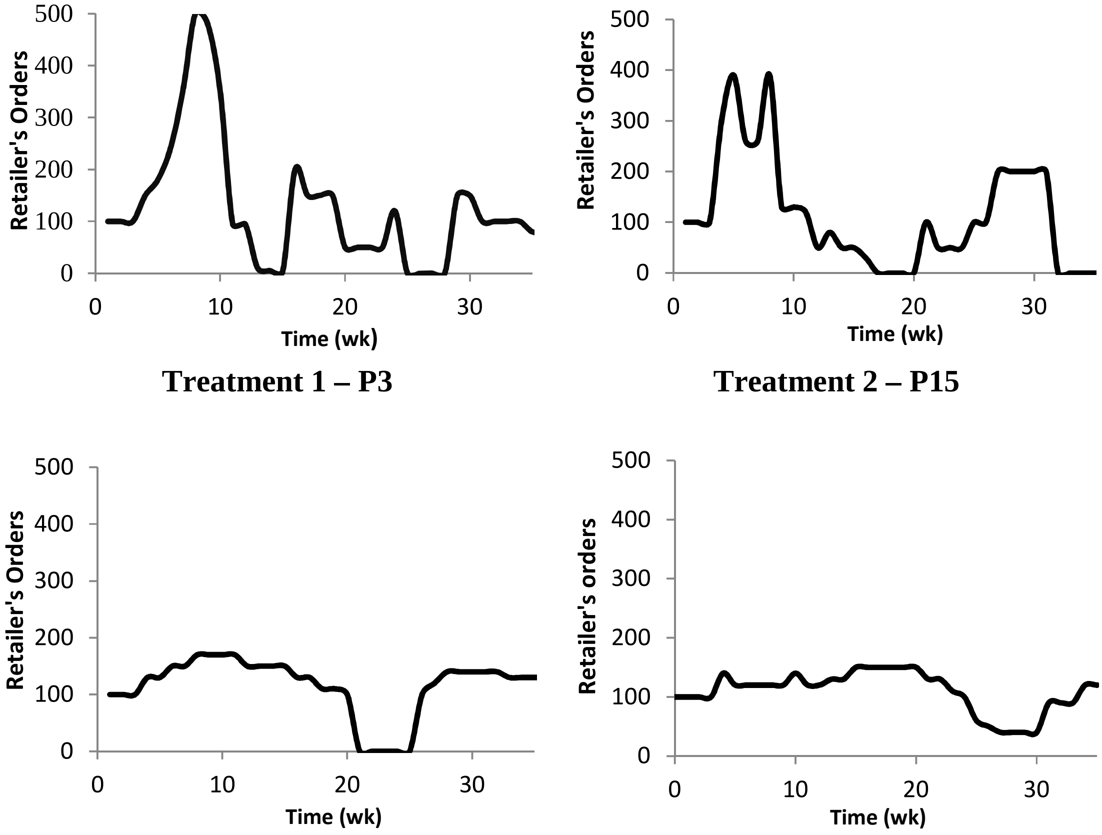


Figure 4. Typical experimental results (*Pj* indicates the subject ID with $j = 1,\ldots,15$)

Figure 4 also shows that subjects in treatments T1 and T2 (with shorter ordering delays) over-order for shorter periods of time (around 10 week). In such treatments the shorter ordering delays allows subjects to more quickly adjust their orders. However, it seems that in treatments T3 and T4 the subjects are more conservative in their initial orders. This could mean that in a certain way subjects are not completely forgetting their supply line.
To compare overall subject behavior in each treatment with the optimal ordering decisions, we compute the average retailer´s orders (AO) for players in each treatment. Figure 5 suggests that subjects fail to place sufficiently large initial orders, and also fail to reduce

them quickly toward the equilibrium value. Instead, subjects place orders with magnitudes averaging half of the desired initial value, but maintain high orders for a longer period than desired. When subjects finally reduce their orders, they do so more than the optimal values. As a result subjects' ordering behavior fluctuates around the optimal trajectory in all treatments. While the pattern presents similarities across treatments, it is also possible to identify differences. The peak in subjects' decision tends to be wider in treatments with longer retailer ordering delays (T3 and T4). Subject's decisions are less stable and take longer to settle in the treatment with higher delays (T4).

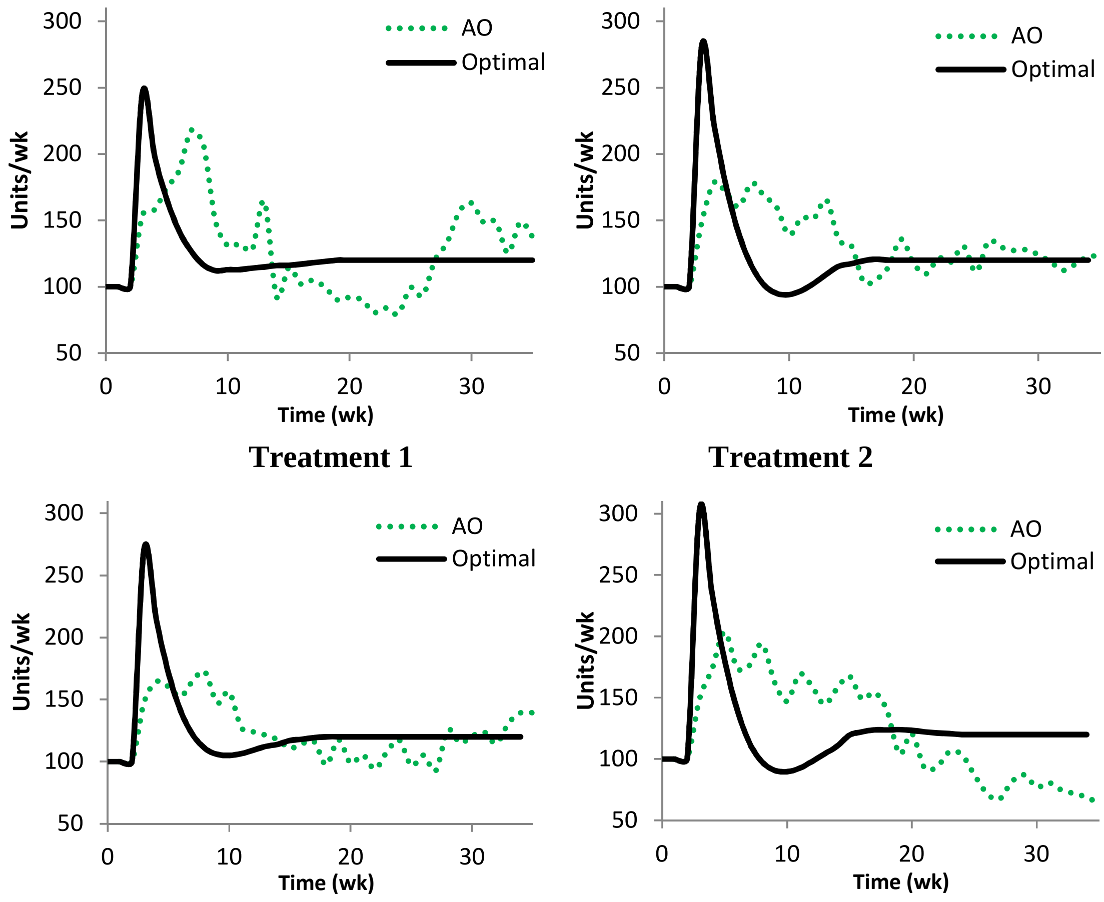


Figure 5. Optimal and Average Subjects' Orders (*AO*) in each treatment

4.2. Subjects' Cost Performance

The main subject's objective in the experiment was to minimize cumulative costs. Table 2 presents total cumulative costs per subject and the average, the minimum and the optimal for each treatment.

A general observation is that most of the subjects perform far from optimal, for all treatments. The lowest total cost achieved by a subject was 33% higher than the optimal of the treatment, which occurred for subject P8 in treatment 2 (T2). The best performances observed in the other treatments were also above optimal costs: 39% above optimal in treatment 1 (T1), 40% above optimal in treatment 3 (T3), and 96% above optimal in treatment 4 (T4).

