Destocking, the bullwhip effect, and the credit crisis: Empirical modeling of supply chain dynamics

Maximiliano Udenio [a,*], Jan C. Fransoo [a], Robert Peels [b]

[a] School of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Den Dolech 2, 5612AZ Eindhoven, The Netherlands
[b] Flostock BV, Horsten 1, 5612AX Eindhoven, The Netherlands

ABSTRACT

In this paper we analyze the strong sales dip observed in the manufacturing industry at the end of 2008, following the bankruptcy of Lehman Brothers and the subsequent collapse of the financial world. We suggest that firms' desire to retain liquidity during these times prompted a reaction characterized by the reduction of working capital, which materialized as a synchronized reduction in target inventory levels across industries. We hypothesize that such a reaction effectively acted as an endogenous shock to supply chains, ultimately resulting in the bullwhip-effect kind of demand dynamics observed. To test this proposition we develop a system dynamics model that explicitly takes into account structural, operational, and behavioral parameters of supply chains aggregated at an echelon level. We calibrate the model for use in 4 different business units of a major chemical company in the Netherlands, all situated 4–5 levels upstream from consumer demands in their respective supply chains. We show that the model gives a very good historical fit of the sales developments during the period following the Lehman collapse. We test the model's robustness to behavioral parameter estimation errors through sensitivity analysis, and the de-stocking hypothesis against an alternative model. Finally, we observe that the empirical data is aligned with experimental observations regarding human behavioral mechanisms concerning target adjustment times.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The world economy experienced a severe, sudden, and synchronized collapse in late 2008. The magnitude of the drop in global trade was the largest since World War II, it was the steepest in recorded history, and it was synchronized: all 104 nations where data is collected by the WTO experienced a drop in imports and exports during the second half of the year (Baldwin, 2009). Following the public collapse of the financial system (starting with the Lehman Brothers bankruptcy in September 2008), firms all over the world observed substantial demand disruptions; sales plummeted across the board, and panic spread. While many consumer markets remained relatively stable (exceptions being consumer durables and capital goods), the manufacturing sector observed almost instantaneous demand drops (Dooley et al., 2010).

In crises such as these, managers are pressured to improve the financial position of the company at the same time that demand levels are dropping dramatically. This typically leads to strategic decisions such as reducing inventories (to reduce the level of working capital), downsizing (to reduce operational expenses), and closing manufacturing facilities (to reduce fixed assets). These decisions, however, have substantial operational consequences when demand increases at a later stage: the reduction of inventory levels, workforce, and manufacturing facilities are decisions that require significant time to be reversed. If the situation that triggered such decisions is temporary and demand recovers faster than the speed at which firms can react, lost sales and general problems with inventory management will appear. Knowledge about the underlying dynamics behind the demand slump is therefore needed to avoid costly mistakes.

These underlying operational dynamics are a focus of extensive study as part of the systems-thinking approach introduced by Forrester (1958). This approach centers on the use of System Dynamics as the preferred methodology to replicate and understand the dynamic behavior of complex systems. System Dynamics models explicitly simulate the behavior of individual components pursuing local results, and exploit the structure of the system to model the interactions between these components. In doing so, System Dynamics allows the modeler to decouple endogenous, exogenous, and structural effects.

With regard to supply chain dynamics, observations are generally made that (a) production variance tends to be greater than

* Correspondence to: Pavijloen F6, Postbus 513, 5600MB Eindhoven, The Netherlands. Tel.: +31 40 247 5042.
E-mail addresses: m.udenio@tue.nl (M. Udenio), j.c.fransoo@tue.nl (J.C. Fransoo), robert.peels@flostock.com (R. Peels).

demand variance, and (b) that this difference increases the further upstream a firm is. This has the effect of greatly amplifying demand fluctuations through a supply chain, and has been termed 'the bullwhip effect' (Lee et al., 1997a). Analytical studies quantify this effect (Chen et al., 2000); empirical studies show evidence of its existence at a firm level (Metters, 1997; Fransoo and Wouters, 2000; Bray and Mendelson, 2012); and a substantial experimental body of work investigates its causes and possible solutions (Sterman, 1989; Croson and Donohue, 2006). Empirical evidence of the bullwhip effect at higher aggregation levels is, however, ambiguous: conclusive evidence of neither variance amplification nor production smoothing has been found in public manufacturing data (see Cachon et al., 2007, for a study based on U.S data). This apparent incompatibility between the predictions of the theory—supported by experimentation– and high-level observations is, however, explained by the effects of data aggregation. Chen and Lee (2012) show that both product aggregation (whereby multiple items are grouped into categories), and temporal aggregation (whereby information is grouped into quarters) mask the magnitude of the bullwhip effect.

In this paper, we argue that firms reacted to the 2008 financial crisis by reducing their working capital targets and, because it was global and synchronized, this reaction introduced a significant shock in the world's supply chains—essentially creating an inventory-driven bullwhip effect. To test our hypothesis, we adopt supply chain modeling, experimentation, and validation methods based on theory from the experimental work by Sterman (1989) and Croson and Donohue (2006)—originally focused on the appearance of the bullwhip effect following demand shocks in a laboratory setting. We develop 4 different supply chain models for a major chemical company in the Netherlands and validate them with demand data from the crisis period. In terms of methodology, our work distinguishes itself from previous studies on inventory dynamics by using extensive empirical data, framing the Lehman Brothers collapse as a natural experiment. We specifically distinguish between the direct estimation of the operational model parameters, such as lead times, and the econometric fitting of behavioral parameters, such as stock adjustment times. In terms of theory, we model aggregates of companies at a particular level of the supply chain in a particular region rather than individual decision makers (as is common in experiments) or firms (as is common in much of the system dynamics literature in supply chain management). The crisis time-frame, through the resulting synchronization in managerial objectives, gives us the opportunity to link aggregate and individual human behaviors.

We show that the combination of declining end-markets and the appearance of a synchronized inventory shock successfully account for a significant portion of the observed long and short term dynamics. Moreover, to increase our confidence in the de-stocking hypothesis, we present an alternative model without the explicit inventory adjustment reaction to the crisis. Our results show that demand drops in the respective end markets were not severe enough to explain by themselves the wild dynamics observed upstream.

In this view, exogenous end-markets drive the overall long-term evolution of sales, while endogenous behavior (such as the inventory decisions taken as a consequence of the crisis) primarily impacts the short term dynamics.

The contribution of this paper to the theory is thus threefold: (1) We identify the 2008 financial crisis as a natural experiment that effectively controls for the masking effects of aggregation. This allows for the usage of a system dynamics framework based on the bullwhip effect literature whereupon we model aggregate echelons. (2) We introduce a de-stocking hypothesis capable of explaining the demand evolution observed by upstream companies following the bankruptcy of Lehman Brothers. (3) We identify the importance of both consumer end-markets and ordering behavior in the evolution of demand patterns through time. By explicitly modeling separate structural, operational, and behavioral parameters, this study quantifies their contribution to the observed transient behavior and allows for a comparison with results obtained from experimental studies on individual human decision making.

By explicitly modeling the impact of the sudden reduction of inventory targets throughout the supply chain, we highlight the impact that locally rational policies can have on overall supply chain performance. From a managerial perspective, we display the value of supply chain models that propagate end-market, and endogenous, dynamics up a supply chain. Whereas an upstream firm cannot avoid the bullwhip-like dynamics that follow shocks of the magnitude of those observed after the onset of the 2008 financial crisis, it can use turning-point forecasts to support strategic decisions.

The remainder of this paper is organized as follows: In Section 2 we introduce several inventory puzzles present in the economics literature, use these to identify the challenges inherent in the study of inventories as part of aggregate models, and develop our de-stocking hypothesis. Section 3 introduces the methodology and model formulation. We extend prior experimental work and frame our models in the crisis time-period by explicitly modeling the managerial decisions behind the hypothesized reduction of inventory targets at an echelon level. In Section 4, we extend the echelon models to four different supply chains, and use empirical data to calibrate and validate these. We then formulate alternative models—without the de-stocking hypothesis—to study the appropriateness of this hypothesis. We conclude in Section 5 with a series of managerial insights.

2. Background and hypothesis development

When looking at the link between inventories and macro economic developments, Blinder and Maccini (1991) point out that interest in inventory behavior seems to follow cycles, not unlike the economy we attempt to explain. Indeed, we observe that research on the role of inventories in the economy peaks throughout history following extraordinary economic happenings such as the post-war period, the late seventies oil crisis, and— relevant to current developments—the financial crisis of 2008.

We refer the reader to Fitzgerald (1997) and Blinder and Maccini (1991) for comprehensive reviews of over 50 years of discussions on inventory theory in the economics discipline and the puzzles they attempt to solve. In his work, Fitzgerald (1997) identifies inconsistencies between theory and data, and the subsequent attempts of researchers to eliminate these discrepancies from their models. Blinder and Maccini (1991) summarize the opposing views of micro and macro economists with regard to the role of inventories: the former discipline sees them as a stabilizing factor, whereas the latter sees them as a de-stabilizing one. Despite these fundamental disagreements, Feldstein and Auerbach (1976) point out, inventory fluctuations have long been recognized as a major endogenous force in American business cycles. In their experience, irrespective of the conceptual contradictions between contemporary models and the real-life processes behind them, most studies of inventory behavior note that about 75% of the cyclical downturn in gross national product (from peak to trough) can be accounted for by the reduction of business inventories. Recognizing these conceptual difficulties, Lovell (1994) reflects upon the inherent challenge of trying to reconcile these views. He poses a series of questions that—for all the body of research available—remain open to this day: "(…) Do firms actually

attempt to smooth production? Is an empirical analysis of industry-level data enough? Is it necessary to analyze firm-level data in order to explain these effects?" These questions read as a research agenda on the mechanisms behind empirical observations on both macro and micro levels, recognizing, among other issues, the potential masking effects of aggregate data. In the operations management literature, inventory theory is often developed in a stylized manner; with strong assumptions that favor mathematical tractability over the inclusion of the myriad factors that are present in real life. The objective of these simplifications is to develop managerial insights that are both rigorous, and useful in the real world. In an exploratory study, Rumyantsev and Netessine (2007) find evidence that many insights from classical inventory models survive aggregation and do, in fact, hold up when analyzing empirical data.

