KIEL CENTRE FOR GLOBALIZATION

Centralized versus Decentralized Inventory Control in Supply Chains and the Bullwhip Effect

Horst Raff and Zhan Qu

No. 6 | November 2017

Centralized versus Decentralized Inventory Control in Supply Chains and the Bullwhip Effect

Horst Raff and Zhan Qu

Abstract

This paper constructs a model of a supply chain to examine how demand volatility is passed upstream through the chain. In particular, we seek to determine how likely it is that the chain experiences a bullwhip effect, where the variance of the upstream firm's production exceeds the variance of the downstream firm's sales. We show that the bullwhip effect is more likely to occur and is greater in size in supply chains in which inventory control is centralized rather than decentralized, that is, exercised by the downstream firm.

Keywords: bullwhip effect, production smoothing, inventory, supply chain, demand uncertainty, stockout avoidance

JEL Classification: L22, L14, M11

Financial support from the Leibniz Association through the Leibniz Science Campus "Kiel Centre for Globalization" is gratefully acknowledged.

Horst Raff
Kiel Centre for Globalization
University of Kiel
Wilhelm-Seelig Platz 1
D-24118 Kiel
Germany
raff@econ-theory.uni-kiel.de

Zhan Qu
Department of Economics
Technical University Dresden
Helmholtzstraße 10
D-01069 Dresden
Germany

About the Kiel Centre for Globalization (KCG): KCG is a Leibniz Science Campus initiated by Christian-Albrechts University of Kiel and Kiel Institute for the World Economy. It works on an interdisciplinary research agenda that evaluates the proliferation of global supply chains as an important aspect of globalization. To this end, the KCG brings together researchers from economics, ethics and management science. KCG is financially supported by the Leibniz Association and the State Government of Schleswig-Holstein. More information about KCG can be found here: **www.kcg-kiel.org**.

1 Introduction

Inventory investment plays a crucial role in generating volatility in the output of individual firms and, at the aggregate level, in international trade and GNP.[1] One of the reasons is that changes in inventory investment may amplify demand shocks so that the variance of production or trade may exceed the variance of sales. This *bullwhip effect*, as it has become known in the operations management literature, has been observed in various supply chains and widely discussed in that literature at least since Forrester (1961).[2] Following Metzler (1941), the bullwhip effect has also featured prominently in the macroeconomic literature, where it has been shown that in many different industries aggregate production is more volatile than aggregate sales.[3] The *bullwhip effect* has also been invoked to explain why, during the recent financial crisis, international trade proved to be more volatile than GDP in the sense that the volume of trade dropped much more and much faster than GDP but also recovered more quickly.[4] In short, it is essential to understand inventory investment, and the bullwhip effect in particular, if one wants to understand firm behavior, analyze the macroeconomics of business cycles, explain the volatility of international trade relative to GDP, or devise strategies for supply chain management (Blinder and Maccini, 1991, p.74; Cachon et al., 2007).

The current paper seeks to examine the microeconomic foundations of the bullwhip effect by constructing a model to explore the conditions under which this effect may arise from the inventory investment decisions of firms in a simple supply chain consisting of an upstream firm (hereafter, for the sake of concreteness, called the *manufacturer*) and a downstream firm (hereafter called the *retailer*). In particular, we ask how the volatility of retail sales

[1]Blinder and Maccini (1991), for instance, find that 87% of the drop in GNP during the average postwar recession in the United States was accounted for by the fall in inventory investment.

[2]The term 'bullwhip effect' first appears in the academic literature in Lee et al. (1997a, b). Hammond (1994) observes an amplification of demand variability in the case of Barilla pasta. Procter & Gamble found that "the diaper orders issued by the distributors have a degree of variability that cannot be explained by consumer demand fluctuations alone." (See Lee et al. 1997a, p.546). The bullwhip effect has also been observed at a soup manufacturer (Lee et al., 1997b).

[3]The term bullwhip effect to describe the phenomenon where the variance of production or trade exceeds the variance of sales has entered the economics literature more recently; see, for instance, Zavacka (2012) and Altomonte et al. (2013). Elsewhere it has been referred to as production counter-smoothing (Kahn, 1987).

Industry level evidence for the United States comes from Blinder and Maccini (1991) for various industries, Blanchard (1983) and Kahn (1992) for the automotive industry, Blinder (1981) for retailing, West (1986) for aggregate manufacturing, and more recently from Wen (2005a). See Fair (1989) for evidence against the bullwhip effect in several industries.

[4]See, for instance, Alessandria et al. (2011), Altomonte et al. (2013), and Novy and Taylor (2014).

and manufacturer production (or shipments), and thus the occurrence of the bullwhip effect, depend on how inventory control is organized in the supply chain. We compare two forms of organization: a vertically integrated supply chain in which inventory control is centralized, and a decentralized supply chain, in which inventory is controlled by the retailer.

Comparing how the bullwhip effect may evolve under different organizational forms is both theoretically interesting and empirically relevant. First, the theoretical literature on vertical control of inventory shows that incentives to invest in inventory and thus the size of inventory generally differ between vertically integrated and decentralized supply chains (see Krishnan and Winter (2007, 2010), and Qu et al. (2017)). Given this relationship between inventory investment and the organization of the supply chain, it seems pertinent to ask whether the volatilities of sales and production—and thus the occurrence of the bullwhip effect—also differ depending on how the supply chain is organized.

Second, several lines of empirical research invoke either differences or changes in inventory control to explain empirical observations regarding the volatility of production or trade. Altomonte et al. (2013) show, using French microdata, that during the financial crisis not only was international trade more volatile than GDP, but that trade between affiliates of multinational enterprises was much more volatile than arm's-length trade. They attribute this to a bullwhip effect occurring within vertically integrated supply chains (i.e., within multinationals) but not in decentralized ones. The macroeconomic literature on the "Great Moderation" attributes an observed decline in aggregate economic volatility after 1985 (and before the recent financial crisis) at least partly to changes in inventory control (see Wen, 2005b, and Davis and Kahn, 2008), although the nature of these changes is not spelled out very precisely.

By constructing a model of inventory investment in supply chains that combines essential features from the microeconomic theories of the bullwhip effect and from theories of vertical control of inventory we seek to provide a framework that may ultimately help to better understand these empirical findings. Two elements of our model are common to both sets of theories: First, inventory has to be ordered from the manufacturer and production has to take place before goods can be sold by the retailer; we assume, for simplicity, that inventory ordered this period only arrives next period. The retailer's sales are hence from inventory. Second, goods have to be ordered and produced before demand is known. If the retailer orders too little and stocks out, sales are lost. Hence there is a stockout-avoidance motive for holding inventory.

A third element of our model comes from the literature on the bullwhip effect: Demand may be positively correlated across periods so that observed demand this period may provide a signal about demand next period and hence about how much inventory should be ordered for next period. The seminal paper by Kahn (1987) explains why this may give rise to the bullwhip effect. An intuitive explanation is as follows: If demand today turns out to be high, the retailer responds by raising sales and by placing orders over and above what would be needed to replenish its inventory, because it takes the positive demand shock as a signal that demand will also be high tomorrow. However, if demand today turns out to be low, the retailer reduces sales and runs down its inventory in anticipation of low demand tomorrow. Due to this *demand persistence effect*, inventory investment is positively correlated with the retailer's sales and, as a result, the variance of retail orders (and hence manufacturer production) exceeds the variance of retail sales.

