THE VARIABILITY OF AGGREGATE DEMAND WITH \((S, s)\) INVENTORY POLICIES

By Andrew S. Caplin

This paper develops a general theory of the aggregate implications of \((S, s)\) inventory policies. It is shown that \((S, s)\) policies add to the variability of demand, with the variance of orders exceeding the variance of sales. Overall, the \((S, s)\) theory contradicts the widely held notion that retail inventories act as a buffer, protecting manufacturers from fluctuating sales.

1. INTRODUCTION

In 1951, Arrow, Harris, and Marschak [3] introduced the \((S, s)\) form of inventory policy. The policies are designed for retailers of finished goods, who face economies of scale when placing orders with their suppliers. To pursue an \((S, s)\) inventory policy, the retailer establishes a lower stock point \(s\), and an upper stock point \(S\). No order is placed until inventories fall to \(s\) or below, whereupon they are restored to the maximum of \(S\). A general proof of the optimality of these \((S, s)\) inventory policies was provided by Scarf [13].

At the microeconomic level, the model has been extensively investigated. Formulae are available to compute optimal policies (e.g., Ehrhardt [6]), and these policies are widely used in industry (e.g., Schwartz (ed.) [14]). In addition, the model has been extended to increasingly complex demand environments (e.g., Karlin and Fabens [11]).

In contrast, little is known about the macroeconomic implications of \((S, s)\) policies. Several recent papers have begun to correct this deficiency. Akerlof has suggested that pursuit of constant threshold money holding policies of the \((S, s)\) variety might be responsible for the observed low short-run income elasticity of the demand for money (Akerlof [1] and Akerlof and Milbourne [2]). In the operations research literature, Ehrhardt, Schultz, and Wagner [7] analyzed the demand environment of a wholesaler supplying several retailers. They required that the distinct retailers have independent sales, ruling out the analysis of common factors in sales. Finally, simulation results of Blinder [4] suggested a role for the \((S, s)\) model in understanding retail sector inventories. However, the theoretical difficulties with the model remained unresolved. Blinder commented:

If firms have a technology that makes the \(S, s\) rule optimal, aggregation across firms is inherently difficult. Indeed it is precisely this difficulty which has prevented the \(S, s\) model from being used in empirical work to date (Blinder [4, p. 459]).

In this paper we present a general theory of the aggregate implications of \((S, s)\) policies. Our central finding is that \((S, s)\) policies add to the variability of demand, with the variance of orders exceeding the variance of sales. This result holds even in the presence of common factors in retail sales. In addition, a close connection

---

1 This is a revised version of portions of my Ph.D. dissertation. I am deeply grateful to Professors Herbert Scarf, William Nordhaus, and James Tobin, and to Philip White, for their many important contributions. I thank two anonymous referees for their constructive suggestions. Financial assistance was provided by the Sloan Foundation.
is established between the size of individual orders, and the variability of aggregate orders.

In the simplest cases, this connection between bulk ordering and variability is transparent. Consider a single retailer observed periodically. Assume that the retailer's minimum order size exceeds the maximum level of sales during any given period. In this case, the retailer either orders an amount in excess of the peak level of sales, or orders nothing. Thus orders are more volatile than sales. In the general case the analysis is more intricate, but the result survives.

The \((S, s)\) theory thus contradicts the widely-held notion that retail sector inventories act as a buffer, protecting manufacturers from fluctuating sales. The evidence supports the \((S, s)\) theory. Blinder [4] found aggregate deliveries to retailers to be more variable than sales, while Holt [9] reported similar findings in a study of the television industry.

The paper is organized as follows. The basic model of the retail sector is presented in Section 2. In Section 3 the long-run behavior of inventories is characterized. Sections 4 and 5 apply earlier results to analyze the variance of retail sector orders. The formulae of Sections 4 and 5 are readily testable. Finally, Section 6 outlines possible extensions of the model.

2. THE MODEL

The model contains an arbitrary number, \(n\), of distinct retailers. Each retailer stocks a single good, which it sells at a fixed price per unit. At this price future demand is uncertain. The demand faced by retailer \(k\) during period \(t\) is denoted \(y_k^t\). It is assumed that future demands of a given retailer are drawn independently through time from a fixed probability density. The nature of the probability density generating future demands varies from retailer to retailer, depending on the particulars of their situation, such as their size and location. For convenience, it is assumed that goods are defined in integral quantities.