The dynamics that stem from the interactions of subsequent echelons along a supply chain have been extensively studied in the operations management literature. The fact that relatively small shocks can introduce severe instabilities in entire systems was shown by Forrester (1958), and is a central idea behind the bullwhip effect. The bullwhip effect has long been analytically and experimentally understood, and its effects and causes have sparked a great amount of research that has delivered valuable managerial insights (Sterman, 1989; Lee et al., 1997b; Croson and Donohue, 2006). This research spans a wide range of methodologies, where the term bullwhip is used to describe a wide range of bullwhip-like phenomena in contexts beyond the formal definition of the classical causes and structural assumptions of the bullwhip effect (see, for example, Lee et al., 2014, for a description of the "green bullwhip"). The work reported in this study is consistent with this wider perspective on bullwhip phenomena. Even though the bullwhip effect itself is significant at the firm level (Metters, 1997; Fransoo and Wouters, 2000; Bray and Mendelson, 2012), attempts to empirically quantify the effect at higher aggregation levels have not been successful: studies have failed to prove it statistically significant at an industry level (Cachon et al., 2007; Bu et al., 2011). The lack of clear empirical evidence is attributed to the influence of factors present in government statistics such as their high level of aggregation (Chen and Lee, 2012), and seasonal adjustment (Gorman and Brannon, 2000). Furthermore, as Rumyantsev and Netessine (2007) point out, extending many structural properties from single-product, single-echelon models to higher aggregation levels

also requires the assumption that products be homogeneous and their inventory control be synchronized.

With this in mind, the financial crisis of 2008 allows us to study empirical data in a different way. Following the bankruptcy of Lehman Brothers on September 2008, the financial world found itself in turmoil; credit dried up almost instantly and many companies in the world shifted their financial priorities according to the "cash is king" motto: liquidity became essential. Freeing up cash in the short term through inventory divestment is one strategy that can be followed by companies in times of distress (Sudarsanam and Lai, 2001). In a recent work, Pesch and Hoberg (2013) conduct an empirical study that shows that firms in financial distress reduce their inventories as part of their turn-around strategy: 70% of the firms in their sample reduce their inventories, with a median reduction of 9.4% of all inventories. We hypothesize that firms all over the world reacted to the financial collapse by significantly reducing their inventory targets. This, combined with the extraordinary synchronization observed during the period (Alessandria et al., 2010) and the ever increasing influence of supply chain dynamics in the global economy (Escaith et al., 2010), introduced a synchronized, endogenous, inventory shock that generated an inventory-driven bullwhip effect. Early studies following the financial crisis seem to confirm this view in the manufacturing sector (Dooley et al., 2010). Using the collapse as a natural experiment, we model supply chains at an aggregate-echelon level, use exogenous end market data to drive those models, and validate them with primary empirical data collected at a major dutch chemical company.

3. Theoretical background and model structure

In this section, we present our echelon model based upon Sterman's managerial decision making and supply chain models (Sterman, 1989, 2000) and follow with an introduction to the de-stocking logic we use to model the hypothesized reaction to the credit crisis.

3.1. Echelon model

An echelon model consists of three decision areas (see Fig. 1): the forecasting and orders sector tracks the incoming customer orders, maintains the echelon sales forecast, and generates
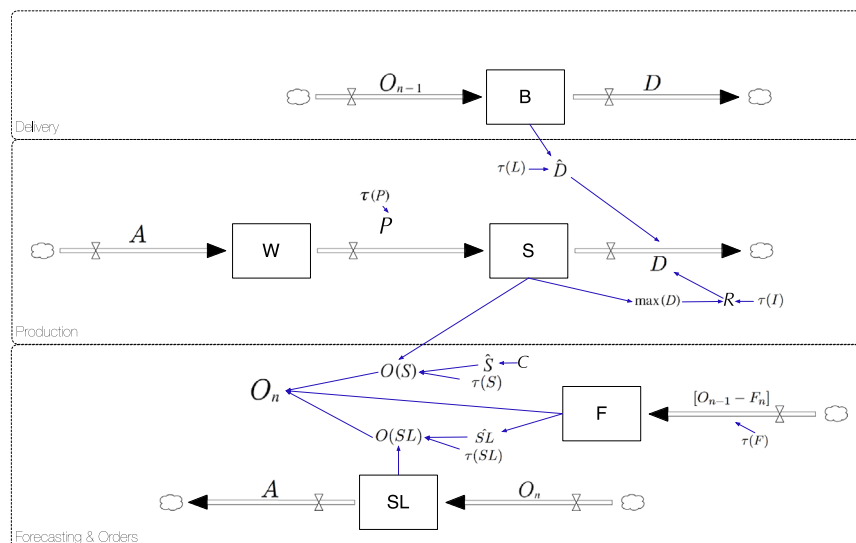

Fig. 1. Overview of a modeled echelon.

material orders. The production sector regulates inventories and production, and the delivery sector keeps track of customer deliveries and backlogs. The model assumes no lost sales, and is based on continuous time system dynamics simulations. There is no sequence of events as such; cause and effect relationships are modeled by differential equations (i.e. we model rates of change), and products are modeled as continuous flows (demand is an outflow, incoming orders an inflow).

Because these echelon models are linked to one another (deliveries from one echelon become material receipts for the echelon immediately downstream in its supply chain), each of the parameters we define has a subscript [$n = (1, ..., N)$] that represents its place in the supply chain. We number echelons from downstream to upstream: the most downstream echelon being 1 and the most upstream N. In the case of diverging supply chains, where one echelon can potentially have several direct customers, we introduce a second index after a period, that indicates the existence of other parallel echelons in the supply chain. Table 1 shows a summary of all parameters and variables in the echelon model.

3.2. Forecasting

The forecasting sector maintains a sales forecast by accumulating the differences between the incoming customer demand (O_{n-1}) and the previous forecast (F_n). When demand exceeds the forecast it is updated upwards and vice-versa. To allow for a smoothing of the forecast, these differences are divided by the forecast adjustment time ($\tau_n(F)$), indicating whether the whole difference or only a fraction is taken into account.

$$\left(\frac{d}{dt}\right)F_n = \frac{O_{n-1} - F_n}{\tau_n(F)}. \tag{1}$$

3.3. Production

The production sector models the flow of material through the echelon. The incoming material rate (A_n) is equal to the delivery rate of the immediately upstream echelon (D_{n+1}),

$$A_n = D_{n+1}. \tag{2}$$

The supply line is the cumulative difference between orders placed and orders received,

$$\left(\frac{d}{dt}\right)SL_n = O_n - A_n. \tag{3}$$

Incoming material is stored as work in process (W_n). In the interest of simplicity we do not model any production release rule. Thus, the work in process stock is not used strategically or as a control variable: all incoming material is committed to production, and the production rate is modeled by applying a fixed delay (equal to the production time PT_n) to the order arrival rate. System dynamics modeling allows for the introduction of this discrete step in the model, which approximates the real production process,

$$P_n = DELAY(A_n, PT_n). \tag{4}$$

Eq. (4) assumes a production model where the manufacturing time is independent of the utilization rate, it also implicitly assumes that there are no capacity limitations for production (the model can be straightforwardly extended to include capacity limitations).

On -hand inventory (S_n) depends on the delivery rate (D_n) and the production rate (P_n),

$$\left(\frac{d}{dt}\right)S_n = P_n - D_n. \tag{5}$$

Material orders are based on an anchor and adjustment heuristic (Tversky and Kahneman, 1974): the sales forecast acts as the anchor, with the adjustment stemming from the difference between actual and target stock (and supply pipeline) levels.