Table 2. Total cumulative, average, and optimal costs across treatment for the experiment.

Subject	T1	T2	T3	T4
P1	$2,331.95	$3,243.45	$1,285.69	$41,569.73
P2	$16,186.75	$1,313.24	$6,349.57	$1,576.23
P3	$17,921.82	$6,806.07	$1,439.31	$1,976.48
P4	$3,995.25	$3,017.32	$2,441.01	$19,297.83
P5	$845.60	$878.90	$8,854.54	$12,619.97
P6	$3,834.15	$899.15	$1,407.55	$37,655.30
P7	$6,805.24	$14,624.58	$2,086.54	$30,220.81
P8	$25,358.16	$854.73	$3,946.30	$4,258.17
P9	$4,056.73	$10,944.48	$2,410.65	$2,214.28
P10	$1,664.46	$2,438.42	$2,958.85	$1,403.54
P11	$1,511.78	$1,002.29	$1,202.67	$2,294.44
P12	$1,193.47	$46,973.16	$1,106.48	$2,649.15
P13	$4,790.34	$2,792.65	$885.04	$13,445.40
P14	$805.86	$7,144.60	$961.68	$35,200.88
P15	$27,068.96	$5,362.77	$1,719.34	$1,388.21
Average	$7,891.37	$7,219.72	$2,603.68	$13,851.36
Min	$805.86	$854.73	$885.04	$1,388.21
Optimal	$579.10	$643.67	$632.40	$707.58

The subjects' average performances vary from 400% to 1900% higher than the optimal. (These results are conservative since we have excluded subjects with outlying ordering behavior.) The lowest benchmark costs is observed for treatments 1 ($579.1) and highest cost is in treatment 4 ($707.0), these results highlight the increasing system difficulty when higher delays are introduced producing lower performances. In general, subjects' decisions have lower total cumulative cost in simpler treatments (shorter delays) and higher total cumulative cost in more complex treatments. For instance, in treatments 1, 2 and 3, most of the cumulative costs are smaller than $10.000, but in treatment 4 we have several values above $20.000. In treatment T3, it seems that given the conservative decisions of the subjects during the first periods and the fast supplier response could lead to a lower long-term costs, but also higher adjustment time (Figure 4). Table 3 shows how cost components contributed to optimal and average subjects' total cost in each treatment. The cost breakdown in the optimal trajectory suggests that most of the costs are given by the ordering component. Hence, the choice of parameters α and β induces orders that minimize the Supply Gap and its associated cost. In contrast, the cost breakdown for the subjects' decisions shows that subjects have a difficult time balancing supply and demand, placing orders that fail to minimize the Supply Gap. Hence, subjects have a disproportionally high fraction of their costs due to the Supply Gap cost component.. As expected, in the most dynamic complex treatment (T4), subjects incur the highest proportion of costs due to the Supply Gap.

Table 3. Costs distribution given by Orders and Supply gap.

	% Cost given by Orders	**% Cost given by Supply Gap**	**% Cost given by Orders**	**% Cost given by Supply Gap**
	T1		**T2**	
Average	26.94%	73.06%	29.25%	70.75%
Optimal	98.52%	1.48%	90.45%	9.55%
	T3		**T4**	
Average	29.69%	70.31%	17.73%	82.27%
Optimal	94.60%	5.40%	85.17%	14.83%

5. MODELING DECISION RULES

For modeling the subjects' decision rules, we test the heuristic proposed by Gonçalves (2003). Gonçalves modeled retailers' orders, R_D, using an *anchor and adjustment heuristic*, where retailers anchor their orders on a demand forecast, and then adjust it up or down to maintain orders at a desired level. The anchor term captures retailers' intention to place sufficient orders to meet their customers' orders. The adjustment term closes the gap between retailers' desired and actual backlog of orders within a specific adjustment time. Gonçalves (2003) also assumes that each retailer adopts the same heuristic with the model capturing total values for customer demand forecast (d), actual backlog of orders (B), desired backlog of orders (B^*), and adjustment time (τ_B). Finally, total retailers' orders are non-negative (no cancellations). Equation (9) shows this heuristics

$$\dot{R}_D = Max\left(0, d + \frac{B^* - B}{\tau_B}\right) \qquad \textbf{(9)}$$

Retailers' desired backlog of orders (B^*) is given by the product of the demand forecast, d and the expected delivery delay to receive orders from the supplier (ED).

$$B^* = d \cdot ED \qquad \textbf{(10)}$$

Gonçalves (2003) assumes that the expected delivery delay is given by a linear function (*f*), with slope α, of the actual delivery delay (*AD*).The function (*f*) captures retailers' delivery delay adjustment, that is, when faced with long delivery delays, retailers set their expected delivery delay (*ED*) above the actual delivery delay (*AD*) quoted by the supplier. Longer expected delivery delays (*ED*) than actual (*AD*) leads to higher desired backlog of orders (B^*) and higher retailers' orders.

$$ED = \alpha \text{ AD, where } \alpha \geq 1 \qquad \textbf{(11)}$$

Where, actual delivery delay (*AD*) is given by the ratio of the order backlog (*B*) to shipments (*S*).

The qualitative similarity of the results shown in the previous section could suggest the subjects use a heuristic with common features (Sterman 1989a). Substituting equations 10 and 11 into 9 we obtain equation 12, which can be used to test if retailers' orders are given by anchoring and adjustment heuristic. The general model behavior base on equations 1 to 12 can be observed in Appendix 8.

$$\dot{R}_D = Max\left(0, d + \frac{d * \alpha * B / K - B}{\tau_B}\right) \qquad \textbf{(12)}$$

The system determined by equation 12 involves a nonlinearity associated with the ratio of the two states: order backlog (*B*) and capacity (*K*). We can linearize the system using a Taylor series approximation of the ratio of the two states (*B*/*K*) around the initial backlog (B_0) and capacity (K_0) and neglect higher order terms. .

$$AD = \frac{B}{K} = \frac{B_0}{K_0} + (B - B_0)\frac{1}{K}\bigg|_{B_0,K_0} + (K - K_0)\frac{B}{K^2}\bigg|_{B_0,K_0} =$$

$$\frac{B}{K} = \frac{B_0}{K_0} + (B - B_0)\frac{1}{K_0} + (K - K_0)\frac{B_0}{K_0^2} \qquad \textbf{(13)}$$

and since in equilibrium we have that the supplier´s initial supplier capacity (K_0) is equal to $K_0 = B_0 / \tau_D$, the linearized form for delivery delays is given by:

$$AD = \frac{B}{K} = \tau_D \left(1 + \left(\frac{B/\tau_D - K}{K_0} \right) \right) \qquad \textbf{(14)}$$

Substituting in 14 in 12, we get:

$$\dot{R}_D = Max\left(0, d + \frac{d * \alpha * \tau_D \left(1 + \left(\frac{B/\tau_D - K}{K_0} \right) \right) - B}{\tau_B} \right) \qquad \textbf{(15)}$$

Finally, grouping terms and taking the linear part we get a linear approximation of the *anchor and adjustment heuristic* proposed by Gonçalves (2003), which could be tested econometrically. We have the model as:

$$\dot{R}_D = Max\left(0, \left(d + \frac{\alpha * d * \tau_D}{\tau_B} \right) - \left(\frac{\alpha * d * \tau_D}{\tau_B * K_0} \right) K + \left(\frac{\alpha * d - K_0}{\tau_B * K_0} \right) B \right) \qquad \textbf{(16)}$$

Equation 16 captures the change in retailers' orders expected when retailers adopt an anchoring and adjustment heuristic in the simulation model. The model will be analyzed using two different methods. First, we estimate the unknown parameters for each subject using the ordinary least squares (OLS); and second, we structured the data from the game as a panel in order to increases the efficiency of the estimations and the representativeness of the resulting rule. It allows us to make estimations across individuals and treatments.