The retailers face economies of scale in placing their orders. Specifically, they obtain supplies at a constant price per unit, but face an additional fixed cost per order, regardless of the amount to be delivered. It is assumed that retailers continuously monitor their inventory levels, and consequently can place orders at any point in time. It is also assumed that orders are delivered instantaneously; in Caplin [5] the model is extended to allow for lags in the delivery of orders. The total of orders placed by retailer \(k\) during period \(t\) is denoted \(r_k^t\).

If a potential customer arrives when the retailer has no stock, the order is backlogged and met from a later delivery. Such stockouts are discouraged by a convex penalty function.

The inventory position of retailer \(k\) at the start of period \(t\) is denoted \(x_k^t\). If there are stocks on hand, \(x_k^t\) is positive; if orders have been backlogged, \(x_k^t\) is negative. The progress of the \(k\)th retailer's inventory position obeys Identity 2.1.

2.1. INVENTORY IDENTITY: For each retailer \(k\), \(1 \leq k \leq n\), the variable \(x_k^{t+1}\) depends on \(x_k^t\), \(y_k^t\), and \(r_k^t\) according to the identity,

\[ x_k^{t+1} = x_k^t + r_k^t - y_k^t. \]
Overall, the incentives to holding inventories are the desire to avoid more frequent payment of the delivery charge, and to avoid costly stockouts. The disincentives to stock holding are that purchase and storage of goods involves the payment of carrying costs, and the sacrifice of interest available on an alternative asset.

In discrete time, optimal policy is of a particularly simple form. The policy is to establish an upper stock point $S_k$ and a lower stock point $s_k$. No order is placed until stocks fall to $s_k$ or below, whereupon they are restored to their maximum of $S_k$. The optimality of such $(S, s)$ inventory policies was established for the multiperiod model by Scarf [13], and by Iglehart [10] for the infinite horizon model.

In the present paper, the retailers can place orders at any point in time. By extension from the discrete time optimal policies, it is assumed that each retailer continuously operates an $(S, s)$ policy. In addition, it is assumed that demands are received in a well-defined order, so that inventories cannot fall from $(S_k + 1)$ to $(s_k - 1)$ without passing through level $s_k$.

In this case, retailer $k$ orders the quantity $(S_k - s_k) = Q_k$ the instant stocks reach the lower limit of $s_k$. The progress of inventories for a retailer applying this form of $(S, s)$ rule is illustrated in Figure 1.

The transition equations for inventories can now be given, describing the progress of inventories through time.

2.2. Transition Equations: For each retailer $k$, $1 \leq k \leq n$, the variables $r_k^t$ and $x_k^{t+1}$ depend on $y_k^t \geq 0$ and $x_k^t$ according to the equations

(a) $r_k^t = m_k^t Q_k$, some integer $m_k^t \geq 0$,
(b) $x_k^{t+1} = s_k + 1$,
(c) $x_k^{t+1} > S_k \rightarrow r_k^t = 0$.

Equation 2.2(a) records that each order placed by retailer $k$ is of size $Q_k$. Equation 2.2(b) reflects the fact that an order is placed the instant inventories reach the lower limit of $s_k$. Finally, Equation 2.2(c) states that if the retailer has placed an order during the period, then inventories cannot be above $S_k$. In
combination, the equations uniquely specify $r_k'$ and $x_k^{t+1}$ as they depend on $y_k^t$ and $x_k^t$.

At this stage, the model contains $n$ retailers operating independently. Some interdependence must be introduced before the model can be applied to macroeconomic issues. In particular, it cannot be assumed that sales of one retailer are independent of sales elsewhere, since this assumption rules out consideration of forces influencing the economy as a whole.

In practice, the level of any retailer's sales depends on many factors, some of which have a general influence. For instance, a short spending spree by consumers raises sales all round. Conversely, a temporary fall in aggregate demand tends to depress sales. On the other hand, if aggregate demand for a product is stable, the various suppliers are fighting for shares of a fixed market, so there is a negative correlation between their sales levels. To achieve the desired level of generality, the joint demand process generating future sales of the $n$ retailers, $\phi(y)$, defined on $y \in \mathbb{Z}^n$, the $n$-dimensional integral lattice, is left unrestricted.

With the introduction of the demand process, the description of the model is complete. Definitions 2.3 summarize the components of the model. The term "$(S, s)$ economy" is introduced to cover the class of models under consideration. To avoid trivialities, it is assumed that $S_k > s_k + 1$.

2.3. Definitions: The $(S, s)$ economy is defined by the number of retailers $n$; $n$ pairs of integers $(S_k, s_k)$ with $S_k > s_k + 1$, all $1 \leq k \leq n$; and a demand process $\phi(y)$ defined on $y \in \mathbb{Z}^n$. Consumer demands arise from independent draws from this density. The progress of inventories is described by Equations 2.2.