To calculate the target stock, we start with the desired on hand inventory coverage measured in time units ($\hat{C}_n$). When this is multiplied by the sales forecast, we obtain the desired on hand stock ($\hat{S}_n$) in units of product.

$$\hat{S}_n = \hat{C}_n F_n. \tag{6}$$

Analogously, there is a supply line level ($\hat{SL}_n$) consisting of the multiplication of the lag (lead time) and the forecasted volumes,

$$\hat{SL}_n = F_n(L_n). \tag{7}$$

3.4. Orders

Once we have calculated the desired levels of on-hand and supply line inventories, we generate adjustment orders with the purpose of closing the gap between the actual values of these inventories, and their desired (target) levels. The inventory adjustment time ($\tau_n(S)$) and supply line adjustment time ($\tau_n(SL)$) represent the time allowed for these quantities to reach the desired levels. These adjustment times model the behavioral aspect of the order generation. Short times imply a nervous buying behavior whereas a long adjusting time is equivalent to a smooth ordering strategy. We define the stock adjustment orders ($O_n(S)$) and supply line adjustment orders ($O_n(SL)$) as

$$O_n(S) = \frac{\hat{S}_n - S_n}{\tau_n(S)}, \tag{8}$$

$$O_n(SL) = \frac{\hat{SL}_n - SL_n}{\tau_n(SL)}. \tag{9}$$

Eqs. (10) and (9) calculate the difference between desired and actual values and spread these in equal parts over the amount of periods specified by the adjustment times. Finally, generated orders (O_n) are calculated as

$$O_n = \max\{0, F_n + O_n(S) + O_n(SL)\}. \tag{10}$$

3.5. Delivery

A backlog is used to keep track of orders. The backlog is calculated as the cumulative difference between the incoming customer order rate O_{n-1} and actual delivery rate (D_n). O_0, the demand observed by the echelon closest to the end market, is the only exogenous input to the model 1,

$$\left(\frac{d}{dt}\right)B_n = O_{n-1} - D_n. \tag{11}$$

The order delivery rate (D_n) is the rate of product that is actually shipped out in response to the incoming customer orders. To calculate this, we first define the desired delivery rate ($\hat{D}$), which is equal to the current backlog divided by the expected delivery delay ($\tau_n(L)$),

$$\hat{D}_n = \frac{B_n}{\tau_n(L)}. \tag{12}$$

The maximum delivery rate ($max(D)_n$) per period depends on the ability of firm to physically prepare the products for shipment, modeled as the minimum time to fill orders ($\tau_n(I)$),

$$max(D)_n = \frac{S_n}{\tau_n(I)}. \tag{13}$$

We calculate the delivery ratio (R_n) as the proportion of outstanding orders that can be shipped from stock,

$$R_n = \min\left\{1, \frac{max(D)_n}{\hat{D}_n}\right\}. \tag{14}$$

Finally, the actual order fulfillment rate is equal to the desired delivery rate multiplied by the delivery ratio,

$$D_n = \hat{D}_n R_n. \tag{15}$$

Alternatively, we can combine Eqs. (13)–(16) and define the order fulfillment rate as

$$D_n = \min \left\{ \frac{B_n}{\tau_n(L)}, \frac{S_n}{\tau_n(I)} \right\}. \tag{16}$$

3.6. Modeling de-stocking decisions

We model de-stocking decisions by decreasing the desired inventory coverage ($\hat{C}_n$) of an echelon n at time T by a fraction d_n (with $0 \le d_n < 1$).

Thus, we can define $\hat{C}_n$ as

$$\hat{C}_n = \begin{cases} \hat{C}_n & \text{if } t < \text{T}, \\ (1-d_n)\hat{C}_n & \text{if } t \ge \text{T}. \end{cases} \tag{17}$$

where C_n is the desired stock coverage in "normal" (non-crisis) situations. It is important to note that de-stocking is a decision to lower target stock levels that are measured in time units. It is not a decision to reduce its absolute value, nor does it imply the destruction or writing-off of inventory. In this way, we separate explicit decisions to lower inventory targets from the implicit reductions that come from a decrease in sales.

3.7. On the equivalence of the ordering policy

The model presented in this section is a straightforward extension of the model found in Sterman (2000). In particular, we introduce an explicit de-stocking decision and a discrete production delay. The structure of this ordering policy, however, is not unique to System Dynamics models, and can be found in other branches of the literature.

In the behavioral operations literature, an equivalent rule is used to model the decision-making behavior of human managers. In this context, the model is presented based upon its equivalence to an 'anchor and adjustment heuristic' (Tversky and Kahneman, 1974). These heuristics are used to describe human decision-making biases: Orders are calculated by selecting an anchor (in this case the forecast) with subsequent adjustments motivated by deviations from the target stock and supply line levels (Sterman, 1989; Croson et al., 2014).

The control theoretic branch of inventory theory also uses a family of models based on the same principles as the model described in this paper. The more general of the models in this framework, the Automatic Pipeline Variable Inventory Order-based Production Control System (APVIOPCS), is a discrete-time, constant-coverage, equivalent of our System Dynamics model. Linking this framework with other branches of inventory theory, Dejonckheere et al. (2003) show that these models are essentially modified Order Up To (OUT) policies. In particular, when the supply line and stock adjustments are taken fully into account every period then this model is equivalent to an OUT policy with a safety lead time proportional to the inventory coverage. The more general cases, when the supply line and stock adjustments are not equal, nor are they taken fully into account every period (as in this paper), can thus be thought of smoothed variations of OUT policies—with the key difference being that OUT policies have only one feedback loop, for the inventory position (the sum of the on-hand stock and the supply line). For a survey of the different policies explored in the control theoretic literature, and a discussion on the implications of independent supply line and stock adjustments, we refer the reader to Ortega and Lin (2004) and Udenio et al. (2013).

4. Results and analysis

In this section, we use the echelon model as a building block to construct, calibrate, and validate 4 different supply chain models based upon data collected at our research company.

The methodology presented thus far concerns the modeling of a single echelon in a supply chain: The input to an echelon model is a customer order and its output is an order placed to a supplier. To model a supply chain, we link echelon models according to the customer/supplier relationships defined by its structure (e.g. number of echelons, linear, divergent) and parameterize the individual echelons. We run the supply chain models using end-market sales data as their exogenous inputs.

Each of the echelon models is defined by operational and behavioral parameters. Operational parameters posses a concrete interpretation in the day-to-day operation of a firm (e.g. target stocks and production times) and are thus set based upon expert interviews. Behavioral parameters (e.g. supply line and stock adjustment times), on the other hand, define the relationship between internal variables—product of explicit or implicit managerial decisions—and are thus estimated through a process of model calibration. The de-stocking decisions we hypothesize, however, do not fall squarely in either of these definitions. While these decisions correspond to the operation of the firm, we could find no hard evidence of the desired inventory reductions. Rather, the de-stocking decisions corresponded to financial recommendations from upper management, which were estimated to be 'on the order of 10–20%'. Similarly, de-stocking decisions do not conform to the definition of a behavioral characteristic of the models. Thus, de-stocking is estimated via scenario analysis based upon expert interviews: Feasible de-stocking quantities (from 5 to 30% reductions of desired stocks) are defined in discrete increments, the calibration is performed for each of these scenarios, and the best fit is chosen. Potentially, the amount of de-stocking could depend on a series of firm characteristics such as the type of product and distance from the end market. However, due to the limitations of our data, we can only quantify the cumulative effect of de-stocking on the uppermost echelon. We therefore use a single de-stocking parameter for each supply chain.

The rest of this section is divided as follows. We explain the model set-up and data collection in Section 4.1. Then, we define the structure of the modeled supply chains and the operational parameters in Section 4.2, and the estimation of behavioral parameters in Section 4.3. Finally, we study the historical fit of the model in Section 4.4, and analyze an alternative model, where the de-stocking hypothesis is suppressed, in Section 4.5.

4.1. Model set-up and data collection

Two distinct flows appear when we link individual echelon models to form a supply chain model: An information flow that travels upstream (orders), and a material flow that travels downstream (deliveries). The information flow of any supply chain originates at the sales point of a finished product (i.e. its end-market). Thus, the demand information observed by an upstream entity is a function of the original signal, generated by the end-market, and transformed—throughout its flow upstream—by the subsequent echelons of the particular supply chain (in the case of divergent supply chains, the combination of end-market signals). We use the 2008 credit crisis as a natural experiment because it allows us to link these end-market signals to the corresponding upstream demand: The synchronization observed during the period effectively controls for the smoothing effects of aggregation. Explicitly, we assume that (a) entities at a given echelon share the same structure, (b) entities at a given echelon share the same behavior during this time frame, and (c) the information

Table 1
Definitions and sources of model parameters and variables.

Parameter/Variable		Dimensions	Source
Parameters			
L_n	Incoming delivery lead time at echelon n	Weeks	Set based on interviews
PT_n	Production time of echelon n	Weeks	Set based on interviews
$\tau_n(SL)$	Supply line adjustment time at echelon n	Weeks	Estimated to fit past data
$\tau_n(S)$	Stock adjustment time at echelon n	Weeks	Estimated to fit past data
$\tau_n(F)$	Forecast adjustment time at echelon n	Weeks	Estimated to fit past data
$\tau_n(L)$	Expected delivery delay at echelon n	Weeks	Set a fortiori and based on interviews
$\tau_n(L)$	Minimum time to fill orders at echelon n	Weeks	Set a fortiori and based on interviews
$\hat{C}_n$	Desired on-hand inventory coverage at echelon n	Weeks	Set based on interviews
d_n	De-stocking fraction at echelon n	Dimensionless	Judgmentally set based on interviews
Variables			
S_n	On-hand stock at echelon n	Units	Endogenous
SL_n	Supply line at echelon n	Units	Endogenous
W_n	Work in process stock at echelon n	Units	Endogenous
P_n	Production rate of echelon n	Units/Week	Endogenous
F_n	Sales forecast at echelon n	Units	Endogenous
$\hat{D}_n$	Desired delivery rate echelon n	Units/Week	Endogenous
$\hat{S}_n$	Desired on-hand inventory at echelon n	Units	Endogenous
$\hat{SL}_n$	Desired supply line at echelon n	Units	Endogenous
$O_n(SL)$	Supply line adjustment of orders at echelon n	Units	Endogenous
$O_n(S)$	Stock adjustment of orders at echelon n	Units	Endogenous
O_n	Orders placed by echelon n	Units	Endogenous
D_n	Delivery rate at echelon n	Units/Week	Endogenous
A_n	Incoming material rate at echelon n	Units/Week	Endogenous
R_n	Delivery ratio echelon n	Units/Week	Endogenous
B_n	Backlog at echelon n	Units	Endogenous

distortion that is observed in the passage of demand information upstream, from its origin in the end-market, corresponds to locally rational policies at each stage of the supply chain (i.e. no demand information is arbitrarily created or discarded by intermediate echelons). We base this approach upon the observation that empirical data shows that, during the credit crisis, turning points in sales and inventories were indeed aligned by tiers, suggesting synchronized behavior by stages (Dooley et al., 2010).