5.1. OLS Analysis

Initially, the model shown in the equation 16 can be rewritten as:

$$\dot{R}_{D(t,j)} = \beta_{(0,j)} + \beta_{(1,j)} K_{(t)} + \beta_{(2,j)} B_{(t)} + \varepsilon_{(t,j)} \qquad \textbf{(17)}$$

Where $\beta_{(i,j)}$ represent the *i-th* coefficient for the treatment *j*, *i=0,1,2* and *j=1..4*, and $\varepsilon_{(t,j)}$ is the error term. According to the formulation of this decision rule, we expect a coefficient of $\beta_{(0,j)}$ larger than 120, $\beta_{(1,j)} < 0$ to be negative and $\beta_{(2,j)}$ to be positive. Estimations of the coefficients using the linearized heuristic are given in Table 4**.** In this estimations we take the equilibrium values used in the model (τ_D=10, τ_B= 4, K_0=100 and d=120).

Table 4. Coefficient estimates.

β_0	β_1	β_2
450	-3.3	0.08

For analyzing the significance and polarity of the parameters using OLS, we estimate the model of equation 17 using the software **R** 2.12.2. Table 5 shows:

1. The estimations for $\beta_{(i,j)}$ for each subject.
2. The estimations for $\beta_{(i,j)}$ after play the experiment using the average decisions obtained per treatment.
3. The average values of $\beta_{(i,j)}$ for all the subjects that have significant values.
4. It shows the r^2 values for all regressions.

A priori, we expected $\beta_{(0,j)}$ to be above than 120, $\beta_{(1,j)}$ to be negative and $\beta_{(2,j)}$ to be positive. We can analyze these parameters mathematically; however, possibly it should intuitive because the lower the supplier's shipments, the higher would be the retailer' orders ($\beta_{(1,j)}$ negative). And, higher backlogs would indicate supplier delivery problems, and therefore higher retailer orders ($\beta_{(2,j)}$ positive).

The results show that a high fraction of models are significant (at 5%). For example, we found significant values for the three parameters in 52% of the whole subjects. We observe that just 40% of R-squares are larger than 0.40 (See appendix 9 for assumptions validation). Table 5 also shows that the R-squares after running the simulator using the average decisions are between 0.47 and 0.67. The coefficient β_1 is consistent with our expectations, with negative values and significant in the large majority. We found significant values for 53%, 80%, 73%, and 67% in treatments T1, T2, T3, and T4 respectively. Also, most of the signs (78%) of this parameter are negative. On the other hand, the coefficient β_2 has positive values in the large majority (77%), with significant values for 60%, 67%, 67%, and 60% in treatments T1, T2, T3, and T4 respectively. The constant β_0 is positive and significant for the 88% of the sample.

Table 5. Coefficient estimates of decision rule for each individual for all treatments

	Treatment 1				**Treatment 2**			
Subject	**β_0**	**β_1**	**β_2**	**R^2**	**β_0**	**β_1**	**β_2**	**R^2**
1	120.00†	0†	0†	0.49	150.91†	**2.57†**	0.21†	0.26
2	51.82	-2.72†	0.31†	0.18	235.50†	-2.17†	0.13†	0.74
3	152.33†	-3.35†	0.30†	0.18	176.03†	-2.62†	0.22†	0.20
4	158.21†	-8.01†	0.77†	0.60	411.27†	-2.95†	0.02	0.21
5	399.56†	1.24	-0.35†	0.60	300.41†	-1.45†	0.01	0.72
6	213.21†	-1.57	0.06	0.02	246.57†	-1.28†	0.03	0.50
7	86.56†	-9.09†	0.93†	0.71	342.49†	-2.06†	0.02†	0.69
8	200.65†	-1.00	0.04	0.02	269.08	1.59	0.35†	0.11
9	149.31†	-5.85†	0.55†	0.31	428.02†	-1.99†	-0.04†	0.89
10	238.34†	0.11	-0.11	0.05	267.14†	-3.89†	0.25†	0.56
11	147.57†	0.09	-0.03	0.06	199.24†	-0.79	0.02	0.25
12	86.49	-8.61†	0.89†	0.76	178.31†	-3.20†	0.29†	0.62
13	19.83	-1.39	0.25	0.72	181.57†	-3.26†	0.28†	0.38
14	209.31†	-0.48	-0.02	0.01	-50.04	1.97	-0.06	0.00
15	276.87†	-4.11†	0.33†	0.20	189.95†	-2.45†	0.18†	0.21
Average*	157.22	-6.08	0.53	0.42	238.88	-2.69	0.17	0.51
Using Average Decisions	204.29†	-6.61†	0.59†	0.59	244.74†	-1.99†	0.11	0.65
	Treatment 3				**Treatment 4**			
Subject	**β_0**	**β_1**	**β_2**	**R^2**	**β_0**	**β_1**	**β_2**	**R^2**
1	221.21†	-4.43†	0.36	0.15	407.04†	0.21	-0.12†	0.89
2	397.79†	0.52	-0.21†	0.51	143.26†	2.48†	0.24	0.27
3	278.59†	-0.29	-0.07	0.04	283.51†	-3.68†	0.22†	0.62
4	196.76†	1.66	-0.24	0.00	50.69	-3.66†	0.40†	0.78
5	78.39†	-8.06†	0.83†	0.55	285.54†	-2.24†	0.14†	0.15
6	493.60†	-3.08†	-0.02†	0.72	258.51†	-2.17†	0.15†	0.14
7	212.80†	-1.09	0.01	-0.03	221.68†	-3.25†	0.25†	0.34
8	177.36†	-7.46†	0.72†	0.28	108.62†	-2.22†	0.23†	0.30
9	213.39†	-3.65†	0.29†	0.13	255.72†	-0.55	-0.06	0.57
10	180.94†	-2.47†	0.17	0.08	124.97†	-1.09	0.08	0.01
11	204.84†	-6.43†	0.56†	0.51	111.33†	1.01	-0.09	0.01
12	159.97†	1.34†	-0.16†	0.35	195.81†	-4.31†	0.38†	0.57
13	203.33†	-5.57†	0.49†	0.44	140.76†	-1.75†	0.15†	0.08
14	301.77†	-3.80†	0.24†	0.46	158.89†	-1.08†	0.09	0.12
15	86.88†	-3.47†	0.38†	0.24	9.40	-0.19	0.08	0.01
Average *	213.28	-4.47	0.37	0.41	213.49	-2.81	0.22	0.31
Using Average Decisions	170.19†	-5.98†	0.56†	0.47	156.86†	-3.20†	0.29†	0.67

5.2. Panel Data Analysis

In order to control for variables we cannot observe or measure and to study omitted variables that vary over time but are constant between treatments, we structured the data from the experiments as a panel. It increases the efficiency of the estimations and improves the representativeness of the decision rule and accounts for individual heterogeneity. In order to explain overall behavior, we assume that the covariance between variables equal to zero and that there is occurrence of random effects across individual for each treatment (Greene, 1997). The panel data showed in table 6 were estimated using **R**2.12.2.

Table 6. Coefficient estimates of decision rule for treatment as panel data

Regressor	Treatment 1	Treatment 2	Treatment 3	Treatment 4
β_0 (intercept)	100.75† (15.36)	133.46† (11.96)	127.85† (16.012)	106.91† (13.197)
β_1 (Capacity)	-3.916† (0.435)	-2.403† (0.191)	-4.37† (0.483)	-1.727† (0.222)
β_2 (Backlog)	0.413† (0.042)	0.226† (0.016)	0.433† (0.047)	0.183† (0.019)
F-statistic	47.7325	93.0629	42.4084	44.9515
P - Value	2.22E-16	2.22E-16	2.22E-16	2.22E-16
R -Square				
within	0.1377	0.2389	0.1228	0.1227
between	0.6276	0.5883	0.6504	0.5936
random	0.1625	0.2745	0.1461	0.1535
N Obsevations	495	495	495	495

† Significant at 1% and Standard error (SE) in parentheses

The significance test of the model using the statistic F shows that all the *P* values are smaller than 0.05, which implies that coefficients are different than zero and we cannot reject the model. The Table 6 also shows the R-squares (within, between, and random) for each treatment. Despite the fact that the overall R-squares are low for within and random, it shows high between values, ranking from 0.58 for T2 to 0.65 for T3. This means that subjects tend to make similar decisions.