The paper by Kahn explores the inventory investment of a monopolist. We extend Kahn's model to a supply chain composed of a manufacturer and a retailer, which allows us to consider how, in a decentralized supply chain, the manufacturer adjusts the producer price in response to a demand shock. We show that the manufacturer will raise the producer price in response to a positive demand shock and lower the producer price in response to a negative demand shock. This *price effect* tends to dampen the variance of production. In fact, we show that with sufficiently small persistence of demand shocks, the *price effect* dampens the variance of production so much that it becomes smaller than the variance of retail sales so that the bullwhip effect is reversed. Following the macroeconomic literature we refer to this case as *production smoothing*. Only when the *demand persistence effect* is big enough to dominate the *price effect* do we observe the bullwhip effect in a decentralized supply chain. By contrast, in a vertically integrated supply chain, in which such price adjustments do not occur, we observe the bullwhip effect for any positive persistence of demand shocks, which mirrors Kahn's results for a monopolist.[5]

In modelling the vertical control of inventory we follow in particular Qu et al. (2017). That paper examines how the volume and intertemporal allocation of inventory differ between supply chains with centralized inventory control and decentralized supply chains, in which inventory is controlled by competitive retailers, respectively an exclusive retailer. Their focus is on how the incentive problems associated with inventory control can best be solved so that the supply chain's aggregate profit can be maximized. By contrast, we explore the effect of different forms of inventory control on the

[5] Kahn's model has been extended to a supply chain and adapted for the operations management literature by Lee et al. (1997a). However, in Lee et al. (1997a), as in the more recent models of the bullwhip effect in the operations management literature (e.g., Warburton, 2004; Gilbert, 2005) prices are typically taken as fixed. These papers hence do not feature a price effect.

See also Reagan (1982) who provides a rigorous mathematical characterization of the optimal pricing and inventory decisions of a monopolist but does not compute or compare the variances of sales and production.

variances of sales and production, and thus on the transmission of demand shocks in supply chains. The theoretical challenge we have to overcome in the current paper is to find a way to compare variances of production and sales not just with each other but across different forms of inventory control. The simple solution we provide can be viewed, from a microeconomic modelling perspective, as a key innovation of the paper.

The rest of the paper is organized as follows. In Section 2 we introduce the model, provide formal definitions of the bullwhip effect and production smoothing, and prove a lemma that shows how we can compare the variances of manufacturer production and retail sales. Section 3 contains results for a general-demand-function specification. The implicit assumptions we make in this section to simplify the analysis are explored in Section 4. There we provide closed-form solutions for a linear demand function and provide explicit sufficient conditions for the assumptions made in Section 3. In Section 5 we explore the robustness of our results to changes in retail market structure and the pricing schemes the manufacturer may use in a decentralized supply chain. We focus on one extension, in particular, namely allowing the manufacturer to use a two-part tariff instead of a simple linear tariff. This extension turns out to be non-trivial, because there is no longer a price effect. We are able to show that all or our results continue to hold, albeit for a different reason. Section 6 concludes, and the appendix contains proofs.

2 The Model

Consider a supply chain consisting of a manufacturer and a retailer. The retailer sells to consumers in two periods, $t = 1, 2$, with consumer demand in each period characterized by an inverse demand function $p(s_t) + \varepsilon_t$, where s_t denotes sales, ε_t is a demand shock, and $p' < 0$, $sp' + p > 0$, $p' + sp'' < 0$. The demand shock may be serially correlated: $\varepsilon_1 = v_1$ and $\varepsilon_2 = \rho\varepsilon_1 + v_2$, where $E(v_t) = 0$, v_t is i.i.d. with cumulative distribution function $F(\cdot)$ and support on $[v_l, v_h]$, and $\rho \in [0, 1]$ measures the persistence of the demand shock. The revenue function is denoted by $R(s_t, \varepsilon_t) = [p(s_t) + \varepsilon_t] s_t$, the marginal revenue is

$$\frac{\partial}{\partial s_t} R(s_t, \varepsilon_t) = p(s_t) + s_t p'(s_t) + \varepsilon_t \equiv m(s_t) + \varepsilon_t.$$

We use the following assumption below:

$$m' + sm'' < 0, \tag{A1}$$

which is satisfied, for instance, for linear demand and whenever the marginal revenue function is not too convex.

All sales are from inventory, which means that the retailer has to place orders with the manufacturer, and goods have to be produced, before the realization of demand has been observed. We assume, in particular, that goods have to be ordered and produced one period before they can be sold. Hence sales in period 1 come from inventory ordered and produced in period 0. Sales in period 2 come either from goods ordered and produced in period 1, or from inventory left over from period 1.

We keep the cost structure as simple as possible. The manufacturer incurs a constant marginal cost, denoted by c, and the marginal cost of retailing is normalized to zero. We further assume that there is no discounting between periods and no cost for storing inventory between periods.

We can now summarize the order of moves as follows: In period 0, the manufacturer sets a producer price P_0, and the retailer orders an amount q_0. In period 1, the order arrives and, after observing the demand shock ε_1, the retailer decides how much to sell in period 1, $s_1 \leq q_0$. After observing ε_1 and s_1, the manufacturer sets producer price P_1, and the retailer then decides how much to order for period 2, q_1. In period 2, the order from period 1 arrives and the demand shock ε_2 is observed. The retailer then sells $s_2 \leq q_1 + (q_0 - s_1)$.

At the heart of our analysis is a comparison of the variance of sales, $Var\left[s_1(\varepsilon_1)\right]$, with the variance of production, $Var\left[q_1(\varepsilon_1)\right]$. If $Var\left[q_1(\varepsilon_1)\right] > Var\left[s_1(\varepsilon_1)\right]$, we say that the supply chain experiences a *bullwhip effect*. *Production smoothing* is said to occur, if $Var\left[q_1(\varepsilon_1)\right] < Var\left[s_1(\varepsilon_1)\right]$. The following Lemma helps us compare these variances by showing how we can link them to the sensitivity of sales and production with respect to demand shocks. It thus represents the key mathematical insight that we apply to prove our results.

Lemma 1 *Let $t_1(\varepsilon_1)$ and $t_2(\varepsilon_1)$ be functions of ε_1. If $t_1'(\varepsilon_1) > t_2'(\varepsilon_1) > 0$ for any ε_1, then $Var\left[t_1(\varepsilon_1)\right] > Var\left[t_2(\varepsilon_1)\right]$.*

Proof: See Appendix.

3 The Bullwhip Effect and Production Smoothing in Supply Chains

3.1 Vertically Integrated Supply Chain

Suppose manufacturing and retailing are vertically integrated, so that the manufacturer controls inventory. The sequence of decisions is then as follows. In period 0, the manufacturer produces q_0. In period 1, after observing ε_1, the manufacturer sells $s_1 \leq q_0$, and produces q_1. In period 2, the manufacturer sells $s_2 \leq q_1 + (q_0 - s_1)$.

To simplify the analysis we make three assumptions that we explore further in the next section. There we explain that the assumptions are satisfied, if the demand shocks are below a certain threshold. First, we assume that the manufacturer sells all inventory on hand in period 2, i.e., $s_2 = q_1 + (q_0 - s_1)$. Second, we assume that the manufacturer produces a sufficient quantity in period 0 to make sure that there is no stockout in period 1. Third, like in the paper by Kahn (1987), we assume that $q_1 > 0$.

Regarding the manufacturer's choice of q_1, it turns out to be convenient to derive this indirectly by solving the maximization problem with respect to sales in period 2, s_2:

$$\max_{s_2} R(s_2, \rho\varepsilon_1) - c\left[s_2 - (q_0 - s_1)\right].$$

The first-order condition,

$$m(s_2) + \rho\varepsilon_1 = c, \tag{1}$$

determines sales in period 2, $s_2(c, \rho\varepsilon_1)$, from which we can compute

$$q_1 = s_2(c, \rho\varepsilon_1) + s_1 - q_0. \tag{2}$$

Consider next how the manufacturer would choose s_1 after ε_1 has been observed. It is easy to see that the manufacturer's optimal choice is determined by setting revenue from selling an additional unit of inventory in period 1 equal to the cost of replacing it:

$$m(s_1) + \varepsilon_1 = c. \tag{3}$$

Writing the manufacturer's optimal sales in period 1 as $s_1(c, \varepsilon_1)$, we observe from (3) that $\frac{\partial s_1(c,\varepsilon_1)}{\partial \varepsilon_1} > 0$.