Conceptually, we model echelons that represent a group of competing companies providing the same product to the same supply chain. This follows what Sprague and Wacker (1996) define as the modeling with a "disaggregation by stages along the inventory stream". They point out that management practice generalizations made in this way recognize the impact of the management of inventory as it progresses through the stages. In particular, we model 4 different supply chains to which a dutch chemical firm (where we collected primary data) belongs to. As a reference, 4 different business units of this firm are situated 4–5 echelons upstream from retail demand in each of these supply chains. The upstream products sold into these supply chains are resins, thermoplastics, and polymers; the end-markets where these can be found are shown in Table 2. Each individual supply chain is defined by a "product family" of the upstream manufacturer. The aggregation level for each of these is defined by each business unit and represents an aggregation level that is used for medium-term decision making in each unit's Supply and Operations Planning (S&OP) procedures.[1] SKU's in the given product families generally only differ in the mix and quantity of the component materials; characteristics such as production time, safety stocks, and lead times are assumed to be equal.

We use monthly, EU27 sales data available from Eurostat as a proxy for the demand for each of the end-markets. The series used are: Construction index, automobile registrations, household

---

[1] Product families analyzed in this study are, among others, water-borne resins used in the production of the paint used in a specific range of construction products, and automotive-grade Ethylene Propylene Diene Monomer (EPDM) rubber used in the manufacturing of engine hoses.

goods retail index, and production indexes for: food products (C10), paper and paper products (C17), glass and glass products (C23.1), basic metal and metal products (C24), and motor vehicles (C29.1). All data is normalized with the average of 2007 = 100. Due to the continuous nature of the System Dynamics simulations, the monthly frequency of the data is approximated to the numerical integration step (daily) through a cubic spline interpolation. This implicitly assumes continuous production, which is reasonable in the context of process industries. The work at each of the four sites of the company began with a kickoff meeting with management where the objectives and scope of the study were explained and defined. Following these, interviews were conducted with employees to formalize data collection procedures. The structure of the supply chain model and the parameterization of operational parameters is based on input from these employees, complemented with information obtained from players distributed along the supply chain. The modeling work was performed on-site, which allowed for additional ad-hoc interviews and further familiarization with the particulars of each individual supply chain.

4.2. Structure and operational parameters

The number of echelons, structure, and end markets of each of the 4 supply chains are all different and can be seen in Fig. 2. The supply chains in this study consist mainly of chemical firms upstream and make-to-stock component suppliers downstream. For further context, we present an appendix where we describe the different stages of supply chain 3 in greater detail. For this study, we consider our research site to be the upstream-most boundary of each supply chain. The parameterization of the operational parameters per echelon is shown in Table 3.

To simplify the models, we assume deterministic lead times and the availability of resources such that order preparation does not introduce significant lags. Thus, the expected delivery delay ($\tau_n(L)$) is equal to its own delivery lead time (L_{n-1}), and the minimum time to fill orders ($\tau_n(I)$) is equal to 1. Due to the absence of disaggregated data, the lead time is defined as the time between placing an order and its receipt (i.e. it encompasses

both the informational and physical components of the delay). We make two additional assumptions regarding the boundary conditions: (1) Orders placed by the uppermost echelon in a supply chain are always served by a supplier with infinite stock, and (2) downstream demand is exogenous and composed of the individual demand signals of the end markets that require the materials produced upstream.

Having defined the structure and operational parameters for each of the supply chain models, we proceed with the analysis of the de-stocking decisions and the estimation of behavioral parameters through calibration.

4.3. Model calibration and behavioral parameters

Oliva (2003) defines model calibration as the process of estimating parameters to obtain a match between modeled and observed behavior and argues that it is, in itself, a stringent test of the validity of the model linking structure and behavior. Nevertheless, he points out that achieving a good historical fit is not enough to confirm the dynamic hypothesis behind the model; the model has to match the observed behavior for the right reasons. Partial model calibration, the process of estimating parameters within a subset of model parameters instead of the entire model parameter space, introduced by Homer (1983) is the preferred calibration strategy for system dynamics models because it "reduces the risk of the structure being forced into fitting the data, increases the efficiency of the estimation (estimators with smaller variances), and concentrates the differences between observed and simulated behavior in the piece of structure responsible for that behavior" (Oliva, 2003). However, we cannot perform partial calibration for our supply chain models because we lack primary sales and inventory data at the intermediate levels, and it is not possible to map secondary empirical data to individual echelons. To overcome this, we perform a full model calibration and follow it with: (i) a sanity check of the estimated parameters (is the model structure sound?) and (ii) the test of an alternative hypothesis (can we achieve the same behavior through a different structure?) to increase the confidence in our model.

We use the corresponding secondary EU27 end-market data as the input for each of the end-markets in our models and primary

Table 2
Summary of end-markets served by the 4 modeled supply chains.

Supply chain	End markets
Supply chain 1 (resins a)	Residential and commercial construction; residential and commercial repair & maintenance.
Supply Chain 2 (resins b)	Residential and commercial construction; residential and commercial repair & maintenance; furniture sales.
Supply chain 3 (polymers)	Automotive sales.
Supply Chain 4 (thermoplastics)	Automotive manufacturing; Glass panel manufacturing; metal manufacturing; Food manufacturing; paper and pulp manufacturing; residential and commercial construction.

Table 3
Operational Supply Chain parameters per echelon.

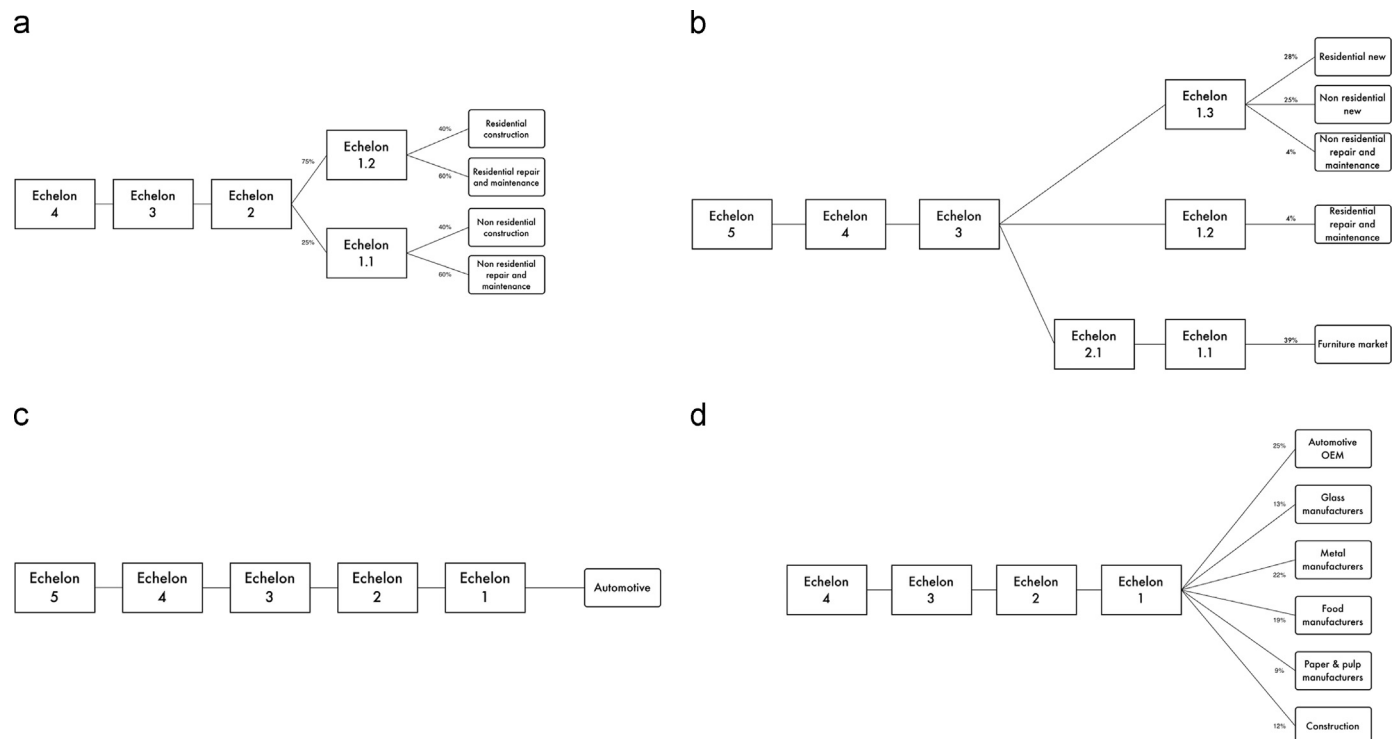|	Supply chain 1				Supply chain 2				Supply chain 3				Supply chain 4		
	$\hat{C}$	L_n	PT_n		$\hat{C}$	L_n	PT_n		$\hat{C}$	L_n	PT_n		$\hat{C}$	L_n	PT_n
1.1	5	10	1	**1.1**	8	4	1	**1**	4	3	1.5	**1**	8	0.25	2
1.2	5	10	1	**1.2**	4	4	1	**2**	2	1	1	**2**	14	0.25	5
2	8	2	1	**1.3**	4	4	1	**3**	1.5	2	1	**3**	10	0.25	4
3	8	2	1	**2.1**	8	4	1	**4**	1	2	1	**4**	8	0.25	1
4	4	2	1	**3**	8	0.25	1	**5**	2	2	1				
				4	3	0.25	1								
				5	2	0.25	1								