3. The General Solution

Analysis of the $(S, s)$ economy involves the elementary theory of Markov chains. This theory can be applied directly to the vector of inventory levels. The progress of this vector, $x^t$, through time is Markovian, because the influence of the past on future stock levels is fully captured in the current stock level. To confirm this, note that the transition equations show $x^{t+1}$ to depend only on $x^t$ and $y^t$, while $y^t$ is independent of its own past values, and thus of $x^{t-1}$.

For the methods of Markov chain theory to be applied, the inventory vectors must first be indexed. Further analysis depends on specifying the transition matrix, $A$, of the Markov chain of inventory vectors. The next stage is to find the state probabilities $\pi(x^t)$ solving the equation system $\pi A = \pi$: the stationary density of the Markov chain of inventory vectors. The stationary density defines the relative frequency with which each inventory vector is observed in indefinitely repeated, independently spaced visits to the economy.

When solving for the stationary density, not all inventory vectors need be considered. Since the stationary density characterizes long-run behavior, transient vectors can be left out of the list of possible states. In particular if a stock vector $x$ is such that some component lies outside the range, $s_k + 1 \leq x_k \leq S_k$, then
Equations 2.2 show that as soon as retailer k's stock level changes, the vector x is never again observed.

The definitions provided in 3.1 below summarize the various components of the Markovian approach to the (S, s) economy.

3.1. Definitions: The state space of the Markov chain of inventory vectors is the set $\tilde{X}$,

$$\tilde{X} = \{x \in \mathbb{Z}_+^n | s_k + 1 \leq x_k \leq S_k, \text{all } 1 \leq k \leq n\}.$$ 

The cardinality of the set $\tilde{X}$ is $C = \prod_{k=1}^{n} Q_k$.

The elements of the set $\tilde{X}$ are ordered in some (arbitrary) manner. The jth state is the element $x^j \in \tilde{X}$, for $1 \leq j \leq C$. The transition matrix of the economy, $A$, is a square matrix of dimension $C$, with $[a_{ij}]$ recording the probability of state $j$ in period $t+1$, following state $i$ in period $t$.

A simple example will serve both to illustrate the method of solution, and to motivate the general result of this section.

Consider the (S, s) economy with two retailers. The retailers operate identical inventory policies, $S_1 = S_2 = 4$ and $s_1 = s_2 = 1$. The demand process $\phi(y_1, y_2)$ assigns the probabilities $\phi(1, 0) = p_1$, $\phi(0, 1) = p_2$, and $\phi(1, 1) = p_3$. With $p_3$ close to 1 there is strong positive correlation between demand levels at the two retailers, while with $p_3$ close to zero the correlation is negative.

The state space of the Markov chain of inventory vectors is illustrated in Figure 2, which also records certain single period transitions between states.

If inventories start period $t$ in state (4, 4), and demand during the period is (1, 0), then retailer 1 loses one unit from stock, while 2 has an unchanged inventory at the end of the period. This explains the arrow marked with $p_1$ (the probability of demand vector (1, 0)) directed from (4, 4) to (3, 4). Similarly, demand of (0, 1) results in a transition from (4, 4) to (4, 3), while demand for one unit from each retailer reduces inventories to (3, 3).
If, however, inventories at the start of period $t$ are in state $(2, 2)$ and demand during the period is $(1, 0)$, then retailer 1 is induced to place an order for 3 units of stock. Thus, with probability $p_1$, stocks of $(2, 2)$ in $t$ are followed by stocks of $(4, 2)$ in $(t+1)$, as in Figure 2.

In Figure 3, similar logic has been applied to all states to complete the single period transition diagram. The remarkable feature of the diagram is the complete symmetry among states. Each state can be entered from precisely three other states, and in all cases one of the entering arrows is vertical ($p_1$), one horizontal ($p_2$), and one diagonal ($p_3$). The symmetry suggests that the stationary density assigns equal likelihood to each state, so that each is observed $1/9$ of the time in the long run.

To confirm this assertion, consider the probability of some particular state, say $(3, 3)$, in $(t + 1)$ given that all states were equiprobable in period $t$. For $(3, 3)$ to occur in $(t + 1)$, the state in $t$ must be $(4, 3)$, $(3, 4)$, or $(4, 4)$. From $(4, 3)$ demand of $(1, 0)$ is required; from $(3, 4)$ demand of $(0, 1)$; from $(4, 4)$ demand of $(1, 1)$. Overall, the probability of $(3, 3)$ at the start of $(t+1)$ is found by adding up the probabilities of these three separate events, yielding

$$\frac{p_1}{9} + \frac{p_2}{9} + \frac{p_3}{9} = \frac{1}{9}$$

as desired.