We know from Lemma 1 that there is a bullwhip effect if $\frac{\partial q_1(\varepsilon_1)}{\partial \varepsilon_1} > \frac{\partial s_1(c,\varepsilon_1)}{\partial \varepsilon_1} > 0$. Noting that q_0 is independent of ε_1, since it is chosen before ε_1 is observed, we obtain from (2) and (1)

$$\frac{\partial q_1(\varepsilon_1)}{\partial \varepsilon_1} - \frac{\partial s_1(c,\varepsilon_1)}{\partial \varepsilon_1} = \frac{\partial s_2(c,\varepsilon_1)}{\partial \varepsilon_1} = -\frac{\rho}{m'(s_2)}. \tag{4}$$

For later reference we call the impact of ε_1 on s_2 displayed in (4) the *demand persistence effect*. It implies that production is more sensitive to demand shocks than sales for any $\rho > 0$, which leads directly to the following result:

Proposition 1 *A vertically integrated supply chain experiences a bullwhip effect if $\rho > 0$.*

This result is our version of Kahn's (1987) finding that a monopolist experiences a bullwhip effect, if there is positive demand persistence. The behavior of a vertically integrated supply chain is simply the same as that of a monopolist. However, as we show in the next subsection, positive demand persistence is not sufficient to generate a bullwhip effect in a decentralized supply chain.

3.2 Decentralized Supply Chain

To characterize the equilibrium in a decentralized supply chain we stick to the implicit simplifying assumptions from the previous case, namely that the manufacturer sells all inventory on hand in period 2, i.e., $s_{r2} = q_{r1} + (q_{r0} - s_{r1})$, where the subscript r denotes the decentralized case, where inventory is controlled by the retailer; that there is no stockout in period 1; and that $q_{r1} > 0$. We provide sufficient conditions for these assumptions to hold in the next section.

To determine whether there exists a bullwhip effect, we again have to examine the derivative of s_{r2} with respect to ε_1. The retailer's choice of s_{r2} now solves

$$\max_{s_{r2}} R\left(s_{r2}, \rho\varepsilon_1\right) - P_{r1}\left[s_{r2} - \left(q_{r0} - s_{r1}\right)\right].$$

The first-order condition w.r.t. s_{r2} is

$$m\left(s_{r2}\right) + \rho\varepsilon_1 = P_{r1}, \tag{5}$$

which implicitly determines $s_{r2}\left(P_{r1}, \rho\varepsilon_1\right)$. The *demand persistence effect* is obviously still at work here: for given P_{r1} an increase in ε_1 raises s_{r2} for any $\rho > 0$, which implies that production is more sensitive to demand shocks than sales so that the supply chain experiences a bullwhip effect.

However, we want to show that there now exists an offsetting effect, the *price effect*: An increase in ε_1 raises P_{r1}, which in turn reduces s_{r2}, since

$$\frac{\partial s_{r2}}{\partial P_{r1}} = \frac{1}{m'\left(s_{r2}\right)} < 0. \tag{6}$$

In other words, the price effect makes production less sensitive to demand shocks than sales, which implies *production smoothing*.

To see that P_{r1} is increasing in ε_1, consider the manufacturer's choice of P_{r1} to maximize its period-1 expected profit

$$E_1\left(\pi_{r1}\right) = \left(P_{r1} - c\right)\left(s_{r2} + s_{r1} - q_{r0}\right).$$

The first-order condition w.r.t. P_{r1} is given by

$$s_{r2} + s_{r1} - q_{r0} + \left(P_{r1} - c\right)\frac{1}{m'\left(s_{r2}\right)} = 0, \tag{7}$$

which implicitly determines $P_{r1}\left(q_{r0} - s_{r1}, s_{r2}, c\right)$. After observing ε_1, the retailer chooses s_{r1} to equalize marginal revenue to next period's expected marginal revenue, which we know from (5) is equal to P_{r1}:

$$m\left(s_{r1}\right) + \varepsilon_1 = P_{r1}. \tag{8}$$

The positive effect of ε_1 on P_{r1} can be now confirmed by totally differentiating equations (5), (7), and (8), and then using Cramer's Rule to obtain

$\partial P_{r1}/\partial \varepsilon_1$ (see the Proof of Proposition 2 in the appendix). The intuition for this positive effect is simple: For given P_{r1} and q_{r0}, an increase in ε_1 raises s_{r1}, which reduces the excess inventory the retailer can take into period 2 and hence induces the retailer to increase the order q_{r1}. The manufacturer responds to this greater demand by raising P_{r1}.

We can now see why the price effect makes production less sensitive to demand shocks than sales. As (6) tells us, the increase in P_{r1} resulting from a positive demand shock dampens the rise in q_{r1} more than the rise in s_{r1}. Similarly, the decrease in P_{r1} associated with a negative demand shock means that q_{r1} falls by less than s_{r1}.

Whether a decentralized supply chain exhibits a bullwhip effect or production smoothing is determined by the relative strength of the demand persistence effect and the price effect. If $\rho = 0$ so that there is zero demand persistence, then the price effect implies that the supply chain exhibits production smoothing. At the opposite extreme, with $\rho = 1$ and thus perfect demand persistence, the demand persistence effect dominates the price effect, so that the supply chain exhibits a bullwhip effect. To see why this is the case, notice that with $\rho = 1$ sales in period 1, s_{r1}, are equal to expected sales in period 2, s_{r2}, which implies that $q_{r1} = 2s_{r1} - q_{r0}$. It follows that orders in period 1 are twice as sensitive to demand shocks as sales: In case of a positive demand shock, the retailer not only has to increase the order to replenish inventory after greater than expected sales in period 1, it also has to order more so that it can satisfy the expected greater demand in period 2. We may hence state:

Proposition 2 *A decentralized supply chain exhibits a bullwhip effect if $\rho = 1$, and production smoothing if $\rho = 0$.*

Proof: See Appendix.

We have shown that under two extreme conditions, namely $\rho = 0$ and $\rho = 1$, production smoothing, respectively a bullwhip effect, will occur. It is natural to try to establish that there exists a cutoff value ρ^*, such that production smoothing is obtained if $\rho < \rho^*$ and a bullwhip effect is obtained if $\rho > \rho^*$. However, within our general demand function framework, this would require an additional assumption regarding the sign of $m'''(s)$ to ensure that the difference between the variances of production and sales is monotonically increasing in ρ. But such a sign restriction has no meaningful economic interpretation. With a linear inverse demand function, to which we turn in the next section, we can easily ensure monotonicity of the variances and hence prove the existence of ρ^*.

4 Closed-Form Solutions

In this section we work with a linear inverse demand function $p(s_t) = 1 - s_t + \varepsilon_t$. We do this for two reasons. First, we want to derive a sufficient condition to ensure that the implicit assumptions we made in the previous section, namely that all inventory is sold in period 2, there are no stockouts in period 1, and the quantity ordered in period 1 is strictly positive, are satisfied. We show that such a condition is given by

$$d < \overline{d} = \min\left\{\frac{1-c}{3(2-\rho)}, \frac{c}{\sqrt{2}+\rho\left(\sqrt{2}-1\right)}\right\}. \tag{A2}$$

Second, we want to obtain a closed-form solution for ρ^*, such that production smoothing is obtained for $\rho < \rho^*$ and a bullwhip effect for $\rho > \rho^*$.

4.1 Vertically Integrated Supply Chain

Using (1), (2) and (3), we can establish that desired sales in periods 1 and 2 are $s_1 = \frac{1-c+\varepsilon_1}{2}$ and $s_2 = \frac{1+\rho\varepsilon_1-c}{2}$, respectively, from which we can confirm that $\frac{\partial s_2}{\partial \varepsilon_1} > 0$ for $\rho > 0$. This guarantees a bullwhip effect for any $\rho > 0$, as stated in Proposition 1. The manufacturer can achieve optimal sales and avoid a stockout in period 1 for any realization of ε_1 by producing in period 0 a quantity equal to $q_0 = \frac{1-c+d}{2}$. By adjusting production in period 1 to the realization of ε_1 and producing a quantity equal to

$$q_1 = \frac{1+(1+\rho)\varepsilon_1 - c - d}{2}$$

the manufacturer ensure optimal expected sales in period 2.