Fig. 2. Supply chain structures. (a) Supply chain 1, (b) supply chain 2, (c) supply chain 3, and (d) supply chain 4.

sales data from our research company as a proxy for upstream demand. This assumption is reasonable all through the historical period used for the calibration (where no capacity shortages were observed). Due to the existence of only one exogenous time series per supply chain, we use 27 months of historical data (Jan 2007–Mar 2009) to calibrate the model, then freeze the model parameters and use further 24 months (Apr 2009–Mar 2011) to run and validate the model. The calibration time period is selected so that it includes the hypothesized de-stocking period. This step is implemented within the simulation software; simulations are performed for each supply chain by generating model runs where all the operational parameters are fixed as established in Section 4.2, while the behavioral parameters are varied. The cumulative sum of squared errors between the estimated demand and the historical sales data is calculated per run and the combination of parameters that minimizes this error is then chosen. Formally, the minimization corresponds to

$$\min_{\tau_n(F),\tau_n(SL),\tau_n(S)} \sum_{t=1}^{k} (O_{N-1}(t)-\tilde{D}_N(t))^2. \qquad (18)$$

where N is the most upstream echelon in the supply chain, and $\tilde{D}_N(t)$ is the historical sales data for time t at echelon N (our research company). In other words, we compare the orders generated by the modeled customers of our research site with the actual historical sales of said firm and search for the parameter values that minimize the error. The minimization is performed through a modified Powell–Brent algorithm (Brent, 2002). For computational purposes and to reduce the search space, $\tau(S)$, $\tau(SL)$, and $\tau(F)$ are estimated through their reciprocals, α_S, α_{SL}, and Θ (with α_S, α_{SL}, $\Theta \in [0, 1]$). Table 4 lists all the parameters estimated through calibration, including the 95% confidence intervals calculated through a sensitivity analysis. The de-stocking fractions, estimated through a combination of interviews and scenario analysis, are also shown in this table.

In all cases, the confidence bounds of the estimations for the uppermost echelon are lax: this is due to the data available for calibration being the historical sales of this echelon. None of the parameters in the model allow a firm to influence its own demand via strategic decisions. Thus, the uppermost echelon can either meet the demand or incur in destabilizing stock-outs. The confidence bounds represent the parameter space that allows for the former. Similarly, the amount of parameters being estimated from a single time series (between 12 and 21, depending on the supply chain) explain the size of the confidence intervals of the supply line adjustment time, which are particularly large. A parameter estimated to be ∞ corresponds to a parameter that is not taken into account in the ordering heuristic. The upper bound for the supply line adjustment time for all but two of the echelons is ∞, which suggests that we cannot reject the hypothesis that firms completely ignore the supply line. On the other hand, the lower bound for 2/3 of the echelons is larger than one, which suggests smoothing of the supply line adjustment.

This is consistent with results from experiments found in the behavioral literature (Sterman, 1989). The gap between desired and actual supply lines is severely underestimated in the ordering decisions: both the means and medians of the supply line adjustment time ($\tau(SL)$) are larger than the respective values for the stock adjustment times (two-sample t-tests on the means, Wilcoxon rank-sum tests for the medians, all $p \le 0.01$) as well as the values for the forecast adjustment time ($p \le 0.1$ when comparing means and $p \le 0.01$ when comparing medians).

When we compare the stock and forecast adjustment times, on the other hand, we find no statistical difference between neither their means nor their medians—suggesting a smoothing of the same order of magnitude for the adjustment of the inventory gap and the forecast updating.

We next compare the adjustment times between the different supply chains to test for any differences in the inherent behavior. We find no significant difference among the different supply chains, implying that the overall behavior, and the mechanisms, behind all the models is comparable.

4.4. Historical fit and structural validity

Following the calibration, we run the four supply chain models driven by the exogenous end-market and the de-stocking

Table 4
Estimated behavioral parameters and de-stocking fractions.

| Echelon	$\tau(S)$	95% CI		$\tau(SL)$	95% CI		$\tau(F)$	95% CI	
Supply chain 1	($d_n = 0.15$)								
1.1	**3.65**	2.17	6.39	**5.51**	2.17	15.07	**6.25**	1.00	29.53
1.2	**9.49**	4.42	14.46	∞	24.46	∞	∞	120.39	∞
2	**9.19**	7.59	11.38	∞	7.59	∞	**8.81**	5.63	14.88
3	**13.70**	11.86	15.97	∞	11.86	∞	**21.75**	16.54	29.12
4	**10.00**	1.30	∞	**60.25**	1.30	∞	**10.00**	1.00	2561.69
Supply chain 2	($d_n = 0.25$)								
1.1	**4.39**	3.20	5.71	∞	7.83	∞	**3.91**	1.00	17.71
1.2	**25.62**	1.00	∞	∞	1.00	∞	**110.74**	1.00	∞
1.3	**540.86**	159.26	∞	**901,432.01**	79.63	∞	∞	204.94	∞
2.1	**5.44**	2.81	7.80	**2.72**	1.40	6.28	**5.02**	2.22	7.77
3	**8.16**	6.67	10.01	**81.62**	3.34	∞	**14.11**	10.01	20.57
4	**31.84**	23.89	44.31	**76100.78**	11.95	∞	**8.60**	6.86	10.76
5	**10.00**	1.00	72.52	**100.00**	1.00	∞	**10.00**	1.00	49.93
Supply chain 3	($d_n = 0.20$)								
1	**12.97**	10.61	15.87	∞	11.02	∞	**1377.31**	124.29	∞
2	**3.11**	1.78	5.04	**10.80**	1.88	∞	**1.00**	1.00	16.04
3	**9.21**	7.81	11.02	∞	37.13	∞	**13.27**	8.40	21.96
4	**12.09**	9.42	16.04	∞	10.09	∞	**11.52**	8.47	16.01
5	**10.00**	1.00	∞	**60.25**	1.00	∞	**6.03**	1.00	∞
Supply chain 4	($d_n = 0.10$)								
1	**14.55**	12.11	17.45	**16.28**	6.06	∞	∞	399.02	∞
2	**8.63**	7.27	10.23	**37,643.27**	15.08	∞	**16.15**	10.55	24.47
3	**11.23**	9.33	13.57	∞	19.70	∞	**8.63**	5.90	12.26
4	**16.88**	1.00	445.38	**56.67**	0.50	∞	**1011.40**	1.00	∞
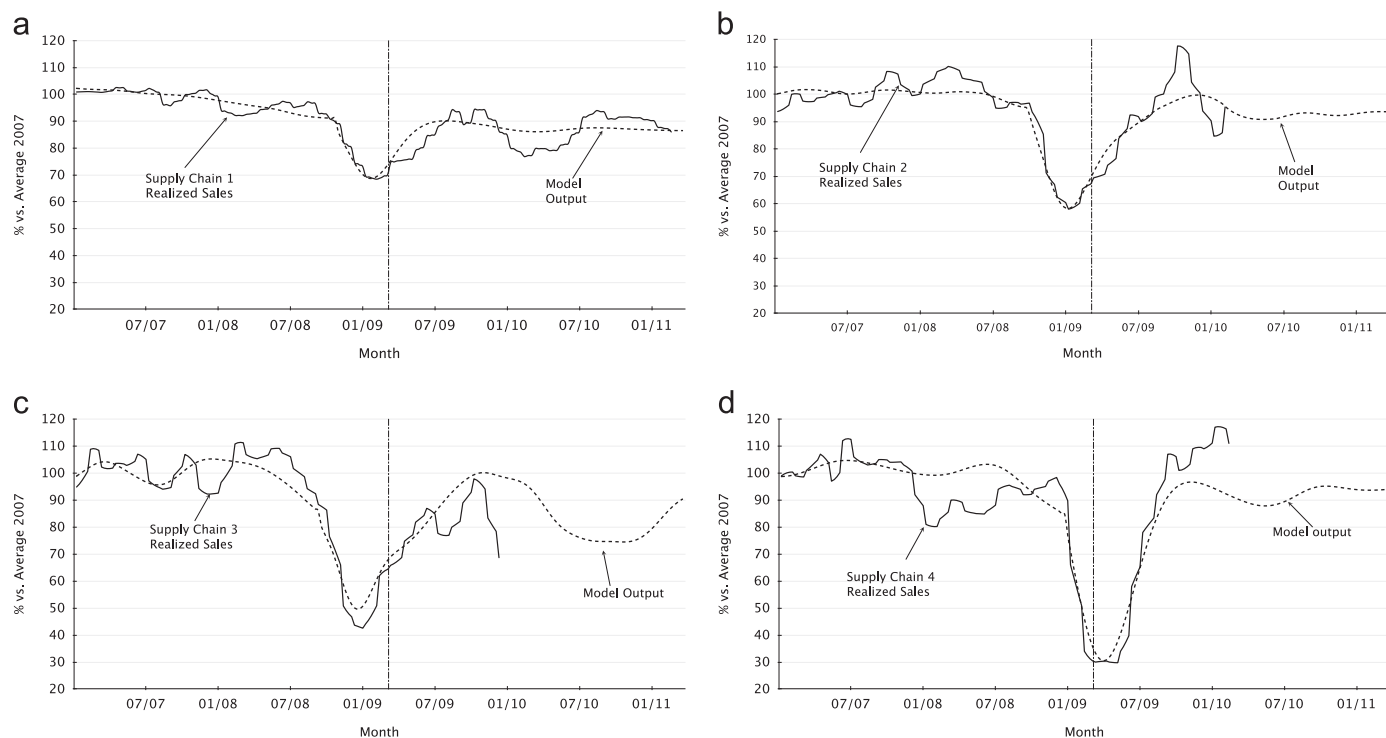
Fig. 3. Model output vs. seasonally corrected sales data. (a) Supply chain 1, (b) supply chain 2, (c) supply chain 3, and (d) supply chain 4.