Similar reasoning applied to the other states shows the uniform probability density to be invariant under the transitions of the $(S, s)$ economy. Thus the equation system $\pi A = \pi$ has $\pi(x^j) = \frac{1}{9}$, all $1 \leq j \leq 9$, as a solution. With any two of the probabilities $p_1$, $p_2$, and $p_3$ strictly positive, this solution is unique, and defines long-run behavior.

The striking conclusion is that the stationary probability density over inventories does not depend on the correlation between demand levels at the two outlets. As long as at least two of $p_1$, $p_2$, and $p_3$ are positive, the inventory levels of the two retailers are mutually independent. Thus, for instance, information on the level of inventories at retailer 1 provides no information on the inventory level of retailer 2, even if their sales are subject to common influences.
The result of the example extends to the general \((S, s)\) economy, provided a certain regularity condition is satisfied. Before proving the main theorem, the regularity conditions must be considered, both in the example and in the general case.

Consider the two retailer economy of the example, with \(\phi(1, 1) = p_3 = 1\). In this case the equation system \(\pi A = \pi\) has multiple solutions. The existence of multiple solutions results in an indeterminacy, with the long-run behavior of inventories being ill-defined.

The problem arises from the lack of communication among the states of the Markov chain. With \(p_3 = 1\) it is impossible, for instance, to arrive at \((4, 2)\) having started at \((4, 4)\). However, with \(0 < p_3 < 1\), all states communicate with one another, so that there is only one probability density solving the equation system \(\pi A = \pi\).

A second difficulty which arises when \(\phi(1, 1) = 1\) is that if the economy is observed every third period, inventories are always in the same state. There is a cycle in the behavior of inventories, with period 3.

To avoid both difficulties, it is assumed that the \((S, s)\) economy satisfies a regularity condition. The definition of regularity is standard in the theory of Markov chains: all states must communicate in an aperiodic fashion (e.g., Feller [8]).

3.2. Assumption: The transition matrix \(A\) is a regular transition matrix.

Proposition 3.3 provides a simple sufficient condition for regularity of the \((S, s)\) economy. The proof is in the Appendix.

3.3. Proposition: Assumption 3.2 is satisfied if there is some \(y \in \mathbb{Z}^n_+\) with \(\phi(y) > 0\), such that for any \(1 \leq k \leq n\), either \(\phi(y + e_k)\) or \(\phi(y - e_k)\) is positive, where \(e_k\) is the \(k\)th unit vector.

These sufficient conditions rely only on the possibility of certain neighboring demand vectors. They would be satisfied when precise pinpointing of sales levels at individual retailers is impossible. Note also that trivial perturbations in the demand process ensure regularity. Regularity Assumption 3.2 is thus innocuous.

The central result of the paper can now be proved.

3.4. Theorem: The stationary density of the Markov chain of inventory vectors assigns probability \(\pi(x^j) = 1/C\) to each state \(x^j \in \bar{X}\).

Proof: To prove the result it suffices that for any \(1 \leq j \leq C\),

\[
(1) \quad \sum_{i=1}^{C} a_{ij} = 1,
\]

i.e., the matrix \(A\) is doubly stochastic.
The Transition Equations 2.2 show that retailer $k$ moves from inventory level $x_k^t$ in $t$ to level $x_k^{t+1}$ in $t+1$ if and only if demand in $t$ takes a value $y_k \geq 0$ satisfying

$$y_k = x_k^t - x_k^{t+1} + m_k Q_k$$

for some nonnegative integer $m_k$.

Applying (2) to all states in $\bar{X}$ and taking the summation yields

$$z_{aij} = \sum_{x' \in \bar{X}} \sum_{m \in \mathbb{Z}_+^n} \phi(x_i^t - x' + m_i Q_i, x_2^t - x_2^t + m_2 Q_2, \ldots, x_n^t - x_n^t + m_n Q_n).$$

To show that the right-hand side of (3) equals 1, it suffices that for the fixed vector $x^t$ and an arbitrarily given vector $y \geq 0$ there is a unique choice of $x' = \hat{x} \in \bar{X}$ and a uniquely specified $m = \hat{m} \in \mathbb{Z}_+^n$, such that equation (2) is satisfied for all coordinates $1 \leq k \leq n$. In this case, for any $y \geq 0$, $\phi(y)$ will appear once and only once on the right-hand side in equation (3), so that the terms sum to unity, as claimed.