The coefficients in all treatments are highly significant. The β_1 coefficients are negative, but the magnitude is lower than the expected value (Table 4, Figure 6). On the one hand, the β_2 coefficients are positive for all treatments but around 4 times higher than the expected value (see Table 4). Finally, the constants β_0 are close to 120 units and highly significant for all treatments.

Figure 6 summarizes the information obtained in the previous sections. First, we built a box-plot using the estimated parameters obtained for all subject in each treatment; and second, the estimated parameters using the heuristic, the mean values of the parameters for each treatment and the panel data parameters estimations are included in each box-plot. The box-plots show a general distribution for each parameter, indicating whether they are skewed or not. For instance, β_2 in T2 has a short plot and it is almost a symmetric plot, which means that the sample is compact and probably the estimations could have less uncertainty. Figure 6 also shows that the 70% of the different kinds of estimations are between the Q1 and Q3 in all treatments. This percentage is not higher because all of the β_0 estimated values (with the heuristic) are higher than the experimental results. It means that the heuristic is creating an overestimation of the independent parameter. However, the results in β_1 (Capacity coefficient) and β_2 (Backlog coefficient) show that the estimated values using the heuristic are in general with the right sign and also in the expected range given by the boxplot and the panel data estimators.

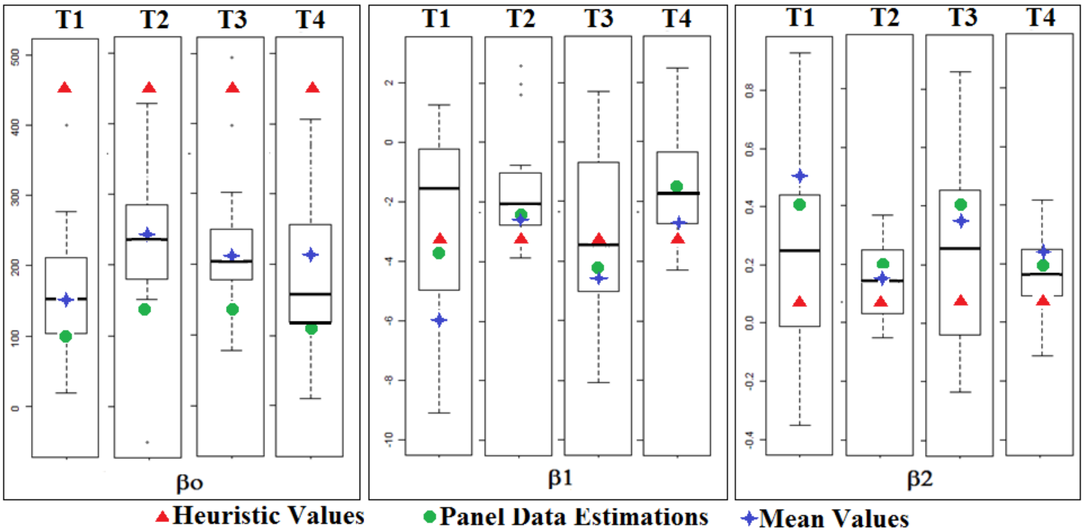


Figure 6. Coefficient estimations of decisions Box plot, Heuristic, mean and panel data estimations

Given the estimations obtained with the panel (Panel) and those obtained using the average OLS (Adjusted AO), we insert and run them into the same model of the experiment. Figure 8 shows the behavior of these runs over time. We observe that these simulations replicate the pattern of behavior given by the actual subjects' behavior (AO). However, the magnitude of the oscillations is much lower in the simulated ones. On the other hand, it can be seen that the proposed heuristic (Heuristic) also replicates the actual subjects' decisions in all treatments, with the underperformance previously described. However, the length of the oscillations is different (Figure 7).

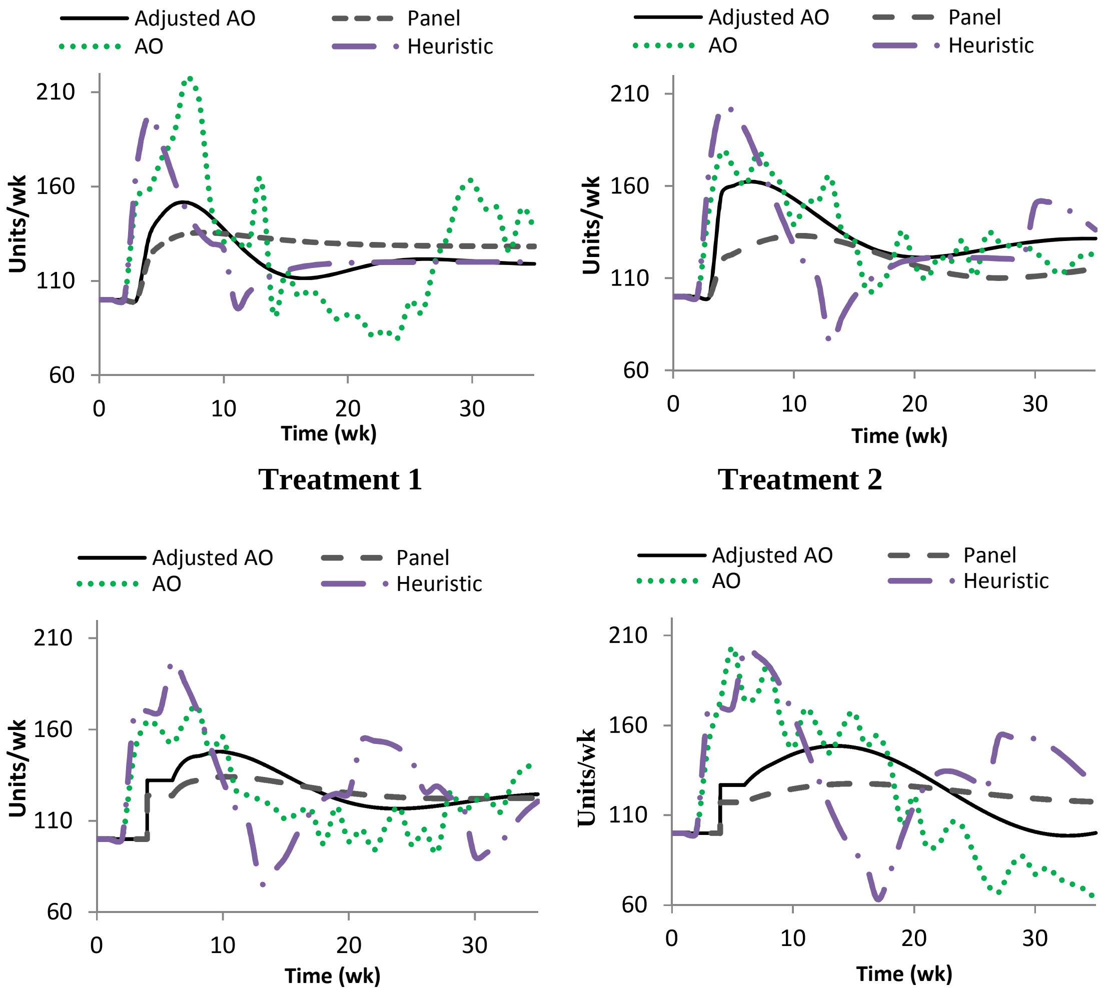


Figure 7. Average Subjects' Orders (AO), Simulation of the proposed heuristic (Heuristic), Adjusted model with Average Subjects' Orders (Adjusted AO) and with Panel parameters (Panel).

6. DISCUSSION AND FURTHER RESEARCH

In this paper, we develop a laboratory experiment to explore how subjects playing the role of a retailer place orders in response to a surge in final customer demand. Subjects must try to minimize cumulative costs, given by the sum of a Supply Gap Cost and an Ordering Cost, under ordering and capacity acquisition delays. We explore the ordering behavior of subjects facing two types of delays: retailer ordering delays and supplier capacity acquisition delays.

To set a performance benchmark, we characterize the optimal ordering trajectory for each experimental treatment. The optimal trajectory is given by a large initial order followed by an exponential decrease that undershoots below initial orders, and a damped oscillation into the final equilibrium of 120 units per week. The magnitude of the peak in the optimal ordering trajectory varies across treatments, increasing with longer system delays. Our selection of cost parameters results in optimal trajectories with total costs driven by the Ordering component (e.g., 98.5% of the total costs in T1), and retailers ability to close any supply/demand gap.