hypothesis. To validate the models, we compare the actual sales realizations of the upstream-most echelon of each of the supply chains with the corresponding modeled demand. Fig. 3 shows the model outputs against the seasonally corrected upstream demand realizations.[2] The vertical axis represents the demand expressed in % of the average 2007 demand and the dotted vertical lines indicate the threshold for the calibration period. Table 5 shows the root mean squared error (RMSE), R^2, and Theil inequality statistics for the data series shown in the figure. These inequality statistics decompose the mean square error into three fractions representing: unequal means (U_m), unequal variances (U_s), and imperfect correlation (U_c) (Theil, 1966). A low U_m indicates a strong correspondence between the modeled mean and the actual mean, and a low U_s indicates a similar correspondence between variances. Therefore, low variance and means statistics indicate that the error is unsystematic, and therefore desirable (Oliva and Sterman, 2001).

The models, driven by one exogenous data series (end custo-mer demand), and the de-stocking policy (desired stock reductions in September 2008) show good tracking of the overall bullwhip-like behavior of the system. The low RMSE values, combined with the unsystematic nature of the errors for all four data series increase the confidence in the model, and hence in the underlying de-stocking hypothesis.

However, a match between observed and simulated behavior is not in itself enough to accept the model and hypothesis. As Oliva (2003) explains, "There is a chance that a set of parameter values might be capable of replicating the observed behavior through a set of unrealistic formulations, and thus generate the right behavior for the wrong reasons". To test the validity of the model, we need to analyze what the estimated parameters say about its structure, and follow this up with an analysis of an alternative

Table 5
Historical fit statistics.

Model	RMSE (%)	R^2	U_m	U_s	U_c
Supply chain 1	4.85	0.60	0.041	0.226	0.733
Supply chain 2	5.70	0.84	0.025	0.087	0.888
Supply chain 3	7.39	0.84	0.020	0.101	0.879
Supply chain 4	9.86	0.82	0.001	0.051	0.948

model to test whether the same behavior can be achieved through a different structure.

As mentioned in Section 4.3, the large confidence bounds for the supply line adjustment time, coupled with the statistically significant underestimation of the supply line ($\tau(SL) > \tau(S)$) are consistent with findings from the behavioral operations literature. In this (mainly) experimental body of work, individual human behavior is analyzed in the context of the beer distribution game. Three of the most salient such studies are Sterman (1989), Croson and Donohue (2006), and Croson et al. (2014). In these experi-ments, students (and professionals) play the beer game under different settings, and their behavior is estimated through the use of regression analysis on a decision rule equivalent to Eq. (10). These studies consistently report underestimation of the pipeline and a smoothing of both the forecast and stock adjustment times. The fact that these characteristics are observed in the behavioral parameters of our calibrated models increases our confidence on its structural validity—the behavior of the models is consistent with prior research.

4.5. Alternative model

The de-stocking hypothesis presented in this paper is moti-vated by a variety of results from the inventory management and economics literature. It has been shown that firms can convert assets into cash in the short-term (Sudarsanam and Lai, 2001), and that lowering inventories is a common response to financial distress (Pesch and Hoberg, 2013). Furthermore, studies focused

---

[2] Note that the time series for actual sales in supply chain 1 has an extra year of data available. This corresponds to the nature of the collaboration agreements, which where negotiated independently with each of the different business units.
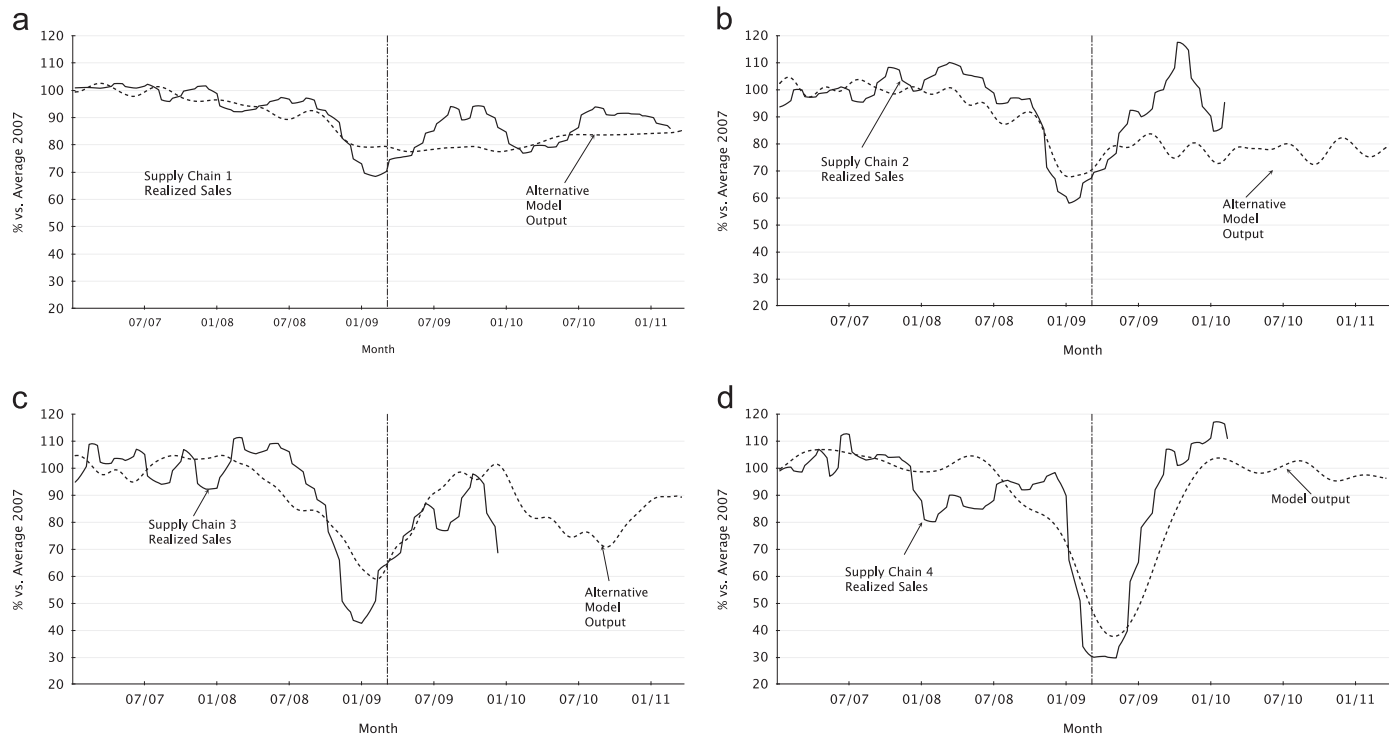
Fig. 4. Alternative model output vs. seasonally corrected sales data. (a) supply chain 1, (b) supply chain 2, (c) supply chain 3, and (d) supply chain 4.

on the mechanisms behind the recent 2008 financial crisis have reported on its extraordinary magnitude and synchronization (Alessandria et al., 2010). In line with this, anecdotal evidence points to decisions having been made to reduce working capital: formal and informal interviews with decision makers across the industry support this view.

However, the empirical validation of our hypothesis not possible: Inventory targets are not explicitly reported, and we cannot use actual inventories as a proxy for inventory targets during the time-frame of this study. The decision to reduce inventory targets triggers a shock that immediately affects orders, but its effect on inventory levels is substantially more complex: The combination of time delays and declining demand caused inventory levels to spike following the start of the crisis, further increasing the gap between target and actual inventory levels.

Therefore, to further test our model and—in particular—the de-stocking hypothesis driving it, we perform additional experiments to rule out alternative explanations. To do so, we repeat the calibration procedure from Section 4.3 this time with a version of the model where Eq. (17) is replaced by a constant $\hat{C}_n$. The structure of the models and the operational parameters of the alternative model remain the same as those of the original model. In Fig. 4 we show the model outputs of these calibrated alternative models; Tables 6 and 7 respectively show the calibrated behavioral parameters and the fit statistics of the alternative models for each supply chain.

With regard to the calibrated parameters of the alternative model we, again, observe relatively large values for the supply line adjustment times coupled with large confidence interval. This suggests a significant under-estimation of the supply line, a result that is consistent with behavioral theory and with the findings of the original model. In particular, we find no significant difference when testing the mean and median values of the supply line adjustment times of both the original and alternative models.