It is now easy to show that $q_1 > 0$ for $d < \overline{d}$. We can also confirm that the manufacturer sells all inventory in period 2 so that $s_2 = q_1 - (q_0 - s_1)$, because $d < \overline{d}$ is sufficient for marginal revenue in period 2 to be positive for all realizations of v_1 and v_2.

4.2 Decentralized Supply Chain

As shown in the Appendix in the proof of Proposition 3, the closed-form solutions for production and sales in a decentralized supply chain are:

$$q_{r1}(\varepsilon_1, \rho) = \frac{2(1-c)}{9} + \frac{(1+\rho)}{6}\varepsilon_1, \tag{9}$$

$$s_{r1}(\varepsilon_1, \rho) = \frac{5(1-c)}{18} + \frac{(2-\rho)}{6}\varepsilon_1. \tag{10}$$

From these solutions we can immediately verify Proposition 2 by comparing $\frac{\partial q_{r1}}{\partial \varepsilon_1} = \frac{1+\rho}{6}$ and $\frac{\partial s_{r1}}{\partial \varepsilon_1} = \frac{2-\rho}{6}$ for $\rho = 0, 1$. However, we can go a step further by noting that $\frac{\partial q_{r1}}{\partial \varepsilon_1}$ is increasing and $\frac{\partial s_{r1}}{\partial \varepsilon_1}$ is decreasing in ρ, which implies:

Proposition 3 *Suppose that demand is linear, and that $d < \overline{d}$. Then in a decentralized supply chain the variance of production is increasing in ρ; and the variance of sales is decreasing in ρ.*

Proof: See Appendix.

This result means that there must exist a critical value ρ^*, such that for $\rho < \rho^*$ the price effect dominates the demand persistence effect and the supply chain exhibits production smoothing, whereas for $\rho > \rho^*$ the demand persistence effect dominates the price effect and the supply chain exhibits a bullwhip effect. In fact, we can compute ρ^* explicitly by setting $\frac{\partial q_{r1}}{\partial \varepsilon_1} = \frac{\partial s_{r1}}{\partial \varepsilon_1}$. We may hence state:

Corollary 1 *In a decentralized supply chain there exists a cutoff value $\rho^* = 1/2$, such that production smoothing occurs if $\rho < \rho^*$, and a bullwhip effect occurs if $\rho > \rho^*$.*

Another immediate consequence of Proposition 3 in the case where $\rho > \rho^*$ is given by:

Corollary 2 *In a decentralized supply chain the bullwhip effect is increasing in ρ.*

Finally it is interesting to compare the variances of production and sales across organizational forms. Comparing $\frac{\partial q_{r1}}{\partial \varepsilon_1}$ and $\frac{\partial s_{r1}}{\partial \varepsilon_1}$ with the respective values for a vertically integrated supply chain, we see that they are both smaller for any value of ρ, which is obviously due to the presence of a price effect in the decentralized supply chain. We may hence conclude:

Proposition 4 *Suppose that demand is linear, and that $d < \overline{d}$. Then the variances of both production and sales are greater in a vertically integrated than in a decentralized supply chain.*

5 Extension: Decentralized Supply Chain with Two-part Tariffs

In our analysis of a decentralized supply chain we assumed that the manufacturer uses linear pricing to sell goods to the retailer. While this is a realistic description of pricing schemes in many chains and thus often taken for granted in the operations management literature (see Krishnan and Winter, 2011), there will certainly be supply chains that adopt more sophisticated pricing schemes. One of the reasons is that linear pricing implies double marginalization and thus a vertical price distortion that, in principle, could be dealt with by using a two-part tariff consisting of a per-unit price and a fixed fee or transfer to the manufacturer. We show in this section that our earlier results continue to hold, even if the manufacturer adopts two-part tariffs. What makes this an especially interesting case to analyze is that there is no longer a price effect to dampen the variance of production and thus to lead to production smoothing for sufficiently low values of ρ. In fact, as we will demonstrate below, the manufacturer will set the per-unit price in period 1 equal to marginal cost and thus no longer adjust the producer price to the demand shock. Production smoothing will instead be driven by a different intertemporal pattern of production.[6]

We continue with same model as in the previous section, making the additional assumption that v_t is uniformly distributed on $[-d, d]$. The only difference as far as decisions are concerned is that in each period $t = 0, 1$ the manufacturer now sets a two-part tariff, consisting of a producer price P_{wt} and a fixed fee or transfer T_{wt}, where the subscript w denotes the two-part-tariff case.

Consider the retailer's choice of q_{w1} for given P_{w1}. Assuming that all inventory is sold in period 2 so that $s_{w2} = q_{w1} + q_{w0} - s_{w1}$, the retailer's maximization problem can be written as

$$\max_{q_{w1}} (1 - q_{w1} - q_{w0} + s_{w1} + \rho\varepsilon_1)(q_{w1} + q_{w0} - s_{w1}) - P_{w1}q_{w1} - T_{w1}.$$

The first-order condition yields

$$q_{w1} = \frac{1 - P_{w1} + \rho\varepsilon_1}{2} - (q_{w0} - s_{w1}), \tag{11}$$

and hence an expected retail price in period 2 equal to $p_2 = \frac{1+P_{w1}+\rho\varepsilon_1}{2}$. The retailer's expected profit in period 1 can hence be calculated as

$$\begin{aligned} E(\pi_{w2}) &= \left(\frac{1 + P_{w1} + \rho\varepsilon_1}{2}\right)\frac{1 - P_{w1} + \rho\varepsilon_1}{2} \\ &\quad - P_{w1}\left[\frac{1 - P_{w1} + \rho\varepsilon_1}{2} - (q_{w0} - s_{w1})\right] - T_{w1} \\ &= \frac{(1 - P_{w1} + \rho\varepsilon_1)^2}{4} + P_{w1}(q_{w0} - s_{w1}) - T_{w1}. \end{aligned}$$

Notice that in period 1 the retailer can guarantee itself an expected profit of at least

$$\pi^{out} \equiv [1 - (q_{w0} - s_{w1}) + \rho\varepsilon_1](q_{w0} - s_{w1}), \tag{12}$$

by not ordering any inventory in period 1, and simply selling in period 2 its excess inventory, $q_{w0} - s_{w1}$, at the resulting expected retail price

[6] The derivation of equilibrium production and sales in this section follows Qu et al. (2017). As that paper makes clear, we could also extend our current model to a multi-retailer setting. In such a setting the *price effect* would still occur, as long as the manufacturer uses a linear pricing scheme. Our key results would thus be unchanged. However, in the presence of a two-part-tariff there is no need for the manufacturer to use more than one retailer.

$[1-(q_{w0}-s_{w1})+\rho\varepsilon_1]$. In fact, even if the retailer orders a positive quantity in period 1, the manufacturer can only set T_{w1} so as to extract the retailer's profit net of π^{out}.

The manufacturer thus chooses (P_{w1}, T_{w1}), such that $T_{w1} = E(\pi_{w2}) - \pi^{out}$ and P_{w1} maximizes

$$\max_{P_{w1}} (P_{w1}-c)\, q_{w1} + \frac{(1-P_{w1}+\rho\varepsilon_1)^2}{4} + P_{w1}(q_{w0}-s_{w1}) - \pi^{out}.$$

Using (11) in the corresponding first-order condition yields $P_{w1} = c$. Thus, the two-part tariff eliminates the price effect.

Now consider the retailer's optimal choice of s_{w1}: At the margin a unit sold in period 1 has to yield the same revenue as holding on to that unit and selling it in period 2. That, is the marginal revenue in period 1, $MR_1 = 1-2s_{w1}+\varepsilon_1$, has to equal expected marginal revenue in period 2, which is determined by the outside option and given by $-\frac{d\pi^{out}}{ds_{w1}} = 1-2(q_{w0}-s_{w1}) + \rho\varepsilon_1$. We hence obtain

$$s_{w1} = \frac{q_{w0}}{2} + \frac{(1-\rho)\,\varepsilon_1}{4}. \tag{13}$$

Notice that the retailer thus sells in period 1 only half the inventory acquired in period 0, adjusted for a term that depends on ε_1, and holds on to the rest to boost π^{out}.