Our experimental results show that subjects' orders deviate widely from the optimal orders. Subjects fail to place sufficiently large initial orders and also fail to reduce them quickly toward the equilibrium. Instead, subjects orders are lower than the amounts initially required but are kept high for longer than optimal. When subjects orders finally are reduced, subjects do so in excess, under-ordering below the optimal levels. Despite having access to complete system information, subjects have limited ability to process and interpret the impact of delays and feedback on overall system behavior. This is commonly known as misperception of feedback (Sterman 1989a). As expected, subject performance differs per treatment, and in particular subjects' performance decreases with longer delays and increased dynamic complexity (Diehl and Sterman 1995). For instance, subjects' orders remain high longer than optimal and as subjects face longer retailer ordering delays (T3 and T4) the duration of high orders increases further deviating from optimal. These results suggest that when possible, retailers should try to decrease the delays inherent in their ordering processes. In doing so, retailers would reduce the complexity of the system and improve their ordering decisions and their ability to manage any mismatches between supply and demand. Our results, those shorter delays lead to simpler, easier to manage systems, which yield to lower costs, are aligned with those of Sterman (1989), Kaminsky and Simchi-Levi (1998), and Gupta et al (2001).

The experimental costs also provide clues about the sources of subjects' underperformance. In general, cumulative costs are closer to optimal in treatments with shorter delays (T1) and further from the optimal in treatments with longer delays (T4). Comparing the costs associated with subjects' decisions with those from optimal ones, we note that subjects' average performances vary from 400% to 1900% higher than the optimal, with the best (i.e., lowest) performance still being 33% higher than the optimal. In addition, subjects fail to minimize Supply Gap during the experiment, incurring the associated long-term costs.

For example, in Treatment 4, around 82% of the subjects' total costs are given by the Supply Gap component.

Given their limited processing and cognitive capability, people make decisions translating complex information into simplistic models, either by capturing essential features from problems and not taking all the features into account, or by developing habits and routines (Lazaric 2000; Simon 1982). Our econometric analysis suggests that subjects use a simple anchoring and adjustment heuristic (Tversky and Kahneman 1974) to place orders.

Analysis of individual regressions had in general good model fit and also suggest that the model explains a significant portion of the variability in the ordering data. In particular, 40% of R-squares were larger than 0.40. Furthermore, the coefficients of capacity (K) and backlog (B) in the individual regressions have the expected signs (negative and positive, respectively) and are significant for more than 50% of subjects. Naturally, while heuristics are simple and useful, they also can lead to consistent biases, limited search and resistance to change (Leonard-Barton, 1992; Lazaric, 2000). In practice, managers should be careful when relying on rules-of-thumb. In our simulated experiment, the adopted heuristic performs substantially worse than the optimal, which suggests significant opportunity for improvement.

The analysis of average decisions produces high significance in the estimated parameters' values and their expected signs. Moreover, the panel data estimations show that the model is significant and the coefficients of the variables are significant and have the expected signs for all treatments. Finally, we conclude that the differences among subjects do not contribute to explain the unexplained variance.

This research could have some limitations regarding the quantitative application in systems under uncertain demand, however we believe the qualitative analyses should hold. On the other hand, in order to extend and improve the results presented here, we thought in the following ideas. First, given the optimal decision trajectory was obtained using simulations, we would like to obtain mathematically a close form solution of this trajectory, and then try to find a robust solution applicable in realistic situations with uncertainty. Second, it could be interesting to find out if subjects can learn to make better decisions and if so, do they retain this learning over time or revert to old patterns as time passes. Third, multiple subjects could interact amongst themselves placing orders to the same supplier, whose responses would be simulated by the computer. Forth, our simulator allocates available supply in proportion to orders placed. Different allocation mechanisms could be explored. Fifth, subjects could be subjected to a different cost function and potentially be exposed to specific information cues. Finally, we could potentially analyze other kinds of heuristics.

REFERENCES

Anderson E., Fine C. 1999. “Business Cycles and Productivity in Capital Equipment Supply Chains.” In Tayur et al. Quantitative Models for Supply Chain Management. Kluwer Academic Publishers, Norwell, MA.

Armony M., Plambeck E. 2005. “The Impact of Duplicate Orders on Demand Estimation and Capacity Investment”. Management Science 51 (10), p. 1505-1518.

Diehl E., Sterman J.D. 1995. “Effects of Feedback Complexity on Dynamic Decision Making” Organizational Behavior and Human Decision Processes. **62**(2): pp. 198- 215.

Foremski, T. 1999. “Intel struggles to meet strong demand for chips” Financial Times (London), November18, p.42.

Friedman D., Sunder S. 1994. “Experimental Methods: A Primer for Economists”. Cambridge University Press: Cambridge.

Friedman D., Cassar A. 2004. “Economics Lab: An Intensive Course in Experimental Economics”. Routledge: London.

Gaither, C. 2001. “Intel Beats Forecast; Warns of Revenue Shortfall,” The New York Times, January 17, C1.

Gonçalves P. 2003. “Phantom Orders and Demand Bubbles in Supply Chains”. Ph.D. Dissertation, Sloan School of Management, MIT, Cambridge, MA.

Gonçalves P., Arango S. 2010. “Supplier Capacity Decisions Under Retailer Competition and Delays: Theoretical and Experimental Results”, 28th International Conference of the System Dynamics Society, Seul-Korea, 2010.

Gupta S., Steckel J.H., Banerj A. 2001. Dynamic Decision Making in Marketing Channels: An Experimental Study of Cycle Time, Shared Information and Consumer Demand Patterns. In: A. Rapoport, and R. Zwick (Eds), Experimental Business Research. Kluwer Academic Publishers, Boston, MA.: 21-48.

Ilkyeong M., Bibhas C., Byungsung K. 2005. “Economic order quantity models for ameliorating/deteriorating items under inflation and time discounting”. Production, Manufacturing and Logistics. European Journal of Operational Research 162. p. 773–785

Kaminsky, P., Simchi-Levi, D., 1998. A New Computerized Beer Game: A Tool for Teaching the Value of Integrated Supply Chain Management. In: H. Lee, and S.M. Ng (Eds), Supply Chain and Technology Management. Production and Operations Management Society: 216-225

Lazaric N. 2000. The role of routines, rules and habits in collective learning: Some epistemological and ontological considerations. European Journal of Economic and Social Systems 14(2): 157-171

Lee H., Padmanabhan V., Seungjin W. 1997a. "Information Distortion in a Supply Chain: The Bullwhip Effect." Management Science, 43(4): p. 546-558.

Lee, H., Padmanabhan, V, and Seungjin W. 1997b. "The Bullwhip Effect in Supply Chains." Sloan Management review, Spring: p. 93-102.

Leonard-Barton D. 1992. Strategy Process: Managing Corporate Self-Renewal. Strategic Management Journal: 111-125.

Mitchell, T.W. 1924. "Competitive Illusion as a Cause of Business Cycles." Quarterly Journal of Economics, 38(4): p. 631-652.

Sterman, J.D. 1989a. "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment." Management Science. 35(3): p. 321-339.

Sterman J.D. 1989b. "Misperceptions of Feedback in Dynamic Decision making." Organizational Behavior and Human Decision Sciences, 43, 3: p. 301-335.

Sterman J.D. 2000. " Business Dynamics: Systems Thinking and Modeling for a Complex World." Chicago, IL, Irwin-McGraw Hill.

Tversky A and Kahneman D. 1974. Judgment under Uncertainty: Heuristics and Biases. Science, 185(4157): 1124–1131.

Zhang C. 2006. "Design and simulation of demand information sharing in a supply chain, Department of Information Management and Information Systems", School of Management, Fudan University, Shanghai-China.