When looking at the output of the alternative model, we see that the alternative model adequately tracks the average, or long-term, demand variations but fails to explain the magnitude of the demand drops and their timing—we see that the output of the alternative model is appreciably more stable than the original one. If we compare the alternative model runs with the original (de-stocking) model runs we can see that the end market sales drive the long-term evolution of upstream sales, while the short term bullwhip-like dynamics are dominated by shocks. This is confirmed by an analysis of the fit statistics of the alternative model (presented in Table 8). Not only does the original model exhibit lower RMSE and larger R^2 (two-sample t-tests for the means $p \leq 0.05$, Wilcoxon rank-sum test for the medians $p \leq 0.05$, and $p \leq 0.1$ respectively), but the Theil statistics denote that the error presents in the alternative models is more systematic than those of the original models.

Behaviorally, it is interesting to note that both the original and alternative models seem to present a consistent picture. Especially when comparing the median values of the calibrated adjustment times, we see that the stock and forecasting adjustment times are of a comparable magnitude, while the supply line adjustment times are significantly larger. This implies that, in order to track medium to long term demand changes, firms tend to smooth their orders. Consequently, short term dynamics seem to stem from other sources of adjustments such as changes in desired stock levels.

5. Conclusions and managerial insights

Behavioral dynamics in supply chains have been widely researched. Initial studies by Forrester (1958) analyzed data at the level of individual or series of companies. Following the work by Lee et al. (1997a), extensive analytical work has been conducted, and more recently, driven by the work by Sterman (1989) and Croson and Donohue (2005), focus has been on laboratory experimentation. On the empirical front, Cachon et al. (2007) do not find conclusive evidence for the existence of the bullwhip effect in aggregate empirical data. Chen and Lee (2012), through analytical work, argue that it is the aggregation of the data that plays an important role in hiding some of the effect, which is observed at a firm level (Bray and Mendelson, 2012).

Table 6
Estimated behavioral parameters for the alternative model.

Echelon	$\tau(S)$	95% CI		$\tau(SL)$	95% CI		$\tau(F)$	95% CI	
Supply chain 1									
1.1	**1.00**	1.00	1.27	**5.80**	5.69	7.47	**1.00**	1.00	2.73
1.2	**35.51**	34.08	37.45	**134,441.23**	18396.22	∞	**4779.57**	4121.37	5923.29
2	**11.23**	7.44	22.01	**187,103.83**	3.72	∞	**7.49**	4.70	12.19
3	**9.67**	6.82	15.90	**59,794.30**	3.41	∞	**7.38**	3.92	13.87
4	**10.00**	1.00	∞	**100.00**	0.50	∞	**6.03**	1.00	3427.92
Supply chain 2									
1.1	**1.98**	1.63	2.62	∞	3.02	∞	**3.92**	1.57	9.77
1.2	**21.03**	1.00	∞	**1666.67**	1.00	∞	**47.00**	1.00	∞
1.3	**49.68**	28.39	96.79	∞	1.00	∞	**84.65**	38.60	343.64
2.1	**1.00**	1.00	3.48	**0.50**	1.00	0.88	**2.86**	1.20	5.85
3	**11.76**	7.51	18.35	**10.00**	1.00	∞	**10.63**	6.75	17.60
4	**3.04**	1.29	6.46	**1.00**	1.00	∞	**13.69**	5.26	45.40
5	**10.00**	1.14	∞	**4.31**	1.00	∞	**10.00**	1.00	174.73
Supply chain 3									
1	**2.32**	2.27	2.46	**139.41**	35.45	∞	**1.00**	1.00	4.17
2	**19.18**	13.44	30.20	∞	6.76	∞	**15.89**	10.33	26.46
3	**18.07**	13.14	26.79	∞	22.45	∞	**16.05**	10.38	26.93
4	**1.01**	1.00	1.40	**4.25**	1.26	6.87	**1.00**	1.00	2.99
5	**10.00**	1.00	∞	**60.25**	1.00	∞	**10.00**	1.00	∞
Supply chain 4									
1	**19.98**	14.20	28.41	∞	12.74	∞	**2.52**	1.00	6.18
2	**17.75**	14.61	21.75	∞	19.95	∞	**9.95**	6.54	13.89
3	**18.23**	14.68	22.88	∞	19.31	∞	**10.75**	7.10	15.09
4	**14.52**	1.00	∞	**18.59**	1.00	∞	**5.22**	1.00	∞

Table 7
Alternative model fit statistics.

Model	RMSE (%)	R^2	U_m	U_s	U_c
Supply chain 1	7.87	0.16	0.245	0.389	0.366
Supply chain 2	12.02	0.43	0.200	0.036	0.764
Supply chain 3	10.93	0.65	0.027	0.193	0.780
Supply chain 4	12.86	0.70	0.005	0.015	0.980

Table 8
Summary statistics for original and alternative models.

Variable	Mean	St. dev	1st Quartile	Median	3rd Quartile
Original model					
$\tau(S)$	36.71	115.71	8.63	10.00	13.70
$\tau(SL)$	84,630.55	258,287.00	13.54	60.25	18,871.62
$\tau(F)$	146.92	386.95	6.25	10.00	16.15
RMSE	6.95	2.21	5.28	6.55	8.63
R^2	0.78	0.12	0.71	0.83	0.84
Alternative model					
$\tau(S)$	13.67	12.00	3.04	11.23	18.23
$\tau(SL)$	29,777.61	59,143.05	5.80	80.13	35,056.11
$\tau(F)$	240.32	1040.25	3.92	9.95	13.69
RMSE	10.92	2.18	9.40	11.48	12.44
R^2	0.49	0.25	0.30	0.54	0.68

In this study, we use observations following the collapse of Lehman Brothers in the Fall of 2008 to develop a hypothesis regarding target level setting and investigate the explanatory power of behavioral dynamics. Our study observes demand at the level of an individual company, but takes into account the hypothesized dynamic decision making behavior at meso-level. With this, our study sets itself apart from previous studies, and not only builds upon the lines of research discussed above, but also on research in economics studying inventory cycles.

Our results show that the theoretical results of (among others) Sterman (1989) and Croson and Donohue (2006) together with an inventory shock, can explain a large part of the dynamic evolution of demand observed upstream in the periods following the start of the recent credit crisis. The endogenous replenishment process drives the evolution of demand throughout the supply chain, determined by structural characteristics of the supply chain (following Forrester, 1958), and the hypothesized human behavior (following Sterman, 1989). The empirical evidence presented shows that slow reaction speeds, and an apparent underestimation of the supply pipeline are prevalent at higher aggregation levels, suggesting that they go beyond being a phenomenon of individual decision-making biases. At this level, the supply line underestimation seems to be caused not from an incorrect estimation of target values, but as a combination of the inherent reaction time of firms and a decision rule that eschews the tracking of the supply line by instead steering on large amounts of on-hand inventory. This finding calls for further study on the ordering behavior of firms; if behavioral biases influence decision-making at the echelon level, *how can—and should—firms overcome them?*. Equally important: *how do these behaviors change over time?*

To increase the confidence in our de-stocking hypothesis in the presence of limited data, we presented an alternative model without the hypothesized de-stocking. Our results in four supply chains show that the underlying behavior of both models is consistent between them and with prior research, and that while the exogenous demand at consumer level and endogenous ordering decisions in the supply chain drive the overall demand evolution, short-term demand bullwhip-like dynamics are mainly driven by the de-stocking response to the crisis.

For managers, our results have implications at both the strategic and tactical levels of decision making. Tactically, for managers it is much more important to keep track of consumer demands, supported by an endogenous simulation of ordering behaviors to make demand forecasts, rather than relying exclusively on information obtained from one or two echelons downstream. These simulation-based forecasts can drive decisions on plant openings and closures, staffing decisions, and aggregate inventory strategies. Additionally, our results highlight the importance of understanding the implications that policy changes
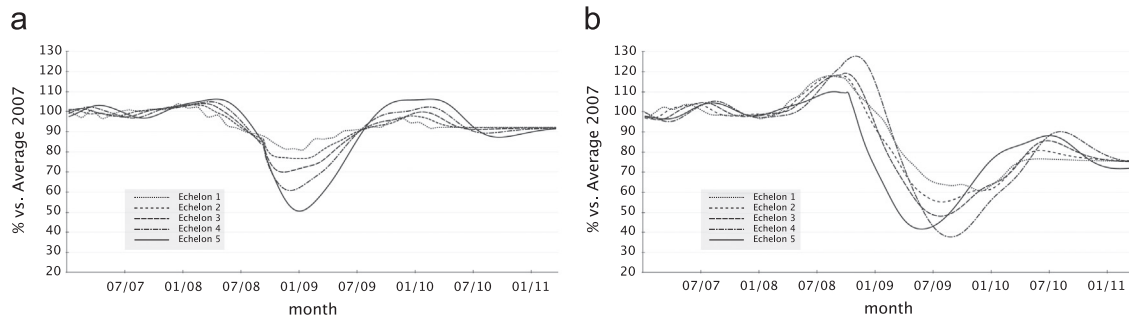
Fig. 5. Model output for sales and inventory at individual echelons. (a) Orders supply chain 3 and (b) inventories supply chain 3.

can bring into a supply chain. It is well known that aggregate inventory levels can serve as an additional way to achieve liquidity targets, however, limited research exists on the implications of such decisions on the stability of the entire supply chain. Strategically, we show that the structure of the supply chain impacts the clockspeed at which the supply chain operates. In this sense, we provide a formal model that can be used (a) to analyze the effects of structural and policy changes in the supply chain, and (b) to potentially become a decision-making tool in which endogenous behavioral changes form the basis of scenario-based forecasting. In this sense, our findings highlight the prospective value of information sharing. In cases such as the period studied in this paper, and consistent with experimental research (Croson and Donohue, 2006), knowledge about the underlying source of the observed demand dynamics (i.e. distinguishing between 'actual' demand drops and inventory adjustments) is crucial so as to adopt the correct response strategy.

