Evaluating Overreaction to Backlog as a Behavioral Cause of the Bullwhip Effect

Rogelio Oliva  
*Mays Business School*  
*Texas A&M University*  
Wehner 301F – 4217 TAMU  
College Station, TX 77843  
Tel (979) 862-3744; Fax (979) 845-5653  
roliva@tamu.edu

Paulo Gonçalves  
*School of Business Administration*  
*University of Miami*  
404 Kosar-Epstein Bldg.  
Coral Gables, FL 33124  
Tel (305) 284-8613; Fax (305) 284-2321  
paulog@miami.edu


Do not cite or quote without permission from the authors.
Evaluating Overreaction to Backlog as a Behavioral Cause of the Bullwhip Effect

We evaluate, in an experiment with the Beer Distribution Game, a complementary behavioral source of the bullwhip effect that has been previously ignored in the literature: overreaction to backlogs. By separating the estimation of the response to inventory and backlog, we find that players treat backlog differently than inventory. Contrary to our expectations, players do not over-order when in backlog; instead, they have a measured response, saturating order adjustment and limiting the amount of amplification they introduce in the order stream. This result is robust across echelons, model specifications, and data sets.

By structuring data from 25 games as a panel, we aggregate data across individuals and echelons, and improve the representativeness of the estimated decision rules. Whereas previous research suggests that the supply line is under-accounted for, we find that most players ignore it and that only players facing short and uncapacitated supply lines are capable of taking it into consideration. We also find that inventory adjustment is more aggressive in upstream echelons of the supply chain, where players face higher order variance. Still, players across the supply chain maintain similar levels of desired inventory. We conclude by discussing the implications of these findings for future research in supply chain management.
1. Introduction
The bullwhip effect, the tendency for the variability of orders to increase as one moves from customers to manufacturers, is a frequent and costly problem in supply chains, leading to excessive capital investment, inventory gluts, low capacity utilization and poor service (Armony and Plambeck 2005; Gonçalves 2003; Lee et al. 1997a; Sterman 2000). For instance, Hewlett-Packard lost millions of dollars in unnecessary capacity and excess inventory following a post-shortage demand surge for its LaserJet printers (Lee et al. 1997b). Cisco Systems incurred more than US$ 2 billion inventory write-off due to a strong inventory built up followed by a drastic decrease in retailer orders (Adelman 2001; Armony and Plambeck 2005). The sources for this amplification in demand variability include operational causes–such as batching of orders, order gaming due to shortages, forward buying due to price discounts, and errors in demand forecasting (Lee et al. 1997a)–and behavioral ones–such as failure to account adequately for the supply line of unfilled orders (Sterman 1989) and the adoption of coordination stocks (Croson et al. 2005).

Motivated by Sterman (1989; 1992), a number of experimental studies have used the Beer Distribution Game (BDG) to explore behavioral causes for the bullwhip effect and methods for dampening it (see Croson and Donohue 2002, for a review.) Kaminsky and Simchi-Levi (1998) find that reducing the ordering and shipment lags decrease overall supply chain costs even though order amplification remains the same. Gupta, Steckel and Banerji (2001) and Steckel, Gupta and Banerji (2004) also show that reducing lead times leads to lower costs. Their results suggest, however, that the impact of sharing POS data on costs depends on the nature of customer demand. Using a stationary and known demand (as proposed by Chen and Samroengraja 1999), Croson and Donohue (2003) find that POS data significantly reduces order oscillation–particularly in upstream echelons–and reduces overall supply chain costs. Croson and Donohue (2006) find a similar result when echelons shared inventory information.