Appendix 1. Model specifications

Variable, Stock or Parameter	Symbol	Initial Value	Units
Backlog of orders	$\dot{B}$	1000	Units
Cumulative Customer Orders	$\dot{D}_R$	100	Units
Cumulative Supplier Shipments	$\dot{E}_S$	100	Units
Retailers' orders	$\dot{R}_D$	100	Units/wk
Supplier Shipments	S	100	Units/wk
Supplier Capacity	K	100	Units/wk
Final Customer Orders	d	100	Units/wk
Retailers' total costs	TC	10	$
Supply Gap Costs	C_{gap}	0	$/wk
Order costs	C_o	10	$/wk
Target Delivery delay	τ_D	10	wks
Time to build Capacity	τ_K	1 or 3	wks
Time Adjust Backlog	τ_B	4	wks
Linear Coefficient	α	1.1	Dimensionless

Appendix 2. Instructions for T3 (in Spanish)

INSTRUCCIONES

POR FAVOR NO TOCAR EL COMPUTADOR HASTA QUE NO SE LE INDIQUE

Bienvenido, a partir de este momento usted hace parte de un experimento de toma de decisiones, en el cual asumirá el papel de gerente de una empresa mayorista. Su responsabilidad es **minimizar los costos acumulados** al final de la simulación del juego (**50** semanas), y de acuerdo a su desempeño obtendrá un pago en dinero efectivo, de un proyecto de investigación patrocinado por la Universidad Nacional de Colombia, Sede Medellín.

Su decisión semanal es definir cuántas unidades ***ordenar*** a su proveedor, con el fin de cubrir toda la demanda de sus clientes (en el experimento, esta decisión se toma en la casilla ubicada al frente de "*Decisión de Pedidos*"). La decisión que usted toma, será recibida por su proveedor tres semanas después de realizada la orden y serán acumuladas en un canal de órdenes pendientes. La capacidad inicial de producción de su proveedor es de 100 unidades por semana, sin embargo, el tiene la capacidad de cambiar su capacidad según la órdenes que usted le realice, a mas órdenes mayor inversión en capacidad. El tiempo de construcción de capacidad de su proveedor es de una semana. En caso de que su proveedor no tenga la capacidad suficiente para satisfacer sus necesidades, este va a empezar a tener retrasos en las entregas de los pedidos (mayores a 10 semanas) y por lo tanto usted también le incumplirá a sus clientes.

Se incurre en costos cada semana por dos componentes:

1. Costo por Ordenar (CO):

$$CO = \frac{1 * \left(Decisión \;\; Pedidos\right)^2}{1000}$$

Con una cantidad inicial de pedidos de 100 unidades, este costo en la primera semana es de $10.000.

2. Costo por Déficit o Inventario (CD):

$$CD = \frac{2 * \left(Déficit\right)^2}{1000}$$

Con un déficit de 0 unidades, el costo en la primera semana es de $0.

De esta manera, el costo total acumulado CTA es la suma de estos costos en toda la simulación, así:

$$CTA = \sum_{t=1}^{T} \left(CO_t + CD_t\right)$$

Inicialmente, usted ordena 100 unidades por semana, lo cual le permite a su proveedor conservar un tiempo de entrega objetivo de 10 semanas como condición inicial. Recientemente, aplicaciones novedosas del producto crearon un aumento en su demanda. El incremento en la demanda será permanente y del orden de 20 unidades por semana. Dado que usted no estaba atento a estas nuevas aplicaciones, el aumento en la demanda lo tomó por sorpresa y se da cuenta que su déficit está aumentando y por lo tanto perdiendo clientes y prestigio.

Usted iniciara por 3 semanas decidiendo 100 unidades como periodo de aprendizaje. Después su tarea es manejar la compañía durante la simulación, decidiendo cuánto ordenar a su proveedor mientras que minimiza el costo total acumulado CTA.

PAGO: El pago será en efectivo al final del experimento. Corresponde a una suma por participación de $10000 mas una suma variable entre $0 y $30000 en función del CTA, a menor costo total CTA mayor pago.

NOTA: Por favor no divulgar información del experimento con sus compañeros para no perder la validez científica del experimento.

GLOSARIO
(ACERCA DE LOS RESULTADOS QUE SE OBSERVAN EN “REPORTES”)

Sección de Operaciones: Da información del sistema del mayorista (usted).

DEMANDA [unds/sem]:	100	→ 1
DÉFICIT [unds]:	0	→ 2
UNIDADES RECIBIDAS [unds]:	100	→ 3
TIEMPO DE ENTREGA [sem]:	10,00	→ 4
UNIDADES PENDIENTES POR RECIBIR [unds]:	1000	→ 5

1. Son las órdenes que usted recibe de sus clientes finales. Esta es la demanda que se debe cubrir cada periodo.
2. Son las unidades que le faltan por entregar (si es **negativo**, indica **inventario**).
3. Son las unidades que le llegan cada periodo al mayorista (a usted) por parte del proveedor y son con las que dispone por periodo para satisfacer la demanda.
4. Es el tiempo de entrega promedio de las órdenes, contados desde el momento en que usted realiza las órdenes hasta el momento en que las recibe. El tiempo de entrega ideal es de 10 semanas.
5. Acumulan la diferencia entre las órdenes realizadas y las recibidas por el mayorista (usted) en el tiempo. Inicialmente usted tiene una cantidad acumulada de pedidos realizados de 1000 unidades, que serían recibidos de a 100 unidades durante 10 semanas.

Sección de Costos: Da información acerca de cada componente de sus costos:

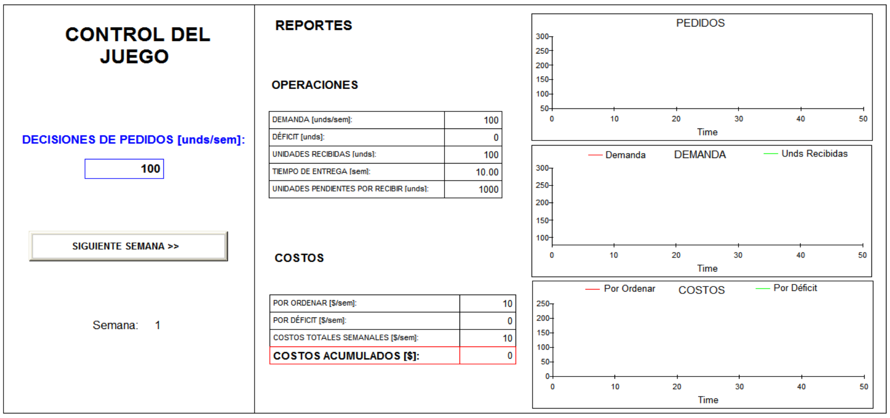
COSTOS

	[$/und]		
POR ORDENAR [$/sem]:	1	10	→ 6
POR DÉFICIT [$/sem]:	2	0	→ 7
COSTOS TOTALES SEMANALES [$/sem]:		10	→ 8
COSTOS ACUMULADOS [$]:		0	→ 9

6. Costos por la decisión tomada cada periodo.
7. Costo por tener inventario o deberle unidades al cliente final ($/semana)
8. Es la suma de los dos componentes de los costos cada semana.
9. Es el costo total acumulado en el tiempo CTA.

Appendix 3. Experiment Enviroment

Appendix 4. Computer Interface (in Spanish)



Appendix 5. The Powersim optimization details

Optimization formulation:

The model equations 1 to 8 reported previously remain valid. Hence, as a **payoff** function we used the variable TC (retailers' total cost), the settings used in this objective function in Powersim Solver are:

Name: TC

Type: Min
Apply time: Stop
Weigth: 1
Divisor: 1

We have created a parameter for each period decision for formulating the decision variables (R_D: Retailers' Orders). Hence each parameter is used just one specific period during the simulation. We also reduce the searching space in the optimization, assigning an upper (500 units/wk) and lower (0 units/wk) bound to these parameters (which are going to be optimized). The settings used in this decision variables in Powersim Solver are:

Name: RDt;   for t: 1, 2… 46
Minimum: 0
Maximum: 500
Apply time: Start

Simulation Settings:
Time unit: period
Time step: 1
Start Time: 0
Stop Time: 46*****
Simulation speed: Maximized speed
Integration: 1st order, Euler
Run count: 1
*The stop time was fixed in a higher number compared with the experiment stop time, because we needed to avoid biased results at the end of the optimization. Hence, we run the optimization with 46 periods, but for the analysis we did not take into account the last 11 times.