Given $y \geq 0$ and $x^t \in \bar{X}$, define $\hat{m} \geq 0$ so that $s_k + 1 \leq x_k^t + y_k - \hat{m}_k Q_k \leq S_k$ for all $1 \leq k \leq n$. Define $\hat{x}$ so that $\hat{x}_k = x_k^t + y_k - \hat{m}_k Q_k$, so that $\hat{x}$ is in $\bar{X}$ by construction. By construction this pair $\hat{m}$ and $\hat{x}$ satisfy equation (2) for all $1 \leq k \leq n$.

It remains to show that $\hat{m}$ and $\hat{x}$ are uniquely determined by $x^t \in \bar{X}$, $y \geq 0$, and satisfaction of equations (2) for all $1 \leq k \leq n$. Any distinct solution $\bar{m}$ and $\bar{x}$ must involve $\bar{m}_k \neq \hat{m}_k$ for some $1 \leq k \leq n$. But in this case satisfaction of (2) requires $\bar{x}_k$ to be either strictly above $S_k$ or strictly below $(s_k + 1)$, contradicting the assumed membership of $\bar{x}$ in $\bar{X}$. This completes the proof. Q.E.D.

Theorem 3.4 has two important implications for the long-run behavior of inventories in the $(S, s)$ economy.

First, each retailer is equally likely to be observed with any of its possible stock levels. Second, in the long-run, the inventory levels of distinct retailers are mutually independent, regardless of the correlation in sales. In this sense, the $(S, s)$ policies serve to insulate inventories from the interdependence in the demand process. It is this second aspect of 3.4 which permits the aggregate implications of $(S, s)$ policies to be characterized.

4. THE VARIANCE OF AN INDIVIDUAL RETAILER'S ORDERS

The orders retailers place during a given period depend both on initial inventory levels, and on demand during the period. The form of this dependence is specified in the transition equations 2.2, which define the vector-valued function $r'(x^t, y^t)$. To study the long-run properties of orders, the relevant probabilities must thus be assigned to pairs $(x^t, y^t)$.

Applying Theorem 3.4, the probability that $x' = x^t \in \bar{X}$ is $1/C$, all $1 \leq j \leq C$. Note in addition that the subsequent vector of consumer demands $y^t = y \in \mathbb{Z}_+^n$ is
independent of \( x' \). Thus, pairs \( x' = x' \in \bar{X} \) and \( y' = y \in \mathbb{Z}_+^n \) occur with relative frequency \( \phi(y)/C \) in the long run. It is these weights which are applied to vectors of orders \( r'(x', y') \) when studying long-run behavior.

In the next two sections, this general method is applied to analyze the variance of orders. The time superscript is suppressed when long-run behavior is being studied. In addition, the subscript \( k \) is suppressed throughout Section 4, in which a single retailer is being considered. Thus \( V(r) \) and \( V(R) \) denote the long-run variance of a single retailer's orders and of aggregate orders respectively.

In the simplest \((S, s)\) economies, \( V(r) \) can be estimated without reference to inventories, as in Proposition 4.1.

4.1. **Proposition:** If \( \sum_{y=0}^{Q-1} \phi(y) = 1 \), then \( V(r) = E(y)[Q - E(y)] \).

**Proof:** Under the assumption, \( r' \) is either \( Q \) or 0. In addition, since \( E(x) \) exists, Identity 2.1 shows that \( E(r) = E(y) \). Thus in the long-run, \( r' \) takes value \( Q \) a proportion \( E(y)/Q \) of the time, so that

\[
V(r) = E(r^2) - [E(y)]^2 = E(y)[Q - E(y)]
\]

as claimed. \( Q.E.D. \)

Unfortunately, this approach cannot be generalized to situations in which orders can take more than two values. The analysis of more realistic cases calls for explicit attention to the behavior of inventories.

Proposition 4.3 provides the general formula for the variance of a single retailer's orders. The formula involves the random variable \( b \), defined as the demand process taken modulo \( Q \).

4.2. **Definition:** For a single retailer, the random variable \( b \) with probability density \( \xi(b) \) is defined by

(a) \( \xi(b) = \sum_{a=0}^{\infty} [\phi(b + aQ)] \) for \( 0 \leq b \leq Q - 1 \),
(b) \( \xi(b) = 0 \) for \( b \geq Q \).

4.3. **Proposition:** \( V(r) - V(y) = E[b(Q - b)] > 0 \).