There are limitations and opportunities for further research. First, we used a single time series of upstream sales per echelon model, which hinders our ability to perform partial model calibration and dissociate the de-stocking decisions according to the supply chain stages. To overcome this we performed the model calibration during a time-frame where supply chain decision-making was particularly synchronized. Further studies with detailed data at every stage of the supply chain can offer more robust statistical tests of firm-level behavior as well as bring insights regarding the influence of firm characteristics (in particular the distance from the end market) in the ordering and de-stocking behavior.

Second, the use of the term bullwhip effect to describe the phenomena investigated in this paper is consistent with the broader use of the term (De Kok, 2012; Disney et al., 2013; Lee et al., 2014). However, it can be argued that the term bullwhip should be restricted to situations where classical assumptions (such as constant behavior and constant parameters) hold.

Third, we make a series of implicit assumptions that may not necessarily apply in other industries or time periods. Our models assume independent echelons with no information sharing among them, with constant market share (nor pricing changes), a stable supply chain structure, no capacity limitations, and aggregate data. We expect these assumptions to be reasonable within the crisis time-frame, but further modeling efforts are necessary to test whether they can be applied during stable times, where the demand dynamics are more subtle.

Such studies, combining fine grained data from multiple echelons in a supply chain, have the potential to take us closer to the objective, both empirical and experimental, of testing whether endogenous mechanisms that we know govern the individual behavior (such as the underestimation of the supply line, and de-stocking and hoarding behaviors) can be consistently found at the aggregate level.

Appendix

In this appendix, we explore details of the model for one of the supply chains under study. In the next section, we introduce the supply chain in greater detail. Then, we analyze the model output for all the intermediate echelons in the supply chain.

Appendix A. Supply chain 3

This supply chain is defined by geographic market boundaries. The unit of upstream aggregation is as defined by the business unit "the production of ethylene propylene diene monomer (EPDM) rubber for the supply of the European automotive market". The steps in this supply chain are:

Echelon 1	OEM. Auto terminal.
Echelon 2	Module assembler. Whereupon products are assembled into modules to be used in the auto terminal.
Echelon 3	Converter. Where the rubber product is modified (through for example slitting, cutting, and extrusion) to form a finished product.
Echelon 4	Compounder. Where the raw materials are mixed with other ingredients to produce usable rubber compounds.
Echelon 5	EPDM producer, our research site.

Appendix B. Model output for intermediate echelons

Fig. 5 shows the demand and inventory dynamics for each subsequent echelon in supply chain 3. We see, as expected, an increase in order variability the further upstream we go in the supply chain. Additionally, the figure illustrates the impact of the delay of information transmission; the timing of the different turning points varies across the supply chain. Note that the assumption of an infinite capacity supplier to echelon 5 results in more responsive inventory dynamics for this echelon—this, however, does not affect the dynamics of its demand.

References

Alessandria, G., Kaboski, J.P., Midrigan, V., 2010. The Great Trade Collapse of 2008-09: an Inventory Adjustment?. NBER Working Papers 16059. National Bureau of Economic Research, Inc.

Baldwin, R., 2009. The Great Trade Collapse, Causes, Consequences and Prospects. Centre for Economic Policy Research, London.

Blinder, A.S., Maccini, L., 1991. Taking stock: a critical assessment of recent research on inventories. J. Econ. Perspect. 5, 73–96.

Bray, R.L., Mendelson, H., 2012. Information transmission and the bullwhip effect: an empirical investigation. Manag. Sci. 58, 860–875.

Brent, R., 2002. Algorithms for Minimization Without Derivatives. Dover Publications; New York.

Bu, X.Z., Zhou, Z.H., Guo, R.f., 2011. An empirical study on bullwhip effect in the supply chain; based on the statistical data from manufacturing industry of china. In: 8th International Conference on Service Systems and Service Management (ICSSSM), pp. 1–6.

Cachon, G.P., Randall, T., Schmidt, G., 2007. In search of the bullwhip effect. Manuf. Serv. Oper. Manag. 9, 457.

Chen, F., Drezner, Z., Ryan, J., 2000. Quantifying the bullwhip effect in a simple supply chain: the impact of forecasting, lead times, and information. Manage. Sci. 46, 436–443.

Chen, L., Lee, H.L., 2012. Bullwhip effect measurement and its implications. Oper. Res. 60, 771–784.

Croson, R., Donohue, K., 2005. Upstream versus downstream information and its impact on the bullwhip effect. Syst. Dyn. Rev. 21, 249–260.

Croson, R., Donohue, K., 2006. Behavioral causes of the bullwhip effect and the observed value of inventory information. Manage. Sci. 52, 323–336.

Croson, R., Donohue, K., Katok, E., Sterman, J., 2014. Order stability in supply chains: coordination risk and the role of coordination stock. Prod. Oper. Manag..

De Kok, A.G., 2012. The return of the bullwhip. Review of Business and Economic Literature 57 (3), 381–393.

Dejonckheere, J., Disney, S.M., Lambrecht, M., Towill, D., 2003. Measuring and avoiding the bullwhip effect: a control theoretic approach. Eur. J. Oper. Res. 147, 567–590.

Disney, Stephen Michael, Hoshiko, L., Polley, L. and Weigel, C., 2013. Removing bullwhip from Lexmark's toner operations. In: Proceedings from Production and Operations Management Society Annual Conference, Denver, CO, USA, 3–6 May 2013. ⟨http://orca.cf.ac.uk/id/eprint/53816⟩.

Dooley, K., Yan, T., Mohan, S., Gopalakrishnan, M., 2010. Inventory management and the bullwhip effect during the 2007–2009 recession: evidence from the manufacturing sector. J. Supply Chain Manag. 46, 12–18.

Escaith, Hubert, Lindenberg, Nannette, Miroudot, S., 2010. International Supply Chains and Trade Elasticity in Times of Global Crisis (February 1, 2010). World Trade Organization (Economic Research and Statistics Division) Staff Working Paper ERSD-2010-08. Available at: SSRN: ⟨http://ssrn.com/abstract=1548424⟩ or ⟨http://dx.doi.org/10.2139/ssrn.1548424⟩.

Feldstein, M., Auerbach, A., 1976. Inventory behavior in durable-goods manufacturing: the target-adjustment model. Brook. Pap. Econ. Activity 7, 351–408.

Fitzgerald, T., 1997. Inventories and the business cycle: an overview. Econ. Rev. Federal Reserve Bank Clevel. 33, 11–22.

Forrester, J., 1958. Industrial dynamics: a major breakthrough for decision makers. Harv. Bus. Rev. 36, 37–66.

Fransoo, J.C., Wouters, M., 2000. Measuring the bullwhip effect in the supply chain. Supply Chain Manag.: Int. J. 5, 78–89.

Gorman, M.F., Brannon, J.I., 2000. Seasonality and the production-smoothing model. Int. J. Prod. Econ. 65, 173–178.

Homer, J.B., 1983. Partial-model testing as a validation tool for system dynamics. Syst. Dyn. Rev. 28, 281–294.

Lee, H.L., Padmanabhan, V., Whang, S., 1997a. The bullwhip effect in supply chains. Sloan Manag. Rev. 38, 93–102.

Lee, H.L., Padmanabhan, V., Whang, S., 1997b. Information distortion in a supply chain: the bullwhip effect. Manag. Sci. 43, 546–558.

Lee, S.Y., Klassen, R.D., Furlan, A., Vinelli, A., 2014. The green bullwhip effect: transferring environmental requirements along a supply chain. Int. J. Prod. Econ. 156, 39–51.

Lovell, M., 1994. Researching inventories: why have not we learned more? Int. J. Prod. Econ. 35, 33–41.

Metters, R., 1997. Quantifying the bullwhip effect in supply chains. J. Oper. Manag. 15, 89–100.

Oliva, R., 2003. Model calibration as a testing strategy for system dynamics models. Eur. J. Oper. Res. 151, 552–568.

Oliva, R., Sterman, J., 2001. Cutting corners and working overtime: quality erosion in the service industry. Manag. Sci. 47, 894–914.

Ortega, M., Lin, L., 2004. Control theory applications to the production-inventory problem: a review. Int. J. Prod. Res. 42, 2303–2322.

Pesch, M., Hoberg, K., 2013. Inventory management under financial distress: an empirical analysis. Working Paper: Kühne Logistics University.

Rumyantsev, S., Netessine, S., 2007. What can be learned from classical inventory models? A cross-industry exploratory investigation. Manuf. Serv. Oper. Manag. 9, 409–429.

Sprague, L., Wacker, J., 1996. Macroeconomic analyses of inventories: learning from practice. Int. J. Prod. Econ. 45, 231–237.

Sterman, J.D., 1989. Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment. Manag. Sci. 35, 321–339.

Sterman, John D., 2000. Business Dynamics: Systems Thinking and Modeling for a Complex World, vol. 19. Boston: Irwin/McGraw-Hill.

Sudarsanam, S., Lai, J., 2001. Corporate financial distress and turnaround strategies: an empirical analysis. Br. J. Manag. 12, 183–199.

Theil, H., 1966. Applied Economic Forecasting. North-Holland Pub. Co, Amsterdam.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185, 1124–1131.

Udenio, M., Fransoo, J., Vatamidou, E., Dellaert, N., 2013. An analysis of the effect of response speed on the bullwhip effect using control theory. Beta working paper 420.