Evolutionary Search Method:
The evolutionary search algorithms used by Powersim Solver are licensed from Dr. Nikolaus Hansen. The general settings used for the optimization are:

Maximum generations: 1000
Parents: 20
Offsprings: 100
Minimum convergence: 1e-10

Appendix 6. General description of the Optimization process: Evolutionary search method.
The evolutionary search algorithms used by Powersim Solver are licensed from Dr. Nikolaus Hansen (Powersim Studio 8). However, the general idea of the Evolutionary search method is presented here:
This method, inspired by Darwin's evolutionary theory, is a goal-seeking process where successive runs take place and where the best inputs from a run are used in the next run to generate new inputs to a simulation and try to find the optimum. This evolutionary research method is based on the collective learning process within a population of individuals (each of which represents a search point in the space of potential solution).
In a given time *t* there are *m* representatives of a given scheme H (set of decisions that are taken into account during the selection step) contented in a population.

$$m = m(H,t) \qquad \textbf{(A1)}$$

If *n* is the population size, *f(H)* is the average sample fitness value of the chains that represents the H scheme in t, and $\sum f_j$ is the sum of the aptitudes of all possible chains, the effect of propagation would be given by:

$$m(H,t+1) = m(H,t)\frac{f(H)}{\bar{f}}\text{; where } \bar{f} = \frac{\sum f_j}{n} \qquad \textbf{(A2)}$$

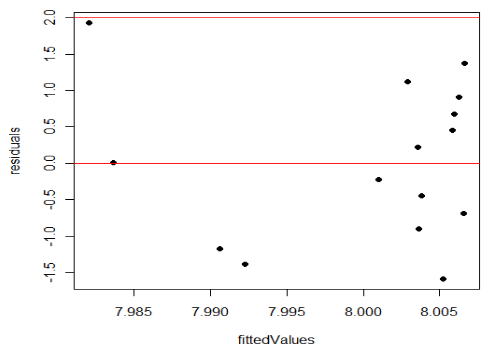
During the calculation of fitness, the evaluation of objective function values is always necessary, such that the information is available and can easily be stored in an appropriate data structure. The population is arbitrarily initialized, and it evolves toward better and better regions of the search space by means of randomized processes of *selection*, *mutation,* and *recombination*. The environment delivers quality information *(fitness value)* about the search points, and the selection process favors those individuals (solutions) of higher fitness to reproduce more often than those of lower fitness (Bäck, 1996).
Including the effects of mutation and recombination we can express the propagation scheme based on the following fundamental theorem:

$$m(H,t+1) \geq m(H,t)\frac{\bar{f}(H)}{\bar{f}}\left[1 - P_c\frac{\partial H}{L-1} - O(H)P_m\right] \qquad \textbf{(A3)}$$

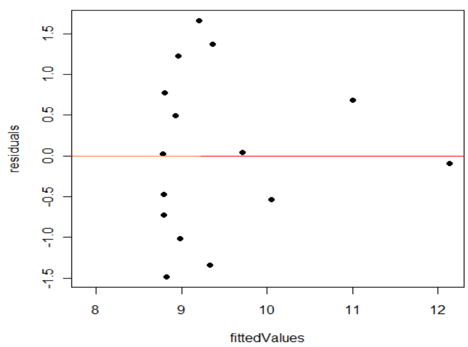
Where P_c is the recombination probability, ∂H is the scheme length, L is the chain length, P_m is the mutation probability and O(H) is the number of fixed positions within a given scheme. The recombination mechanism allows the mixing of parental information while passing it to their descendants, and mutation introduces innovation into the population. Both mutation and recombination can essentially be reduced to local operators that create one individual when applied (Bäck, 1996). The simulation model is run many times with new sets of input values to produce a satisfactory result. The method discards the poorest results, and selects the best results as new scheme. This scheme is used to form offspring for the next generations. An evolutionary search algorithm is repeated until the optimal solution is found, reaches a user-specified limit of generations or until the result is below the specified minimum convergence rate.

Appendix 7. Residual analysis for each subject performance.

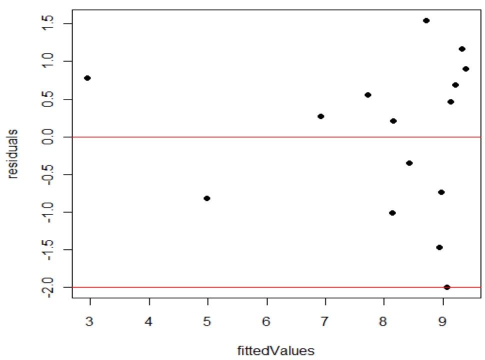
While the experiment was run with more than 15 people per treatment, we removed a few outliers through statistical analysis. The behavior of excluded subjects' deviated markedly from those of other subjects. In particular, some excluded subjects placed a few orders of a dramatic magnitude (e.g., order decision of 9000 units/week, 90times above initial). The removal of these subjects does not change the overall results obtained, but they prevent our statistical analysis to be swayed by these extreme cases. In the presentation of our results, we consider the 60 remaining subjects, 15 per treatment (as shown in table 2). The figures below graphically summarize for each treatment each subject's residuals after econometrically fitting their orders by the proposed heuristic. The graphs for each treatment display the 15 remaining subjects' residuals, between 2 and -2 standard deviations, whose data were used in our statistical analysis.

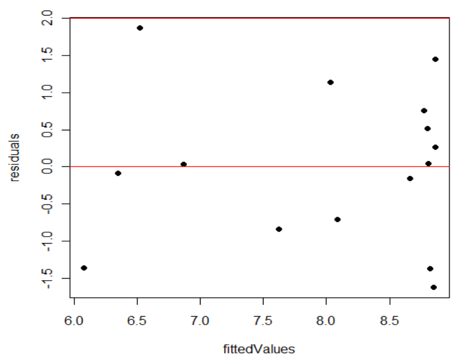


Treatment 1



Treatment 2





Treatment 3 **Treatment 4**

Figure A1. Residual analysis for each subject performance

Appendix 8. Model Behavior

In order to analyze the model behavior, we simulate the model for 30 weeks. Initially, during the first 3 weeks, the model is set in dynamic equilibrium. For this reason, the backlog, retailer's orders and capacity are fixed at their initial values. Thereafter, we introduce a permanent 20% step increase in final customer demand (see Figure A2 – black dashed line) in order to gain intuition about model behavior.

With the anchoring and adjustment heuristic for retailer's ordering behavior the increase in final customer demand leads to an over-reaction of retailers' orders (see Figure A2 – blue dashed line). These retailers' orders exceed supplier capacity and cause an increase in backlog. As backlog increases, retailers experience longer delivery delays, the supplier cannot meet all retailers' orders with the normal delivery delay, and therefore, retailers inflate more their orders. Due to this increase in the backlog, the supplier invests in capacity orders (see Figure A2 – continuous green line) to meet the increase in retailers' orders, but it takes time before the investment in capacity allows the supplier to meet actual demand.