In period 0, as is straightforward to see (and shown formally in the appendix), the manufacturer will use T_{w0} to extract the retailer's total expected profit, including the outside profit π^{out} that the retailer can retain in period 1. To capture the retailer's maximum profit, the manufacturer optimally sets $P_{w0} = c + \left(\frac{3}{2}-\sqrt{2}\right) d\,(1+\rho)$, which implies that the quantity produced in period 0 equals $q_{w0} = 1-c-\left(\frac{3}{2}-\sqrt{2}\right) d\,(1+\rho)$.[7]

What this implies for sales in period 1 can be determined by using $q_{w0} = 1-c-\left(\frac{3}{2}-\sqrt{2}\right) d\,(1+\rho)$ in (13) to obtain

$$s_{w1}(\varepsilon_1,\rho) = \frac{1-c-\left(\frac{3}{2}-\sqrt{2}\right) d\,(1+\rho)}{2} + \frac{(1-\rho)\,\varepsilon_1}{4}. \tag{14}$$

If $\rho = 1$, so that the expected demand shock in period 2, $E(\varepsilon_2)$, is exactly equal to the observed demand shock in period 1, ε_1, the retailer will set $s_{w1}(\varepsilon_1,\rho) = \frac{q_{w0}}{2}$ independently of ε_1. Thus the variance of sales is zero at $\rho = 1$, and given that $\frac{\partial^2 s_{w1}}{\partial\rho\partial\varepsilon_1} < 0$, we see that it increases as ρ is reduced below 1 (an analogous proof has been shown in the appendix in the Proof of Proposition 3).

[7] We show below that this quantity is large enough to cover retail sales in both periods 1 and 2 if ε_1 is small. This can be seen most easily, if we we let $d = 0$. In this case, the manufacturer would set $P_{w0} = c$, and the retailer would order $q_{w0} = 1-c$, which is exactly equal to optimal monopoly sales over two periods.

Turning to q_{w1}, we show in the appendix that the retailer does not order any inventory in period 1, if $\varepsilon_1 < \varepsilon_{w1}^z = -\left(3-2\sqrt{2}\right)d$. This is intuitive: Given that the retailer has ordered enough inventory in period 0 to cover expected sales in both period 1 and period 2, there is no need to order more, when demand in period 1 turns out to be low. Only if $\varepsilon_1 > \varepsilon_{w1}^z$, does the retailer order a positive amount in period 1, which we show in the appendix is equal to $(1+\rho)\frac{(3-2\sqrt{2})d+\varepsilon_1}{4}$; this amount is required to make sure that it has enough inventory in period 2 to realize optimal expected sales. Thus we obtain

$$q_{w1}(\varepsilon_1,\rho) = \begin{cases} 0 & \text{if } \varepsilon_1 \in [-d, \varepsilon_{w1}^z] \\ (1+\rho)\frac{(3-2\sqrt{2})d+\varepsilon_1}{4} & \text{if } \varepsilon_1 \in [\varepsilon_{w1}^z, d] \end{cases}. \tag{15}$$

By inspecting (15), we can see how the intertemporal pattern of retail orders and thus manufacturer output dampens the variance of production. In particular, the fact that a large quantity of output is produced in period 0 implies that for any demand shock in the interval $[-d, \varepsilon_{w1}^z]$, $q_{w1}(\varepsilon_1,\rho)$ is equal to 0; only if $\varepsilon_1 \in [\varepsilon_{w1}^z, d]$ does production vary with the demand shock. Hence there is a positive probability, equal to $F(\varepsilon_{w1}^z)$, that production is constantly equal to zero. From (14) and (15), we see that for $\rho = 0$, the variance of production has to be smaller than the variance of retail sales, since $\frac{\partial q_{w1}}{\partial \varepsilon_1} \leq \frac{\partial s_{w1}}{\partial \varepsilon_1}$, and in the interval $[-d, \varepsilon_{w1}^z]$ this inequality is strict. For $\rho = 1$, the opposite has to be true, as the variance of production is positive and the variance of retail sales is zero.

Moreover, notice that the variance of production is increasing in ρ, due to $\frac{\partial^2 q_{w1}}{\partial \rho \partial \varepsilon_1} \geq 0$, and this inequality is strict in the interval $[\varepsilon_{w1}^z, d]$. Since the variance of retail sales is decreasing in ρ, we know that there exists a critical value $\hat{\rho}$, such that production smoothing occurs if $\rho < \hat{\rho}$ and a bullwhip effect occurs if $\rho > \hat{\rho}$. As before, we can also show that the bullwhip effect is increasing in ρ, and the variances of production and retail sales are smaller in a decentralized supply chain with two-part-tariffs than in a vertically integrated supply chain.

6 Conclusion

In this paper we constructed a simple model of a supply chain consisting of an upstream firm (the manufacturer) and a downstream firm (the retailer) to study how final-demand volatility is transmitted upstream in the chain. In particular, we wanted to know under which circumstances this volatility is enhanced so that the variance of upstream production is greater than the variance of downstream sales (the *bullwhip effect*), or dampened so that the variance of production becomes smaller than the variance of downstream sales (*production smoothing*).

We showed that the transmission of demand shocks, and thus the occurrence of a bullwhip effect or production smoothing, depends on the interplay of two factors, namely the way in which inventory control is organized and the degree of persistence of demand shocks. If the supply chain is vertically integrated so that inventory control is centralized, then a bullwhip effect occurs for any positive degree of demand persistence. In a decentralized supply chain, the bullwhip effect appears only if the persistence exceeds a critical value. Otherwise the decentralized supply chain exhibits production smoothing.

The economic mechanism behind the bullwhip effect in a centralized supply chain is the same as that explored by Kahn (1987): as long as demand shocks are persistent, a positive demand shock today raises production by more than sales, because not only has inventory to be replenished to account for today's greater sales, but inventory also has to rise in anticipation of greater demand tomorrow. In case of a negative demand shock, this mechanism works the other way round so that production decreases by more than sales. As a result of this *demand persistence effect,* the variance of production exceeds the variance of sales. Moreover, we show that, if there is a bullwhip effect, then it becomes stronger the greater is the persistence of demand shocks.

In a decentralized supply chain, the demand persistence effect is not sufficient to induce a bullwhip effect, because it is counteracted by a *price effect*, namely the manufacturer's producer price adjustment. In fact, if the persistence of demand is sufficiently weak, the price effect dominates the demand persistence effect and the supply chain exhibits production smoothing. Why this happens can be seen most easily, if the demand persistence is zero. Thus, when there is a positive demand shock, the retailer sells more today which raises its demand for inventory for next period. The manufacturer reacts by increasing the producer price. The opposite happens in case of a negative demand shock: because the retailer has more excess inventory, its demand for additional inventory for next period falls, and the manufacturer responds by reducing the producer price. This price effect dampens the variance of production relative to the variance of sales so that production smoothing occurs.

Interestingly, the results of the paper continue to hold, if the manufacturer is allowed to use a two-part tariff instead of a simple linear producer price. This is the case despite the fact that there is no price effect, as the manufacturer finds it optimal to keep the producer price in period one equal to marginal cost and thus not adjust it in response to demand shocks. Production smoothing in a decentralized supply chain now comes about due to a different equilibrium order pattern. In particular, the retailer orders such a large quantity in period zero that it only has to reorder goods in period one if demand happens to be sufficiently big. In other words, there is a positive probability that production in period one is zero and thus unresponsive to the demand shock, which naturally dampens the variance of production.

By providing a simple microeconomic model of the transmission of demand shocks in a supply chain our paper may help shed light on a wide range of empirical studies. Most directly related are studies of how demand volatility is passed upstream through different echelons of a supply chain from retailers to wholesalers to various layers of production. This research goes back at least to Holt et al. (1968) and has been taken up most prominently by the operations management literature already cited above. We show, among other things, that the extent of the pass-through of volatility depends on where and how in the chain inventory is managed.