**Proof:** Note that \( V(r) - V(y) = E(r^2) - E(y^2) \) so that,

\[
V(r) - V(y) = \sum_{a=0}^{\infty} \sum_{b=0}^{Q-1} (E[(r')^2 | y' = aQ + b] - [aQ + b]^2)\phi[aQ + b].
\]

From the Transition Equations 2.2, when \( y' = aQ + b, r' = aQ \) if \( x' > s + b + 1 \), and \( r' = (a + 1)Q \) if \( x' \leq s + b \).
Thus, applying Theorem 3.4,

\[ E[(r')^2 | y' = aQ + b] = (aQ + b)^2 + b(Q - b). \]

Substituting (2) in (1) yields,

\[ E(r^2) - E(y^2) = \sum_{a=0}^{Q-1} \sum_{b=0}^{Q-1} [b(Q - b)] \phi(aQ + b) = E[b(Q - b)]. \]

Finally, \( \xi(0) = 1 \) only if demand always comes in multiples of \( Q \), which is ruled out by regularity Assumption 3.2. Thus \( V(r) - V(y) \) is strictly positive, completing the proof.

The formulae illustrate the close connection between the extent of the micro-economic indivisibility and the variability of orders. Proposition 4.1 suggests that \( V(r) \) increases linearly with the order size \( Q \). The proposition also shows that when orders are placed on a sufficiently large scale, \( V(r) \) can be estimated by a simple formula depending only on the lot size and average sales per period.

In practice, 4.1 applies when successive orders by the retailer are typically separated by several time periods. Caplin [5] exhibits an even simpler formula which applies when the time between successive observations is longer, so that the retailer typically places several orders within each period.

Further discussion of 4.1 and 4.3 is deferred until the problem of aggregation has been addressed.

5. THE VARIANCE OF AGGREGATE ORDERS

Entirely new issues are raised when more than one retailer is under study. Consider the simplest case of two identical retailers, with \( Q_1 = Q_2 = Q \). Assume in addition that \( y'_1 \) and \( y'_2 \) take values close to \( Q/2 \), so that both retailer 1 and retailer 2 order practically every second period.

Even in this simple economy, the variance of aggregate orders depends on how often the retailers order in the same period, and how often they order in different periods. At one extreme, if they almost always order together, then periods in which aggregate orders total \( 2Q \) alternate with periods in which orders are zero. At the other extreme, if the retailers place orders in alternate periods, then total orders are very stable, rarely deviating from the level \( Q \).

This second case presumes a strong negative correlation between the orders placed by the two retailers. Can such correlations persist in the \((S, s)\) economy, and if so, does aggregation result in the variance of the sector's orders falling below that of its sales?

The expansion,

\[ V(r_1 + r_2) - V(y_1 + y_2) = V(r_1) - V(y_1) + V(r_2) - V(y_2) + 2[Cov (r_1, r_2) - Cov (y_1, y_2)] \]
shows that the relationship between $\text{Cov}(y_1, y_2)$ and $\text{Cov}(r_1, r_2)$ is critical. If $\text{Cov}(r_1, r_2)$ is less than $\text{Cov}(y_1, y_2)$, then aggregation serves to smooth the pattern of orders relative to sales. With $\text{Cov}(r_1, r_2)$ above $\text{Cov}(y_1, y_2)$, aggregation actually adds to the variability of orders. In economic terms, the question is how an association between sales at two outlets is reflected in their orders. This issue must be settled before further progress can be made.

The general answer is provided in Theorem 5.1. The result shows that correlations between sales at distinct retailers are transformed into identical correlations between their orders. Thus aggregation in the $(S, s)$ economy is neutral with respect to variance, neither increasing nor diminishing the impact of individual retailers on their own sales processes.

5.1. **Theorem:**

$$V(R) - V(Y) = \sum_{k=1}^{n} [V(r_k) - V(y_k)].$$

**Proof:** In general,

(1) $$V(R) - V(Y) = \sum_{k=1}^{n} [V(r_k) - V(y_k)] + \sum_{j \neq k} \text{Cov}(r_j, r_k) - \text{Cov}(y_j, y_k),$$

where

(2) $$\text{Cov}(r_j, r_k) - \text{Cov}(y_j, y_k) = [E(r_j r_k) - E(y_j y_k)] - [E(r_j)E(r_k) - E(y_j)E(y_k)]$$

$$= E(r_j r_k) - E(y_j y_k),$$

since $E(r_k) = E(y_k)$, all $1 \leq k \leq n$.