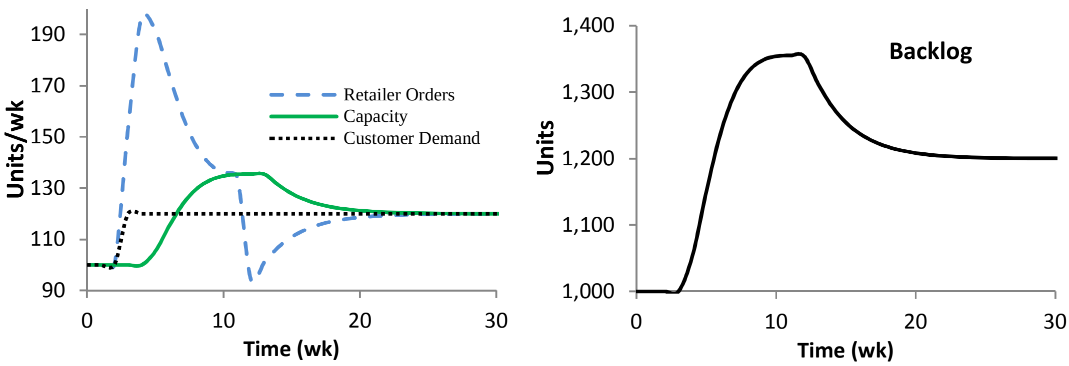


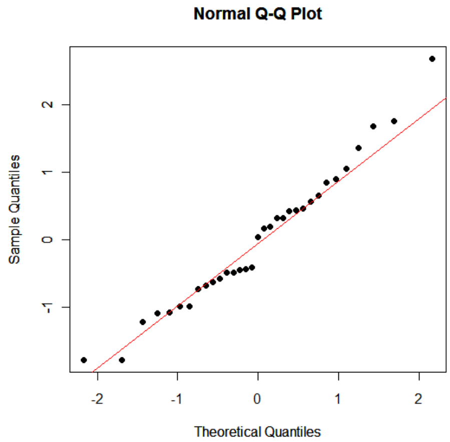
Figure A2. Model behavior (with $\tau_B = 6, \alpha = 1.2, \tau_K = 1$)

When capacity increases high enough such that outgoing shipments equate incoming orders, backlog reaches the maximum value. This retailer orders peaks is mostly influenced for the shipments receive from the supplier, for the long delivery delays faced and for the quantity of orders previously placed. At this point, desired capacity also peaks and capacity investment turns to divestment. As the supplier tries to meet inflated orders from retailers, it dramatically overinvests in capacity. The boom-and-bust in supplier's capacity and backlog represent the reference mode of the system. While final customer demand increases by 20%

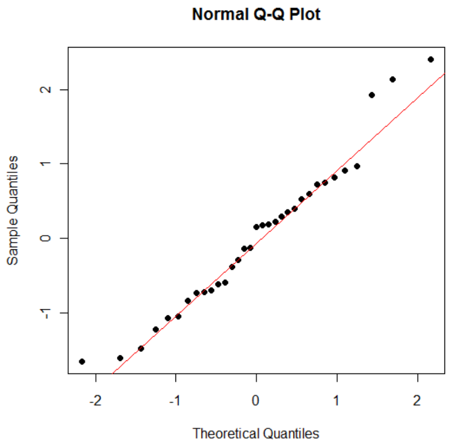
for one year, supplier backlog and capacity increase above 40% to final equilibrium levels (Figure A2).

Appendix 9. Validation of econometrical model assumptions

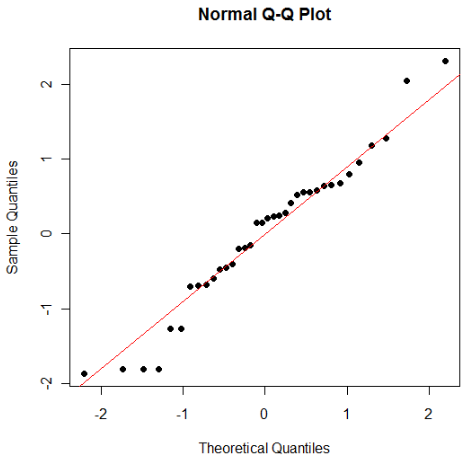
In order to accept the model and the found parameters, we should also assure that the error has normal distribution and constant variance. In the next figure, we use a simple graphical normality test for the residuals. The residual adjust to a straight line, especially in the middle of the graph. So we can say the normality assumption might be satisfied for these data.



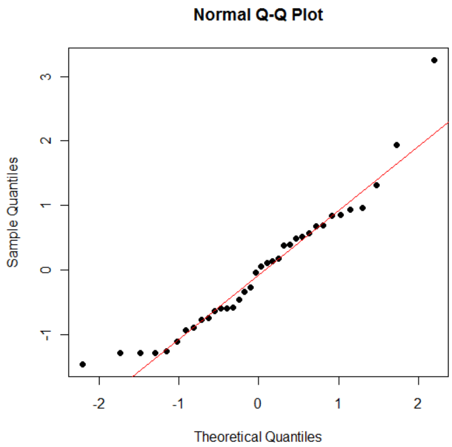
Treatment 1



Treatment 2



Treatment 3



Treatment 4

Figure A3. Normality Analysis per treatment

In the other hand, we use another graphical test in order to analyze the residuals variance. The next figure shows that the points in the plots seem to be fluctuating randomly around zero in an un-patterned fashion. Thus, the plots do not suggest violations of the assumptions of zero means and constant variance of the random errors.

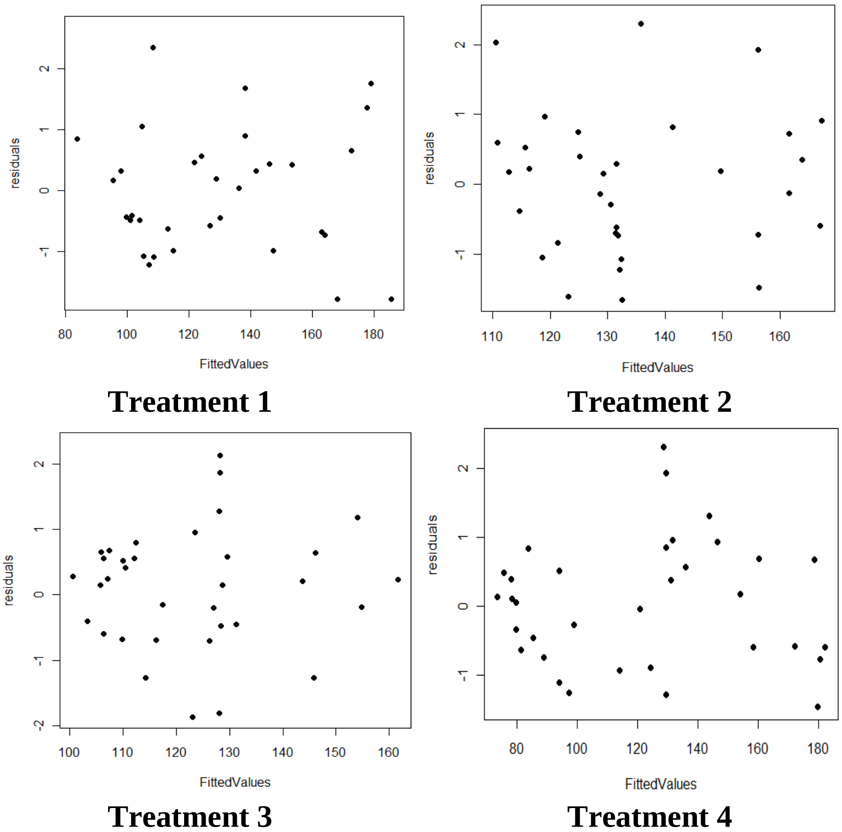


Figure A4. Variance Analysis per treatment

As we show, the parameter estimations are significant and the residuals satisfy all the normality restrictions. In fact, we would like to assure that there is not exist multicolinearity between the variables, because multilinear variables contain the same information about the dependent variable and actually quantify the same phenomenon.

In the next table, we find the largest and smallest Eigenvalues for each data and the ratio between them (condition number). Condition numbers in the table shows numbers less than 100, it suggests there are no serious problems with multicolinearity.

Table A1. Econometrical model selection

	Max Eigen Value	Min Eigen Value	Condition number (Max/min)	AIC Initial	AIC -x1	AIC -x2
T1	1,956	0,044	44,619	196,560	223,880	230,490
T2	1,838	0,162	11,375	170,150	179,180	198,760
T3	1,976	0,024	83,463	203,990	221,530	224,840
T4	1,871	0,129	14,490	231,300	269,640	269,960

Finally, the table also shows the *Akaike information criterion* (AIC), it is a measure of the goodness of fit of a statistical model. The AIC is not a test of the model in the usual sense of hypothesis testing; but, it is a tool for model selection. In the table, we have three values of AIC: 1) The model using x1 and x2, 2) the model without x1 and 3) the model without x2. The best of the three models is chosen according to the lower AIC for each treatment. From the AIC values one may also infer that the top models are roughly using x1 and x2.