Our model can rationalize not only the observation that international trade was more volatile than GDP during and after the financial crisis, but also the micro-level evidence by Altomonte et al. (2013) that intra-firm trade was more volatile than arm's-length trade, which directly corresponds to our result that shipments by the manufacturer should be more volatile in a vertically integrated than in a decentralized supply chain. The explanation we provide is that in a decentralized supply chain the volatility of shipments will be dampened by the manufacturer's price adjustment.

Another line of research, as mentioned in the introduction, concerns the macroeconomic literature on the "Great Moderation" that has identified a decline in aggregate economic volatility after 1985 (and before the recent financial crisis). As already mentioned above, this observation is partly attributed to changes in inventory control (see Wen, 2005b, and Davis and Kahn, 2008), but also to a decline in the persistence of demand shocks (Ramey and Vine, 2006). Our model could rationalize the decline in output volatility through two possible micro-level channels. First, a moderation of production volatility would come about for a given level of persistence, if inventory control in a supply chain becomes decentralized. This might, for instance, happen if the retailer, or more generally a downstream firm, outsourced production to an independent supplier, but kept control of inventory. Second, in a decentralized supply chain, a change from the bullwhip effect to production smoothing would occur, if the persistence of demand shocks declined, which is what Ramey and Vine argue has happened in the US automobile industry.

7 Appendix

7.1 Proof of Lemma 1

To prove Lemma 1 we make use of the following result from Gurland (1967):

Lemma 2 *Let X be a random variable, and let n, m be continuous functions on $\mathbb{R}$. If n is monotonically increasing and m monotonically decreasing, then*

$E[n(X)m(X)] \leq E[n(X)]E[m(X)]$. *If n, m are both monotonically increasing or decreasing, then $E[n(X)m(X)] \geq E[n(X)]E[m(X)]$. Moreover, in both cases, if both functions are strictly monotone, the inequality is strict.*

Consider the function $g(\lambda,\varepsilon_1) = \lambda t_1(\varepsilon_1) + (1-\lambda)t_2(\varepsilon_1)$. Easily we can check that $E\left(\frac{\partial[g(\lambda,\varepsilon_1)]^2}{\partial\lambda}\right) = \frac{\partial E\left[g(\lambda,\varepsilon_1)^2\right]}{\partial\lambda}$, and $\frac{\partial}{\partial\lambda}E[g(\lambda,\varepsilon_1)] = E\left[\frac{\partial}{\partial\lambda}g(\lambda,\varepsilon_1)\right]$. Consider how the variance of $g(\lambda,\varepsilon_1)$ changes with λ: $\frac{\partial}{\partial\lambda}Var[g(\lambda,\varepsilon_1)] = E\left[2g(\lambda,\varepsilon_1)\frac{\partial}{\partial\lambda}g(\lambda,\varepsilon_1)\right] - E[2g(\lambda,\varepsilon_1)]E\left[\frac{\partial}{\partial\lambda}g(\lambda,\varepsilon_1)\right]$.

Furthermore, $\frac{\partial}{\partial\varepsilon_1}g(\lambda,\varepsilon_1) = \lambda\frac{\partial t_1(\varepsilon_1)}{\partial\varepsilon_1} + (1-\lambda)\frac{\partial t_2(\varepsilon_1)}{\partial\varepsilon_1} > 0$ and $\frac{\partial}{\partial\varepsilon_1}\left[\frac{\partial}{\partial\lambda}g(\lambda,\varepsilon_1)\right] = \frac{\partial t_1(\varepsilon_1)}{\partial\varepsilon_1} - \frac{\partial t_2(\varepsilon_1)}{\partial\varepsilon_1} > 0$. Using Lemma 2, we can state that $\frac{\partial}{\partial\lambda}Var[g(\lambda,\varepsilon_1)] > 0$, which implies $Var[g(1,\varepsilon_1)] > Var[g(0,\varepsilon_1)]$, that is, $Var[t_1(\varepsilon_1)] > Var[t_2(\varepsilon_1)]$.

7.2 Proof of Proposition 2

Consider first the second-order condition for the manufacturer's choice of P_{r1}. This condition is satisfied if:

$$\begin{aligned} & 2\frac{\partial s_{r2}}{\partial P_{r1}} + (P_{r1}-c)\frac{\partial^2 s_{r2}}{\partial(P_{r1})^2} \\ = \quad & \left[m'(s_{r2})\right]^{-1}\left[2-(P_{r1}-c)m''(s_{r2})\left[m'(s_{r2})\right]^{-2}\right] < 0 \end{aligned}$$

which requires that $2-(P_{r1}-c)m''(s_{r2})[m'(s_{r2})]^{-2} > 0$. In fact, equation (7) means that

$$P_{r1} - c = -m'(s_{r2})(s_{r2}+s_{r1}-q_{r0}) < -m'(s_{r2})s_{r2}$$

and therefore

$$1-(P_{r1}-c)m''(s_{r2})\left[m'(s_{r2})\right]^{-2} > 1 + s_{r2}m''(s_{r2})\left[m'(s_{r2})\right]^{-1} > 0,$$

in which the last inequality follows from assumption (A1).

Totally differentiating the equations (5), (7) and (8), we obtain

$$\begin{bmatrix} 0 & m'(s_{r2}) & -1 \\ 1 & 1-\frac{(P_{r1}-c)m''(s_{r2})}{[m'(s_{r2})]^2} & \frac{1}{m'(s_{r2})} \\ m'(s_{r1}) & 0 & -1 \end{bmatrix}\begin{bmatrix} \partial s_{r1} \\ \partial s_{r2} \\ \partial P_{r1} \end{bmatrix} = \begin{bmatrix} -\rho \\ 0 \\ -1 \end{bmatrix}\partial\varepsilon_1$$

Using Cramer's Rule, we have

$$\begin{aligned}
\frac{\partial s_{r1}}{\partial \varepsilon_1} &= \frac{(\rho-1)\left\{1-(P_{r1}-c)\left[m'(s_{r2})\right]^{-2} m''(s_{r2})\right\}-1}{m'(s_{r2})+\left\{2-(P_{r1}-c)\left[m'(s_{r2})\right]^{-2} m''(s_{r2})\right\} m'(s_{r1})} \\
\frac{\partial s_{r2}}{\partial \varepsilon_1} &= \frac{(1-\rho)-\rho m'(s_{r1})\left[m'(s_{r2})\right]^{-1}}{m'(s_{r2})+\left\{2-(P_{r1}-c)\left[m'(s_{r2})\right]^{-2} m''(s_{r2})\right\} m'(s_{r1})} \\
\frac{\partial P_{r1}}{\partial \varepsilon_1} &= \frac{\rho\left\{1-(P_{r1}-c)\left[m'(s_{r2})\right]^{-2} m''(s_{r2})\right\} m'(s_{r1})+m'(s_{r2})}{m'(s_{r2})+\left\{2-(P_{r1}-c)\left[m'(s_{r2})\right]^{-2} m''(s_{r2})\right\} m'(s_{r1})}
\end{aligned}$$

We can show that $\frac{\partial s_{r1}}{\partial \varepsilon_1} > 0$ and $\frac{\partial q_{r1}}{\partial \varepsilon_1} = \frac{\partial s_{r1}}{\partial \varepsilon_1} + \frac{\partial s_{r2}}{\partial \varepsilon_1} > 0$. For $\rho = 0$, we find that $\frac{\partial s_{r2}}{\partial \varepsilon_1} < 0$, and for $\rho = 1$ we obtain $\frac{\partial s_{r2}}{\partial \varepsilon_1} > 0$. Using Lemma 1 completes the proof.