To evaluate the right-hand side in (2), expand as

(3) $$E(r_j r_k') = \sum_{y \in \mathbb{Z}_+} [E(r_j r_k'|y' = y) \phi(y)].$$

For a particular $y' = y$, there is one and only one pair of integers $a_j$ and $b_j$, with $0 \leq b_j \leq Q_j - 1$, such that $y_j = a_j Q_j + b_j$, all $1 \leq j \leq n$. Similarly $y_k = a_k Q_k + b_k$ for unique integers $a_k$ and $b_k$, $0 \leq b_k \leq Q_k - 1$. With $y' = y$, the Transition Equations 2.2 show the values $(r_j, r_k)$ to depend on the initial inventory levels, $s_j + 1 \leq x_j \leq S_j$ and $S_k + 1 \leq x_k \leq S_k$ as in (4),

(4) $$(r_j', r_k') = \begin{cases} [a_j Q_j, a_k Q_k] & \text{for } x_j' \geq s_j + b_j + 1, x_k' \geq s_k + b_k + 1 \\ [(a_j + 1)Q_j, a_k Q_k] & \text{for } x_j' \leq s_j + b_j, x_k' \geq s_k + b_k + 1 \\ [a_j Q_j, (a_k + 1) Q_k] & \text{for } x_j' \geq s_j + b_j + 1, x_k' \leq s_k + b_k \\ [(a_j + 1) Q_j, (a_k + 1) Q_k] & \text{for } x_j' \leq s_j + b_j, x_k' \leq s_k + b_k \end{cases}.$$
Applying Theorem 3.5, all pairs of values \((x_j', x_k')\) with coordinates obeying \(s_j + 1 \leq x_j' \leq S_j\) and \(s_k + 1 \leq x_k' \leq S_k\) are equiprobable, so that

\[
E(r_j'r_k'|y' = y) = a_ja_kQ_jQ_k + a_kb_jQ_k + a_kb'_kQ_j + b_kb'_k
\]

\[= (a_jQ_j + b_j)(a_kQ_k + b_k) = y_jy_k.
\]

Substituting (5) in (3) shows that for any \(j \neq k, 1 \leq j, k \leq n\), 
\[E(r_j'r_k') = E(y_j'y_k'),\]
so that \(\text{Cov}(r_j, r_k) = \text{Cov}(y_j, y_k)\), and the proof is complete. \(\text{Q.E.D.}\)

Theorem 5.1 is an aggregation theorem. In determining the increase in variance between sales and orders, interdependences among the retail units can be ignored. Instead, the economy can be considered as consisting of \(n\) separate units, each responsible for a particular magnification of the variance of its own sales process. The magnification of variance in the aggregate is simply the sum of these \(n\) components.

The impact of the theorem can be illustrated for a model 100 retailer economy. There are 80 retailers of type \(A\), each placing orders in lots of size 20, and 20 retailers of type \(B\), placing orders of size 40. All that is known about sales is that type \(A\) retailers sell between 5 and 15 units per period, while type \(B\) retailers sell between 10 and 30 units per period. Even with this limited information it is possible to make close estimates of the increase in variance between retail sector sales and orders.

The first step is to apply Theorem 5.1.

\[V(R) - V(Y) = \sum_{k=1}^{100} [V(r_k) - V(y_k)].\]

To place bounds on individual terms on the right-hand-side in (1), the product \(b(Q - b)\) must be considered, with \(b\) as in 4.2. For a retailer of type \(A\), the product \(b(Q - b)\) is maximized when \(b = 10\), and minimized when \(b = 5\) or 15. Thus \(E^A[b(Q - b)]\) lies between 75 and 100. Similarly, for a retailer of type \(B\), \(E^B[b(Q - b)]\) lies between 300 and 400. Combining these bounds, and applying Proposition 4.3 yields,

\[(80.75 + 20.300) \leq [V(R) - V(Y)] \leq (80.100 + 20.400),\]

\[12,000 \leq [V(R) - V(Y)] \leq 16,000.\]

An important feature of the example is the limited sales information needed to place bounds on \([V(R) - V(Y)]\). For instance, the average sales level \(E(Y)\) is only restricted to the range \(600 \leq E(Y) \leq 1800\).

The example can also be used to illustrate the close connection between order sizes and the magnification of variance. Consider the effect on the variability of orders if all type \(A\) retailers suddenly switch to placing orders in lots of size 30. Applying Proposition 4.3, this change adds between 50 and 150 to the value
of $V(r)$ for an individual retailer of type $A$, and so adds between 4000 and 12,000 to the variance of aggregate orders.