Interestingly, the bullwhip effect occurs even when demand is fixed, commonly-known, and players start at the optimal inventory level (Croson et al. 2005). The authors suggest that players build inventory to protect against coordination risk (i.e., the risk that others will deviate from optimal behavior.) Investigating the effect of learning and communication on the bullwhip effect, Wu and Katok (2006) find that order variability decreases when team players are allowed to formulate strategies collaboratively. All these studies, when estimating the ordering decision rule for individuals, arrive at a common source of supply chain instability: players underestimate the supply line of unfilled orders. This work contributes to this line of empirical research by articulating and analyzing a complementary behavioral source of the bullwhip effect that has been overlooked by previous research: overreaction in response to shortages. Overreaction implies that subjects order more aggressively (e.g., have a stronger reaction) when they face shortages than when they hold inventory.

As Mitchell suggests, when competing with other retailers for scarce supplies (i.e., horizontal competition), retailers inflate their orders to manufacturers to improve their chances of obtaining the
Retailers find that there is a shortage of merchandise at their sources of supply. Manufacturers inform them that it is with regret that they are able to fill their orders only to the extent of 80 percent. ... Next season, 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.” (1924, p. 645)

Since there is no horizontal competition in the BDG, we cannot justify overordering as a rational consequence of the rationing game proposed by Lee et al. (1997a). However, we hypothesize that players overreact in response to backlogs motivated by Tversky and Kahneman’s (1974) availability heuristic (i.e., the tendency to overreact to dramatic or vivid events). A backlog is a dramatic event in the beer game because (a) is twice more costly than inventory and (b) causes great disruption to the supply chain.

To increase the efficiency of estimates and the representativeness of the resulting rule, we structure the data as a panel (cross-sectional time series), as opposed to all previous studies that estimate decision rules at the individual level, allowing us to make estimations across individuals and echelons. Our estimated ordering rule provides stronger evidence than previous studies that players underestimate the supply line. Contrary to our expectations, we find that players do not overreact when in backlog, instead their correction saturates at a maximum value; a policy more stable than the linear response to shortfall suggested in previous studies.

The remainder of the paper is structured as follows. In §2 we present the experimental design and methods, in §3 our models and results. We conclude with a summary of our findings and implications for future research in supply chain management.

2. Experimental Design
Our experiment utilizes a web-based version of the Beer Distribution Game (BDG) developed at Harvard Business School that maintains the essential structure of the board game (Sterman 1989). The game represents a serial supply chain with four echelons: retailer, wholesaler, distributor, and factory (R, W, D, and F, respectively). Each supply chain is independent of the other and managed by a team charged with minimizing the supply chain cost. Each echelon incurs an inventory holding cost of $0.50 per unit/week and a backlog cost of $1.00 per unit/week. Shipment and order delays between echelons are two weeks and factories incur a one-week production delay with no capacity constraints. Each simulated week players face the following sequence of events: (1) receive shipments; (2) fill customer orders, if sufficient inventory is available, otherwise accumulate a backlog; and (3) place an order with its supplier.

The game is initialized in flow equilibrium: order and shipment flows are 4 units/week and each echelon starts with an initial inventory level of 12 units. Subjects are not informed about the shape of demand. A single time increase in retailer orders (a step input) is introduced in the second period (week), bringing orders to 8 units/week. To avoid end-of-horizon behavior the experiment is announced to run for a simulated year, but is, in fact, terminated after 36 weeks. The web-based version, by virtue of its automatic computation of order receipts, incoming orders, shipments, and inventory-backlog levels, can
be run with less time pressure than the board version of the game on which a facilitator imposes the pace. The automatic recording of transactional data avoids reporting errors, although data entry (i.e., “typing”) errors are still possible. Because of the cascading effects to other players, we did not attempt to correct typing errors.

