On the Modeling of Demand Spill for a Stochastic Demand System under Competition

Derek Atkins
Operations and Logistics, Sauder School of Business, University of British Columbia, Vancouver, Canada, V6T 1Z2, derek.atkins@sauder.ubc.ca

Xuan Zhao¹
Operations and Decision Sciences, School of Business and Economics, Wilfrid Laurier University, Waterloo, Canada, N2L 3C5, xzhao@wlu.ca

Abstract

When customers for a product from N substitutable alternatives find their first choice sold out, they might “spill” to their second-most preferred products. The existing literature typically assumes an exogenous spill rate. We develop a surprisingly simple model that links the spill rate to economic factors associated with direct demand systems.

Key words: Spill rate; Representative consumer; Quadratic utility function; Index of differentiation/substitutability

¹The corresponding author.
Phone: 519-884-0710 Ext. 2814
Fax: 519-884-0201
1 Introduction and Contribution

The purpose of this paper is to assist in the modeling of spill between sellers or products in stochastic demand systems under competition. Because two key elements of a firm’s competitive strategy are product availability and product variety, the result of this paper contributes in at least two ways. First, it helps sellers to have a better understanding of the spill and actual demand they face. Demand spill usually occurs between sellers selling substitutable products, who have to decide on capacity or inventory in advance of actual demand. A survey by P&G found about 50% of its customers switching to another retailer after a stockout [1]. The inaccurate estimation of spill demand can lead to incorrect inventory decisions, causing service level (availability) to decrease and operational costs to increase. The second reason relates to product variety and assortment. Offering a wide variety of items has many advantages from a marketing perspective, but it can increase the costs of inventory, shipping and other merchandise. Firms need to optimize their assortment of products. Anupindi et al. [2] reported a significant substitution rate for vending machine products and mention a 82%-88% spill for grocery customers. Verbeke et al. [3] found 34% of Coca-Cola customers switch after a stockout. The possibility of substitution across items impacts the demand distribution of each item and customer service. For firms, “[it] may be beneficial to stock only the most popular or most profitable items, or it may make more sense to stock an item that is attractive to many customers as a substitute, even though it may not be the first choice for most customers” ([4], p 20). A wise assortment decision needs to incorporate demand-switching behavior and substitution effects between items.

The core of the problem we address is how demand spill relates to important factors associated with classical economic models that derive demand functions. “One major approach in product differentiation theory is the ‘representative consumer’ model ([5, 6]). This model assumes that the behavior of consumers with different tastes can be described by the choices made by a single individual who has a preference for diversity” ([7], p 461). In the
literature of economics, many utility forms have been assumed to describe a representative consumer; in this paper, we focus on one of the most commonly-used forms—the quadratic utility function—which has been widely used to derive the linear-demand system, a major tool for understanding market demand in economics, marketing and operations ([8–10]).

Given the significant impact of demand substitution on inventory management and assortment decisions, research on demand substitution has been attracting substantial attention (see [11] for a detailed discussion of the literature). There has been extensive literature that explores the impact of demand substitution on inventory decisions for competing firms. Two approaches have been used to model substitution. The first assumes that, when a stock-out occurs, customers are divided according to a proportional rule, and customers remain unsatisfied if the substitute is stocked-out as well. For example, Lippman and McCardle [12] considered both deterministic and random splitting rules. Parlar, Netessine and Rudi, Cachon and Netessine, and Zhao and Atkins [13–16] assumed an exogenous spill rate. The second approach, which was used in [17], explicitly models the dynamic consumer process, so the sequentially incoming consumer purchases the available product that maximizes his or her utility. Based on [17], Hopp and Xu [11] assumed that unsatisfied consumers have multiple attempts at product substitution and approximate the dynamic demand-substitution process with a static fluid network model, which significantly simplifies the technical difficulty faced by modeling dynamic demand spill. There has been also a vast amount of literature on the impact of demand substitution on inventory and assortment decisions of a single firm that carries multiple products, for example, the earlier work [18] and the more recent work [4]. Most of this literature assumes an exogenous spill rate between items. Hopp and Xu [11] applied their demand-spill model in a monopolist’s assortment decision as well.

Although many studies have been done on demand substitution, there is little explicit modeling of consumers’ switching behavior by considering the underlying economics. It is still not clear how demand spill relates to the classical economic models that for decades have been used to describe demand systems. This paper contributes a spill-rate model that
is derived from the classical representative consumer theory that employs a quadratic utility function.

To anticipate our result, we shall prove that the spill rate is given by a function of the parameters describing the representative consumer. We further use the linear-demand function derived from this representative consumer theory, and find a spill-rate formula that is remarkably simple and has nice economic interpretations: the spill rate is highly associated with the degree of product differentiation and self-price sensitivity. This simple but effective spill rate formula will, we hope, facilitate further research on the assortment of products, pricing and inventory management, and will provide a basic step for research on strategic consumer behavior and its impact on operations.

We would like to introduce how the rest of the paper proceeds. Section 2 provides basic assumptions and definitions; Section 3 derives the spill-rate formula, investigates its properties and provides a numerical example; Section 4 provides an application and concludes the paper.