7.3 Proof of Proposition 3

Using (5), (7) and (8), we can derive explicit expressions for s_{r1}, s_{r2}, P_{r1} and also q_{r1} as functions of q_{r0} and ε_1. They are given by:

$$\begin{aligned}
s_{r1} &= \frac{1}{6}\left[1+(2-\rho)\varepsilon_1-c\right]+\frac{1}{3}q_{r0} \\
s_{r2} &= \frac{1}{6}\left[1+(2\rho-1)\varepsilon_1-c\right]+\frac{1}{3}q_{r0} \\
P_{r1} &= \frac{2+(1+\rho)\varepsilon_1+c}{3}-\frac{2}{3}q_{r0} \\
q_{r1} &= \frac{1}{6}\left[2+(1+\rho)\varepsilon_1-2c\right]-\frac{1}{3}q_{r0}
\end{aligned}$$

A stockout occurs if $s_{r1} \geq q_{r0}$, which would happen if ε_1 were to exceed the cutoff value $\varepsilon_{r1}^s = \frac{4q_{r0}-1+c}{2-\rho}$. We will show below that ε_{r1}^s is always larger than d, given Assumption (A2). But first we derive a closed-form solution for q_{r0}.

In period 0, the retailer maximizes:

$$\max_{q_{r0}} \int_{-d}^{d} \left[(1-s_{r1}+\varepsilon_1)s_{r1}+(1-s_{r2}+\rho\varepsilon_1)s_{r2}-P_{r1}q_{r1}\right] dF(\varepsilon_1) - P_{r0}q_{r0}.$$

The first-order condition (FOC) with respect to q_{r0} is given by

$$\int_{-d}^{d} \left[\frac{8+(4+4\rho)\varepsilon_1+c}{9}-\frac{8}{9}q_{r0}\right] dF(\varepsilon_1) - P_{r0} = 0.$$

In the period 0, the manufacturer's revenue is hence equal to

$$P_{r0}q_{r0} = \int_{-d}^{d} \left[\frac{8 + (4 + 4\rho)\,\varepsilon_1 + c}{9} - \frac{8}{9} q_{r0} \right] q_{r0} dF(\varepsilon_1),$$

and the manufacturer's total expected profit across the two periods is given by

$$\begin{aligned} & E\left[(P_{r0} - c)\,q_{r0} + (P_{r1} - c)\,q_{r1}\right] \\ = & \int_{-d}^{d} \left[\frac{8 + (4 + 4\rho)\,\varepsilon_1 - 8c}{9} - \frac{8}{9} q_{r0} \right] q_{r0} dF(\varepsilon_1) \\ & + \int_{-d}^{d} \left[\frac{2 + (1 + \rho)\,\varepsilon_1 - 2c}{3} - \frac{2}{3} q_{r0} \right] \left[\frac{2 + (1 + \rho)\,\varepsilon_1 - 2c}{6} - \frac{1}{3} q_{r0} \right] dF(\varepsilon_1). \end{aligned}$$

Deriving the FOC with respect to q_{r0} and simplifying yields $q_{r0} = \frac{1-c}{3}$. Using this solution we can rewrite $q_{r1}(\varepsilon_1)$ and $s_{r1}(\varepsilon_1)$ to obtain (10) and (9).

Now notice from (10) and (9) that $\frac{\partial q_{r1}}{\partial \varepsilon_1} = \frac{1+\rho}{6}$ is increasing in ρ (i.e. $\frac{\partial^2 q_{r1}}{\partial \rho \partial \varepsilon_1} > 0$), and that $\frac{\partial s_{r1}}{\partial \varepsilon_1} = \frac{2-\rho}{6}$ is decreasing in ρ (i.e. $\frac{\partial^2 s_{r1}}{\partial \rho \partial \varepsilon_1} < 0$). Therefore for any ρ_1 and ρ_2, such that $0 < \rho_1 < \rho_2 < 1$, we have $0 < \left.\frac{\partial q_{r1}}{\partial \varepsilon_1}\right|_{\rho=\rho_1} < \left.\frac{\partial q_{r1}}{\partial \varepsilon_1}\right|_{\rho=\rho_2}$ and $\left.\frac{\partial s_{r1}}{\partial \varepsilon_1}\right|_{\rho=\rho_1} > \left.\frac{\partial s_{r1}}{\partial \varepsilon_1}\right|_{\rho=\rho_2} > 0$, which means that $\frac{\partial q_{r1}(\varepsilon_1,\rho_2)}{\partial \varepsilon_1} > \frac{\partial q_{r1}(\varepsilon_1,\rho_1)}{\partial \varepsilon_1} > 0$ and $\frac{\partial s_{r1}(\varepsilon_1,\rho_1)}{\partial \varepsilon_1} > \frac{\partial s_{r1}(\varepsilon_1,\rho_2)}{\partial \varepsilon_1} > 0$ for any ε_1. Lemma 1 hence tells us that $Var\left[q_{r1}(\varepsilon_1,\rho_2)\right] > Var\left[q_{r1}(\varepsilon_1,\rho_1)\right]$ and $Var\left[s_{r1}(\varepsilon_1,\rho_2)\right] < Var\left[s_{r1}(\varepsilon_1,\rho_1)\right]$

Using these solutions we can verify that for $d < \overline{d}$ there is no stock-out, because $\varepsilon_{r1}^s = \frac{4q_{r0}-1+c}{2-\rho} = \frac{1-c}{3(2-\rho)} > d$, and that $s_{r1}, s_{r2}, q_{r1}, P_{r1} > 0$. Moreover $MR_2 = P_{r1} + v_2 > 0$ for all realizations of v_1 and v_2 if $d < \overline{d}$.

7.4 Derivation of $q_{w1}(\varepsilon_1, \rho)$ and $s_{w1}(\varepsilon_1, \rho)$

Using (13) in (11) we find

$$q_{w1} = \frac{2(1-c) - 2q_{w0} + (1+\rho)\,\varepsilon_1}{4}.$$

Notice that $q_{w1} = 0$, if

$$\frac{2(1-c) - 2q_{w0} + (1+\rho)\,\varepsilon_1}{4} \leq 0,$$

and thus $q_{w1} > 0$ if ε_1 exceeds a cutoff value given by

$$\varepsilon_{w1}^z = -\frac{2 - 2c - 2q_{w0}}{(1+\rho)}.$$

If $q_{w1} = 0$, the retailer will use its initial inventory of q_{w0} to cover sales in both periods, $s_{w1}^z + s_{w2}^z = q_{w0}$, and will allocate q_{w0} between periods so as to equalize marginal revenue in in period 1 with the expected marginal revenue in period 2:

$$1 - 2s_{w1}^z + \varepsilon_1 = 1 - 2s_{w2}^z + \rho\varepsilon_1.$$

Solving for s_{w1}^z, we obtain the same solution as in (13), namely

$$s_{w1}^z = \frac{q_{w0}}{2} + \frac{(1-\rho)\,\varepsilon_1}{4}.$$

Hence whether it orders goods or not in period 1, the retailer's total expected profit in period 0 is given by

$$E\left\{(1 - s_{w1} + \varepsilon_1)\, s_{w1} + \pi^{out} - P_{w0}q_{w0} - T_{w0}\right\},$$

which can be rewritten as

$$\int_{-d}^{d} \left[\left(1 - \frac{1}{2}q_{w0}\right) q_{w0} + \frac{(1-\rho)^2\,\varepsilon_1^2}{8}\right] \frac{1}{2d} d\varepsilon_1 - P_{w0}q_{w0} - T_{w0}.$$

The FOC yields $q_{w0} = 1 - P_{w0}$, which means that the retailer earns an expected profit of

$$E\left[\frac{(1-P_{w0})^2}{2} + \frac{(1-\rho)^2\,\varepsilon_1^2}{8}\right] - T_{w0}.$$