Our data set consists of a sample of 116 pairs of first-year Harvard MBA students (class of 2004) that played the game as part of the introductory course in operations management. The students were on average 27 years old and had about two years of work experience in diverse areas. Prior to the game, players received a five-page document describing the structure of game and the sequence of events players will be facing. Less than two percent of the players expressed prior knowledge of the game. Due to the reduced number of “experts,” they were not excluded from our sample. Players were randomly assigned, in pairs, to echelons (R, W, D, or F) and teams. Team members interacted via a computer screen and, in contrast to the board version of the game, lacked both visual access to the state of the supply line and knowledge of who other teammates were. We eliminated four games from our original sample. These games included one or more players showing anomalous ordering behavior (consistently not ordering when in backlog or placing high orders when holding large inventory) suggesting they had misunderstood the stock management task. Our analysis is based on the remaining 25 games.

2.1. Methods
We treated the BDG’s non-negativity constraint on orders as censored data. That is, we assumed that an order for zero could represent situations in which a subject wished to cancel a previously placed order (a negative order) but was restricted by the rules of the game to a minimum order of zero. Accordingly, we estimated the model using a tobit model (Tobin 1958). Also, to estimate a decision rule that reflects the full range of observations available, we structured the data from the games as a panel (cross-sectional time-series data set) with individual players the cross-sectional unit (i) and week of decision the time index (t). There being no reason to suspect that individual differences can be captured by changes in the constant term, and subjects being clearly a sample from a larger population, we assumed random effects across individuals (Greene 1997).

3. Estimation of ordering policies
In a setting similar to the BDG with stationary and commonly-know demand distribution Chen (1999) demonstrated that a base stock policy—where orders placed equal those received—minimizes total supply chain cost and avoids demand amplification. Since the demand in our experiment is both non-stationary and unknown to players (a step increase), it is not possible for players to calculate an optimal strategy prior to the game. While there is no reason to expect a base stock policy to be optimal, it still provides a very reasonable ordering policy for players. Therefore, we test the base stock policy as a base model to estimate players’ ordering policies and incrementally change it to include heuristic that players may use up to testing the hypothesis of overreaction to backlog.
### 3.1 Base stock policy

A model that estimates the ability of a base stock policy to explain the variability in orders is given by:

$$O_{it} = \text{MAX}(0, \beta_L L_{it-1} + u_i + \varepsilon_{it})$$  \hspace{1cm} (I)

where, to be consistent with the BDG, orders are constrained to be nonnegative; $L_{it-1}$ represents orders received by the $i$th subject in the last period ($t-1$); $u_i$ is the random disturbance characterizing the $i$th subject, and $\varepsilon_{it}$ is an additive disturbance term. According to Chen’s (1999) base stock policy, orders placed ($O_{it}$) must equal those received in the previous period ($L_{t-1}$), thus we expect $\beta_L$ to be equal to one.

Model I in Table 1 shows the estimated parameters for the base model together with the model’s log-likelihood value, significance ($\chi^2$), $R^2$, and root mean percent error. The model is highly significant ($p<0.001$), explains 54% of the variance in orders, and differences among players do not contribute to explain unexplained variance in orders ($r^2_{su2}+s_{e2} = 0.0$). While a base stock policy (using a lag forecast as a predictor for orders) provides a good fit for players’ ordering policy, the fact that the $\beta_L$ coefficient is slightly greater than 1 ($p=0.03$ for $H_0 : \beta_L = 1$) suggests that the rule does not fully capture all the adjustments being made by the players.

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
<th>Model IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_L$ Expected Loss ($L_{t-1}$)</td>
<td>1.02</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_S$ Inventory/Backlog ($S_t$)</td>
<td>-0.11</td>
<td>-0.21</td>
<td>-0.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01)**</td>
<td>(0.01)**</td>
<td>(0.01)**</td>
<td></td>
</tr>
<tr>
<td>$\beta_B$ Backlog ($S_tB_t$)</td>
<td>0.21</td>
<td>0.21</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)**</td>
<td>(0.02)**</td>
<td>(0.02)**</td>
<td></td>
</tr>
<tr>
<td>$\beta_{SL}$ Supply line ($SL_t$)</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>$\beta_0$ Constant</td>
<td>0.33</td>
<td>1.69</td>
<td>1.60</td>
<td>1.60</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.17)**</td>
<td>(0.25)**</td>
<td></td>
</tr>
</tbody>
</table>