An even more precise estimate of the effects of such an alteration is possible if the expected value of sales is known. For example, assume that each type $A$ retailer has expected sales of 10 units per period. In this case Proposition 4.1 can be applied to show that when a type $A$ retailer changes $Q$ from 20 to 30, $V(r)$ increases by precisely 100. Thus $V(R)$ increases by 8000, so that $20,000 \leq [V(R) - V(Y)] \leq 24,000$.

Overall, the predictions of the $(S, s)$ theory appear readily testable, using data on sales, order sizes and orders. Initial tests might involve small groups of similar retailers, to avoid hazards in the definition of the size of an order.

While specific tests have not yet been performed, the general results of Blinder [4] are promising. He found that for the retail sector as a whole, and seven of eight subsectors, the variance of deliveries to retailers exceeded the variance of final sales. The $(S, s)$ theory provides a plausible explanation for these findings.

6. CONCLUDING REMARKS

The paper had two main goals: first, to understand the macroeconomic implications of the $(S, s)$ model of operations research; second, to explain the empirical finding that retail sector orders are more variable than final sales.

The results have been encouraging on both counts. The aggregate implications of $(S, s)$ policies were assessed in a quite general model of the retail sector. In particular, it was shown that there is a close connection between economies of scale in transactions and the variability of aggregate orders. Thus the $(S, s)$ theory is both workable, and broadly consistent with the facts.

The impact of $(S, s)$ policies on the pattern of production and income depend on the interactions among manufacturers, distributors, and consumers. For instance, with constant returns to scale in production, the pattern of output will closely mirror that of orders, so that production will also be more variable than final sales. However, if manufacturers face steeply increasing marginal costs, as in the buffer stock model (Lovell [12]), it is manufacturers’ inventories rather than production that will be directly affected by the variability of orders.

A fuller analysis of these topics requires the explicit introduction of both consumers and manufacturers into the model. In such closed models, it may be possible to investigate government policies aimed at reducing order size, and thus damping fluctuations in production and employment. The approximation formula of Ehrhardt [6] suggests that one possible policy would be to raise interest rates. The possibility that interest rates might influence the variability of demand by changing the scale of orders is intriguing.

More immediately, the $(S, s)$ theory demonstrates that economies of scale in transactions can provide the basis for a distinctive theory of inventories. There are many other areas of macroeconomics in which transactions costs are prominent, such as the theory of money demand, the theory of price adjustment, and
the theory of wage contracts. Further research extending the analysis to these topics appears warranted.

Harvard University

Manuscript received June, 1984; revision received November, 1984.

APPENDIX

The Appendix contains a proof of Proposition 3.3. It is convenient to term the set of \( y \in Z^n \) with \( \phi(y) > 0 \) as possible demands.

3.3. PROPOSITION: The \((S, s)\) economy is regular if there is a possible demand \( y \) such that, for any \( 1 \leq k \leq n \), either \( y + e_k \) is possible, or \( y - e_k \) is possible.

PROOF: It must be proved that all states in \( \mathbf{X} \) communicate in an aperiodic fashion.

Full communication is proved by picking first arbitrary vectors \( x^1 \) and \( x^2 \) in \( \mathbf{X} \). Lemma A.1 is a useful preliminary. The vector \( Z \) in A.1 is defined in two parts by \( Z_k = \frac{4}{4} - \frac{1}{4} \) if \( 4_k \in X \), and \( Z_k = Q_k + x_k - x_k^1 \) if \( x_k^1 < x_k^1 \).

A.1 LEMMA: \( x^2 \) can be reached from \( x^1 \) if some sequence of possible demands \( y' \), \( 1 \leq i \leq T \), satisfies the conditions that:

\[ y'_T - y' (\text{mod } Q_k) = Z_k, \quad \text{all } 1 \leq k \leq n. \]

PROOF: From the Equations 2.3, note that transitions over \( T \) periods depend only on the cumulative total demand over all \( T \) periods. Thus it suffices that Lemma A.1 hold for \( T = 1 \). This follows immediately from the Transition Equations 2.3, establishing Lemma A.1. Q.E.D.

To apply A.1, arrange indices so that for \( 1 \leq k \leq n \), \( y + e_k \) is possible, while for \( n + 1 \leq k \leq n \), \( y - e_k \) is possible. For \( 1 \leq k \leq n \), repeat \( y + e_k \) a total of \( Z_k \) times. For \( n + 1 \leq k \leq n \), repeat \( y - e_k \) a total of \( (Q_k - Z_k) \) times. Finally, repeat \( y \) a total of \( \sum_{k=1}^{n} y_k (\text{mod } Q_k) = Z_k, \quad \text{all } 1 \leq k \leq n. \)

REFERENCES