Since the manufacturer will use T_{w0} to capture the retailer's total expected profit, including the π^{out} that the retailer retains in period 1, the manufacturer thus maximizes

$$\max_{P_{w0}} (P_{w0} - c)(1 - P_{w0}) + E\left[\frac{(1-P_{w0})^2}{2} + \frac{(1-\rho)^2\,\varepsilon_1^2}{8}\right]$$
$$+ \int_{\varepsilon_{w1}^z}^{d} \left[\frac{(1 - c + \rho\varepsilon_1)^2}{4} + c(q_{w0} - s_{w1}) - \pi^{out}\right] \frac{1}{2d} d\varepsilon_1$$

where the first line corresponds to the manufacturer's expected profit in period 0. The second line represents the expected profit that can be extracted from the retailer in period 1, given that an order is placed only if the realization of demand in period 1 is higher than ε_{w1}^z. The FOC is

$$-(P_{w0} - c) - \frac{1}{2}\int_{\varepsilon_{w1}^z}^{d} \left[c - P_{w0} - \frac{(1+\rho)\,\varepsilon_1}{2}\right] \frac{1}{2d} d\varepsilon_1 = 0$$

Solving this FOC, we have

$$P_{w0} = c + \left(\frac{3}{2} - \sqrt{2}\right) d\,(1+\rho)\,.$$

This implies

$$q_{w0} = 1 - c - \left(\frac{3}{2} - \sqrt{2}\right) d\,(1+\rho)\,.$$

Using this solution in the expression of ε_{w1}^{z}, we find that $\varepsilon_{w1}^{z} = -\left(3 - 2\sqrt{2}\right) d$ for any value of ρ. Thus, if $\varepsilon_1 < \varepsilon_{w1}^{z}$, the retailer does not order any inventory in period 1 ($q_{w1} = 0$); and if $\varepsilon_1 > \varepsilon_{w1}^{z}$, the retailer orders

$$q_{w1} = (1+\rho)\,\frac{\left(3 - 2\sqrt{2}\right) d + \varepsilon_1}{4}.$$

We can now obtain (15) and (14).

Finally, notice from (14) that $s_{w1} > 0$, if $d < \frac{(1-c)}{\left[\left(2-\sqrt{2}\right)-\left(\sqrt{2}-1\right)\rho\right]}$, which also guarantees that there is no stockout in period 1 (i.e. $q_{w0} > s_{w1}$) and that $s_{w2} > 0$. Next, consider the implicit assumption that the retailer sells all inventory in period 2 so that $s_{w2} = q_{w1} + q_{w0} - s_{w1}$, which requires that $MR_2 > 0$. If $\varepsilon_1 > \varepsilon_{w1}^{z}$, we have $s_{w2} = \frac{1-c+\rho\varepsilon_1}{2}$, so that $MR_2 > 0$ provided that $d < c$; if $\varepsilon_1 < \varepsilon_{w1}^{z}$, we have $s_{w2}^{z} = \frac{1-c-\left(\frac{3}{2}-\sqrt{2}\right)d(1+\rho)}{2} - \frac{(1-\rho)\varepsilon_1}{4}$, so that $MR_2 > 0$ provided that $d < \frac{c}{\sqrt{2}+\rho\left(\sqrt{2}-1\right)}$. It is easy to check that these inequalities are satisfied under Assumption (A2).

References

[1] Altomonte, C., Di Mauro, F., Ottaviano, G., Rungi, A., Vicard, V. (2013). Global Value Chains during the Great Trade Collapse: A Bullwhip Effect? in: Sjoerd Beugelsdijk, Steven Brakman, Hans van Ees and Harry Garretsen (eds.), *Firms in the International Economy*, MIT Press, Cambridge MA.

[2] Alessandria, George, Joseph P. Kaboski and Virgiliu Midrigan (2011). US Trade and Inventory Dynamics, *American Economic Review* 101(3), 303-307.

[3] Blanchard, Olivier J. (1983). The Production and Inventory Behavior of the American Automobile Industry, *Journal of Political Economy* 91, 365-400.

[4] Blinder, Alan S. (1981). Retail Inventory Investment and Business Fluctuations. *Brookings Papers on Economic Activity* 2, 443-505.

[5] Blinder, Alan S. and L. Maccini (1991). Taking Stock: A Critical Assessment of Recent Research on Inventories, *Journal of Economic Perspectives* 5, 73-96.

[6] Cachon, Gerard, Taylor Randall and Glen M. Schmidt (2007). In Search of the Bullwhip Effect, *Manufacturing and Service Operations Management* 9, 457-479.

[7] Davis, Steven J. and James A. Kahn (2008). Interpreting the Great Moderation: Changes in the Volatility of Economic Activity at the Macro and Micro Levels, *Journal of Economic Perspectives* 22(4), 155-180.

[8] Fair, Ray C. (1989). The Production-Smoothing Model is Alive and Well, *Journal of Monetary Economics* 24, 353-370.

[9] Forrester, J. W. (1961). *Industrial Dynamics*, MIT Press, Cambridge, Massachusetts.

[10] Gilbert, Kenneth (2005). An ARIMA Supply Chain Model, *Management Science* 51(2), 305-310.

[11] Gurland, J., 1967. An Inequality Satisfied by the Expectation of the Reciprocal of a Random Variable, *The American Statistician* 21(2), 24-25.

[12] Hammond, J. 1994. Barilla SpA (a). HBS Case9-694-046, Harvard Business School, Boston, MA.

[13] Holt, Charles C., Franco Modigliani and John P. Shelton (1968). The Transmission of Demand Fluctuations through a Distribution and Production System, the TV-Set Industry, *Canadian Journal of Economics* 1(4), 718-739.

[14] Kahn, James A. (1987). Inventories and the Volatility of Production, *American Economic Review* 77 (September), 667-679.

[15] Kahn, James A. (1992). Why is Production More Volatile Than Sales? Theory and Evidence on the Stockout-Avoidance Motive for Inventory-Holding, *Quarterly Journal of Economics* xx, 481-510.

[16] Krishnan, Harish and Ralph Winter, 2011. The Economic Foundations of Supply Chain Contracting, *Foundations and Trends in Technology, Information and Operations Management* 5, 3-4, 147-309.

[17] Krishnan, Harish and Ralph Winter, 2010. Inventory Dynamics and Supply Chain Coordination, *Management Science* 56(1), 141-147.

[18] Krishnan, Harish and Ralph Winter, 2007. Vertical Control of Price and Inventory, *American Economic Review* 97(5), 1840-57.

[19] Lee, H., V. Padmanabhan, S. Whang (1997a). Information Distortion in a Supply Chain: The Bullwhip Effect, *Management Science* 43, 546-558.

[20] Lee, H., V. Padmanabhan, S. Whang (1997b). The Bullwhip Effect in Supply Chains *MIT Sloan Management Review* 38, 93-102.

[21] Metzler, Lloyd A. (1941). The Nature and Stability of Inventory Cycles, *Review of Economics and Statistics* 23, 113-125.

[22] Novy, Dennis and Alan M. Taylor (2014). Trade and Uncertainty, NBER Working Paper 19941.

[23] Qu, Zhan, Horst Raff and Nicolas Schmitt (2017). Incentives through Inventory Control in Supply Chains, CESifo Working Paper 6705.

[24] Ramey, Valerie A. and Daniel J. Vine (2006). Declining Volatility in the U.S. Automobile Industry, *American Economic Review* 96(5), 1876-1889.

[25] Reagan. Patricia B. (1982). Inventory and Price Behaviour, *Review of Economic Studies* XLIX, 137-142.

[26] Warburton, Roger D. H. (2004). An Analytical Investigation of the Bullwhip Effect, *Production and Operations Management, 13, 150-160.*

[27] Wen, Yi (2005a). Understanding the Inventory Cycle, *Journal of Monetary Economics* 52, 1533-1555.

[28] Wen, Yi (2005b). Durable good inventories and the volatility of production: explaining the less volatile U.S. economy, Federal Reserve Bank of St. Louis Working Paper 2005-047.

[29] West, Kenneth D. (1986). A Variance Bound Test of the Linear Quadratic Inventory Model, *Journal of Political Economy* 94, 374-401.

[30] Zavacka, Veronika (2012). The Bullwhip Effect and the Great Trade Collapse, European Bank for Reconstruction and Development Working Paper 148.