| Log-likelihood value            | -9769.9 | -9535.8  | -9426.0   | -9425.8  |
| Wald $\chi^2$                   | 10395.0*** | 398.4*** | 562.5***  | 559.1*** |
| $R^2$                           | 0.54    | 0.58     | 0.60      | 0.60     |
| RMSE                            | 4.83    | 4.60     | 4.46      | 4.46     |
| $\rho$                          | 0.00    | 0.02     | 0.04      | 0.04     |
|                                  | (0.00)  | (0.01)*  | (0.01)**  | (0.01)** |
| Backlog effect ($\beta_S + \beta_B$) | 0.00    | 0.00     | 0.00      | 0.00     |
| $p(\beta_{SL} + \beta_B=0)$    | 0.92    | 0.73     |           |          |
| Observations                    | 3500    | 3500     | 3500      | 3500     |
| Censored ($O=0$)                | 520     | 520      | 520       | 520      |
| Number of players               | 100     | 100      | 100       | 100      |

Standard errors (SE) in parentheses: * significant at 10%; ** significant at 1%; *** significant at 0.1%.

(a) $R^2=r^2$, where $r$ is the simple correlation between estimated and actual orders (Wooldridge 2002).

† We found some evidence of autocorrelation in the residuals ($p=0.08$). Tests on the linear model (i.e., without the non-negativity constraint), however, indicated that, for all regressions, the underestimation on the SE resulting from the autocorrelation was not large enough to affect the reported significance of the estimates (in most cases the
standard error of the estimates from the linear model matched that of the tobit model, and, when adjusting for
autocorrelation in the residuals in the linear model, the change in the SE of the estimates was in the third significant
digit). Since estimates with autocorrelation are unbiased, the reported results can be safely interpreted.

Table 1. Estimation results

3.2 Stock management policy
We revise the base stock policy to incorporate inventory and supply line adjustments suggested by
Sterman (1989). Due to the difficulty in finding the optimal ordering policy in the traditional BDG,
Sterman proposes a simple, self-correcting ordering heuristic that uses information locally available to the
decision maker and presumes no knowledge of the structure of the system. Specifically, managers are
assumed to size orders to (1) replace expected losses from stock, (2) reduce the discrepancy between
desired and actual stock, and (3) maintain an adequate supply line of unfilled orders. The decision rule is
formalized as:

\[ O_t = \hat{L}_t + \alpha_S (S^* - S_t) + \alpha_{SL} (SL^* - SL_t) \] (1)

where \( \hat{L}_t \) represents the expected loss from the stock, \( S_t \) and \( SL_t \) the inventory and supply line positions
at time \( t \), \( S^* \) and \( SL^* \) the desired levels for stock and supply line, and the parameters \( \alpha_S \) and \( \alpha_{SL} \) the
fractional adjustment rate for inventory and supply line, respectively.

Sterman (1989) assumed adaptive expectations for the formation of the expected loss according to the
exponential smoothing equation

\[ \hat{L}_t = \theta \hat{L}_{t-1} + (1 - \theta) \hat{L}_{t-1} \] (2)

and obtained, for each player, maximum likelihood estimates for the simultaneous equations 1 and 2,
subject to the constraints \( 0 \leq \theta \leq 1 \) and \( \alpha_S, \alpha_{SL}, S^*, SL^* \geq 0 \). The joint estimation of these equations,
however, has the potential of shifting variance between the stock replenishment and forecasting
equations, eqs. 1 and 2 respectively. Lower values of \( \theta \) make the forecast series more stable and shift the
residual variance to the replenishment decision, thus potentially biasing its parameter estimates (Oliva
2003).

We assume a simple lag forecast \( (\hat{L}_t = L_{t-1}) \), an implied \( \theta = 1 \) in the exponential smoothing model in
equation 2, and an intuitive and plausible model of expectation formation (Kleinmuntz 1993). The simple
lag forecast assumption has been used before in empirical research with the BDG (Steckel et al. 2004);
generates a reasonable forecast (0% Median Absolute Percent Error (MdAPE) and 23% Mean Absolute
Percent Error (MAPE)) for a series with a coefficient of variation of 0.61; and has a Root Mean Square
Error (RMSE) only 9.5% higher than the optimal exponential smoothing forecast. The change in
forecast, the introduction of the non-negativity constraint on orders, and the expansions for panel data and
additive disturbances yields the model:

\[ O_{it} = \text{MAX}(0, L_{it-1} + \beta_0 + \beta_S S_{it} + \beta_{SL} SL_{it} + u_i + \epsilon_{it}) \] (II)

where \( \beta_0 = \alpha_S S^* + \alpha_{SL} SL^* \), \( \beta_S = -\alpha_S \) and \( \beta_{SL} = -\alpha_{SL} \) in equation 1.

The results of the estimation of model II are presented in Table 1. The model is highly significant and
explains 58% of the order variance, a 4% improvement over the base stock model. In this case 2% of the unexplained variance in orders is explained by the differences among individuals ($\rho = \sigma_s^2 / (\sigma_s^2 + \sigma^2_e)$) and this panel-level variance is marginally significant when compared to the pooled tobit model (likelihood-ratio test of $\sigma_s^2 = 0$). The estimated fractional inventory adjustment ($\alpha_s = -\beta_s = 0.11$) is highly significant and has the expected sign. The estimated fractional adjustment to the supply line ($\alpha_{SL} = -\beta_{SL} = 0.01$), on the other hand, is not statistically significant, suggesting that players ignore the supply line. The dampened responsiveness of the estimated adjustment policies when compared to Sterman’s (1989) previous estimations of the stock management model for individual players ($\alpha_s = 0.26$ and $\alpha_{SL} = 0.09$) suggests that our assumption of the simple lag-forecast is leaving less variance for the stock adjustment policies to explain.

3.3 Response to backlog policy

Previous estimations of ordering policies (Croson and Donohue 2006; Sterman 1989) treat backlog as negative inventory and assume a linear response to the gap between current and desired inventory. Because the cost of backlog is twice the holding cost for inventory, it is possible that subjects reacted differently to backlog than to excess inventory. To test, with the simplest model possible, for the possibility of a different reaction to backlog, we assumed a piecewise linear model (Pindyck and Rubinfeld 1998), introducing a dummy variable ($B_t$) to reflect the backlog condition ($B_t = 1$ if $S_t < 0$; 0 otherwise). Accordingly, the response to inventory is modified to $\alpha_s (S^* - S_t) + \alpha_B S_t B_t$, where $\alpha_s$ represents the fractional adjustment rate for the inventory and $\alpha_B$ the incremental adjustment due to backlog (i.e., the response to backlog is $-\alpha_s + \alpha_B$). Since $S_t B_t \leq 0$, $\alpha_B < 0$ indicates a stronger reaction to the backlog condition (see Figure 1 for expected response to the inventory position and reaction to shortages). Retaining the assumption of a simple lag as the expected loss from the stock, the introduction of the non-negativity constraint on orders and the expansions for panel data and additive disturbances lead to the following model:

$$O_{it} = \text{MAX}(0, L_{it-1} + \beta_0 + \beta_S S_{it} + \beta_B S_{it} B_{it} + u_i + \epsilon_{it})$$

(III)

where $\beta_0 = \alpha_s S^*$, $\beta_S = -\alpha_s$, and $\beta_B = \alpha_B$.

![Figure 1. Order response to the inventory/backlog position](image)

The estimated model is highly significant (see model III in table 1). The introduction of the backlog
response explains another 2% of the order variance, increasing $R^2$ to 0.60, and the estimated fractional inventory adjustment ($\alpha_s = -\beta_s = 0.21$) is significantly higher (i.e., more aggressive) than the estimate found without the backlog adjustment (model II), suggesting that indeed players treat inventory differently than backlog. We were surprised, however, to find a positive coefficient for the backlog response ($\beta_B = 0.21$). The combined response to a situation when inventory is in backlog ($\beta_s + \beta_B$) is not significantly different from zero (see Backlog effect and its test in Table 1). This suggests that players in backlog place orders equal to the expected loss plus a constant amount ($\beta_0$) proportional the desired inventory level $S^*$ (see Figure 2 for a schematic of the order response to the estimated model). Instead of “over-reacting” to a backlog situation as we had expected, players seem to ignore the backlog information cue and respond only to the $S_t = 0$ signal.

\[ \beta_s + \beta_B = 0 \]

\[ S_t = \beta_0 \]

\[ L_{t-1} = \beta \]

\[ \beta_s + \beta_B = 0 \]

\[ S_t = \beta_0 + \beta_{L_t} \]

\[ S_t = \beta_0 / \beta_s \]

\[ S_t = \beta_0 + \beta_{L_t} / \beta_s \]

**Figure 2. Estimated model’s response to the inventory/backlog position**

When the sample is split by echelon (see R, W, D, and F models in Table 2), model III is significant for all positions and all the estimates have consistent signs and are significant (the model with dummies for each position yields the same results as Model III and the coefficients for the three dummies are not significant). The seemingly paradoxical result that both the $R^2$ and RMSE of the models increase as we move up the supply chain is explained by the fact that each successive stage the supply chain faces a demand stream with higher variance (see $\sigma_{L_{t-1}}$ row in Table 3) and, although the models explain a higher fraction of that variance ($R^2$), the magnitude of the errors (RMSE) is increasing. In terms of parameter estimates, as a result of the differences in variance in demand each echelon faces, the aggressiveness of the inventory fractional adjustment ($\alpha_s = -\beta_s$) increases as we move up the supply chain: the higher the variance of $L_{t-1}$, the more aggressive corrections to deviations in inventory. Parameter estimates for the three first echelons (R, W and D), however, are not statistically different and the pooled model for these non-factory echelons (~F in Table 2) yields more efficient estimates, all within the standard error of the estimates for the separate models. Factories, whose stock management problem is structurally different from that of the other echelons—their delivery delay is shorter, and, because they are uncapacitated, it remains constant—have a significantly different replenishment rule with fractional adjustments to inventory more than twice as aggressive as the other echelons. The decision rule is, however, consistent across echelons as the other parameters adjust to accommodate the required aggressiveness of the
fractional inventory adjustment—estimated values for desired inventory ($S^*=\beta_0-\beta_S$) for each sample partition are not statistically different (see Table 3) and are consistent with one week of orders at the increased consumption rate—and maintain an almost-flat response to the backlog condition ($\beta_5+\beta_{SL}=0$).

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Model III</th>
<th>Model IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full</td>
<td>R</td>
</tr>
<tr>
<td>$\beta_S$ Inventory/Backlog ($S_t$)</td>
<td>-0.21</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td>(0.01)***</td>
<td>(0.01)***</td>
</tr>
<tr>
<td>$\beta_B$ Backlog ($S_t, B_t$)</td>
<td>0.21</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>(0.02)***</td>
<td>(0.03)***</td>
</tr>
<tr>
<td>$\beta_{SL}$ Supply line ($SL_t$)</td>
<td>1.69</td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td>(0.17)***</td>
<td>(0.29)***</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood value</td>
<td>-9426.0</td>
<td>-2161.7</td>
</tr>
<tr>
<td>Wald $\chi^2$</td>
<td>562.5***</td>
<td>110.7***</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.60</td>
<td>0.16</td>
</tr>
<tr>
<td>RMSE</td>
<td>4.46</td>
<td>2.94</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>(0.01)**</td>
<td>(0.04)**</td>
</tr>
<tr>
<td>Backlog effect</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>0.92</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>3500</td>
<td>875</td>
</tr>
<tr>
<td>Censored ($O_t=0$)</td>
<td>520</td>
<td>30</td>
</tr>
<tr>
<td>Number of players</td>
<td>100</td>
<td>25</td>
</tr>
</tbody>
</table>

Standard errors in parentheses: * significant at 10%; ** significant at 1%; *** significant at 0.1%.

$R^2=r^2$, where $r$ is the simple correlation between estimated and actual orders (Wooldridge 2002).

† See note on Table 1.

Table 2. Estimation results†

To test the robustness of the lack of response to backlog we tested each of the assumptions made on our model specifications. First, we tested the forecasting assumption and replaced the simple-lag forecast with static expectations ($\hat{L}_{it}=L^*$), extrapolative expectations ($\hat{L}_{it}=L_{t-1}+\gamma(L_{t-1}-L_{t-2})$, where $0\leq\gamma\leq1$), and optimal adaptive expectations ($\hat{L}_{it}=\theta_iL_{it-1}+(1-\theta_i)\hat{L}_{it-1}$, where $\theta_i$ minimizes the forecast error for each player). The almost-flat response to the backlog condition held under all forecasting assumptions. Second, we tested the non-linear shape of the response to inventory-backlog by testing different breakpoints for the piecewise linear model (the best fit was obtained when breakpoint is at $S=0$); introducing separate intercepts to the two line segments to decouple the inventory to the backlog response (the second intercept was not significant); and testing different non-linear continuous responses (e.g., quadratic and logistic models)—all models generated a flat response to the backlog condition and none
was as intuitive as model III presented above. Third, we tested the censored data assumption and ran the model without the tobit constraint and found no significant change to the backlog response. Fourth, we tested the aggregation assumption and ran the model, first, as a pooled data set (as expected from the small value of \( \rho \), there was no significant change to the estimates), and then for each player independently: 84% of the players showed underreaction to backlog, and the model was significant for 67% of those players. Finally, we tested the model with Sterman’s (1989) data and an additional data sample and found similar results, ruling out explanations differences from our data collection methods. The next section explores the impact of the supply line on the almost-flat response to the backlog condition.

### Table 3. Demand variance and estimated parameters—base model and by echelon

<table>
<thead>
<tr>
<th></th>
<th>Full</th>
<th>R</th>
<th>W</th>
<th>D</th>
<th>F</th>
<th>~F</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L_{i-1} )</td>
<td>7.59</td>
<td>7.56</td>
<td>7.68</td>
<td>7.78</td>
<td>7.76</td>
<td>7.67</td>
</tr>
<tr>
<td>( \sigma_{L_{i-1}} )</td>
<td>4.78</td>
<td>1.26</td>
<td>3.23</td>
<td>5.11</td>
<td>7.30</td>
<td>3.56</td>
</tr>
<tr>
<td>( S^* = \beta_{0i}/\beta_S )</td>
<td>7.94</td>
<td>9.44</td>
<td>9.63</td>
<td>8.12</td>
<td>7.56</td>
<td>8.57</td>
</tr>
<tr>
<td>S.E.</td>
<td>0.67</td>
<td>1.65</td>
<td>1.21</td>
<td>1.91</td>
<td>1.19</td>
<td>0.83</td>
</tr>
</tbody>
</table>

(1) Calculations based on the “delta method” (Oehlert 1992) values might differ from calculations based on coefficients from Table 2 because of rounding.

3.4 Response to supply line

Model IV (tables 1 and 2) shows the estimated parameters and performance statistics for a decision rule that incorporates, in addition to the inventory and backlog adjustments, the supply line signal:

\[
O_{it} = \text{MAX}(0, L_{i-1} + \beta_0 + \beta_S S_{it} + \beta_B B_{it} + \beta_{SL} SL_{it} + u_{it} + \epsilon_{it}) \quad (IV)
\]

For the full sample, the coefficient for the supply line is not significant and introducing it into the regression has no effect in the other estimates. When we split the sample in factories and non-factories, the \( \beta_{SL} \) coefficient becomes significant, taking the expected negative value for factories but a positive value for non-factories. Nevertheless, the introduction of the supply line as a regressor does not have a significant impact in the inventory and backlog adjustment fractions for the two sub-samples, and the models’ overall performance does not improve. When estimating model IV for individual players only 21% of the non-factories and 52% of the factories showed a significant negative response to the supply line. These results indicate that players consistently ignore the supply line, a finding stronger than Sterman’s (1989) that players underestimate the supply line.

4. Discussion

We explored, in an experimental serial supply chain, the causes of the bullwhip effect by proposing a complementary behavioral source of order amplification in supply chains: overreaction to backlogs. The paper contains several contributions relative to previous work in this area. Specifically, refinements in assumptions and estimating techniques shed light on important aspects of the estimated decision rules for BDG players. First, by structuring the data as a panel (cross-sectional time series), we use all the data available for estimating the replenishment decision rule, thereby increasing the efficiency of estimates and the representativeness of the resulting rule. Our estimated ordering policies compare well to the
policies suggested by previous studies with the BDG that estimate parameters for individual players (Croson and Donohue 2003; 2006; Sterman 1989). While Sterman’s (1989) study yielded average R² and RMSE of 0.71 and 2.86 respectively, it required four parameters per player (i.e., 176 parameters for 44 players) to achieve this result. Our most parsimonious model achieves an R² of 0.60 and an RMSE of 4.46 with only three parameters for 100 players, confirming that it is possible to make some inferences across individuals. The panel data structure also allowed us to perform analyses by echelon in the simulated supply chain. We found that the inventory adjustment fraction became more aggressive as players faced increasing order variance (e.g., upstream in the supply chain) but they did not change their desired inventory levels.

Second, by decoupling the estimation of the forecast rule from the stock replenishment rule, and removing constraints in the feasible space for all parameters, we found a weaker inventory correction response and even stronger evidence that players underestimate the supply line. Whereas previous research suggests that the supply line is under-accounted for, we only found a significant effect of the supply line in the decision rule for factories, which face shorter and uncapacitated supply lines.

Regarding our core hypothesis, we found that players treated backlog and inventory differently. Contrary to our expectations, however, players do not over-order when in backlog; instead, they have a measured response, saturating order adjustment at a maximum value and limiting the amount of amplification they introduce in the order stream. This result held across echelons and is robust across model specifications and data sets.

Our findings have several implications for future research on ordering policies for supply chains. Our main finding, that decision-makers treat inventory and backlog differently, suggests a revision to behavioral models for replenishment decisions to account explicitly for the backlog condition. Clearly, our finding of under-ordering when in backlog needs to be further explored since it indicates a larger departure from optimal performance. Among the potential explanations of this result are the small asymmetry in inventory and backlog costs used in the BDG and expectation adjustments that players could be making when faced with consistent under-delivery from suppliers. Finally, we introduce a new framework and process for the estimation of the decision rules that makes use of all the available data and allows for generalizations across individuals and comparisons among games and echelons. We believe that use of these tools can further the insights derived from empirical work in operations management.

Acknowledgments
We are grateful to John Sterman, Benito Flores, Alan Love, Arzu Onar, two anonymous reviewers, and participants of the INFORMS, POMS, and System Dynamics Conference, for valuable feedback on earlier versions of this work, and to Lisa Noonan for her help running the games and capturing the data.

References
Arony, M., E. Plambeck. 2005. The Impact of Duplicate Orders on Demand Estimation and Capacity


Hill, New York.