2 Assumptions and Definitions

We summarize the assumptions below.

1. Both the primary (direct) and secondary (spill) markets occur at one period of time: we do not consider situations where retailers might adjust prices for a secondary “spill market.”

2. There is only one chance of spill and unsatisfied consumers choose a substitute based on their preference: If the second most preferred product is sold out, then the consumer remains unsatisfied.

3. Also consistent with the one-period-of-time assumption is that we assume spill consumers to be representative of all consumers; anyone might be a spill. For instance, it could be the case that consumers with very strong preferences might arrive early, leaving fairly indifferent consumers to arrive later and find the product sold out. In this case, fairly indif-
ferent consumers would be disproportionately represented in the spill. We do not consider this detail; spill consumers are representative of all primary consumers.

In order to introduce Assumption 4, we need to set up some notation. The entire consideration set for consumers is denoted by \( \Omega = \{1, 2, ..., N\} \). When product \( j \in \Omega \) is not available, we denote it by \( \Omega \setminus j = \{1, 2, ..., j-1, j+1, ..., N\} \). Consumers have preferences for products given the prices at which they are offered. If a consumer's valuation for the product is no less than its price, then it is acceptable to the consumer. The product that provides the largest consumer surplus is the most preferred. All consumers who prefer product \( i \) the most constitute the direct demand of product \( i \), denoted by \( \tilde{q}_i \). If product \( j \) is not available and the consideration set is \( \Omega \setminus j \), then the direct demand of product \( i \) is denoted by \( \tilde{q}_{i|\Omega \setminus j} \).

We are now ready for Assumption 4.

4. Following the literature that models price-sensitive stochastic demand, we assume that the randomness in demand is price-independent, and we model it in an additive fashion as in [19, 20]: \( \tilde{q}_i = q_i + x_i \), where \( q_i \) is the deterministic portion of the direct demand, and \( x_i \) is the random portion with mean zero. Similarly we assume \( \tilde{q}_{i|\Omega \setminus j} = q_{i|\Omega \setminus j} + x_{i|\Omega \setminus j} \).

5. Although not needed until Section 3, to keep all assumptions together, we assume that the underlying economy can be modeled through the representative consumer theory and that the representative consumer has a quadratic utility function. See Section 3 for more discussion on this.

If a consumer’s most-preferred item \( j \) has run out, he/she might switch to his/her second-most preferred product \( i \). Now \( (\tilde{q}_{i|\Omega \setminus j} - \tilde{q}_i) \) is the number of customers whose first choice is product \( j \) and whose second choice is product \( i \) and would be willing to purchase product \( i \) at \( p_i \) if product \( j \) is not available. This paper is concerned with the spill rate from product \( j \) to product \( i \)—the proportion of those consumers in the direct demand of product \( j \). Using Assumptions 2 and 3, we define the spill rate from product \( j \) to product \( i \) if product \( j \) is not available.
**Definition 1** Define the spill rate as \( \gamma_{ji} = \frac{E[\tilde{q}_i|\Omega_{\neg j} - \tilde{q}_i]}{E[\tilde{q}_j]} \).

In words, the spill rate is defined as the ratio between the expected number of consumers who would switch from \( j \) to \( i \) if \( j \) is not available and the expected direct demand for product \( j \). From the above expression, we must have \( 0 \leq \gamma_{ji} \leq 1 \). As \( E[x_i] = E[x_i|\Omega_{\neg j}] = 0 \) (Assumption 4), the next lemma provides another expression for \( \gamma_{ji} \).

**Lemma 1** The spill rate \( \gamma_{ji} \) is given by

\[
\gamma_{ji} = \frac{(q_{i|\Omega_{\neg j}} - q_i)}{q_j}.
\] (1)

Note that so far, we do not assume any particular form of the deterministic demand function \( (q_i) \). As long as we know the expected demand of products \( i \) and \( j \) when all products are available and the expected demand of product \( i \) when product \( j \) is not available, we can calculate the spill rate from product \( j \) to \( i \) through (1). So, we can empirically estimate the spill rate through experiment or from demand data if available.

Theoretically, one of the major approaches to derive demand functions is the representative consumer theory [21]. Given a particular utility function of the representative consumer, we can derive the demand functions when all products are available by maximizing the surplus of the representative consumer. Similarly, we can derive the demand functions when a particular product (say \( j \)) is not available, and then the spill rate between product \( j \) and other products can be consequently derived.

### 3 Derivation of the Spill Rate Under the Representative Consumer Theory

According to [8] and [9], if a representative consumer for an economy demands an amount \( q = (q_1, q_2, q_3, ..., q_N) \) from sellers 1, 2, ..., \( N \) at price \( p = (p_1, p_2, p_3, ..., p_N) \) respectively, the
utility function of the representative consumer is \( U(q) = \alpha^t q - \frac{1}{2} q^t C_N q \), where \( C_N = [c_{ij}] \) is an \( N \times N \) symmetric \((c_{ij} = c_{ji})\) positive definite matrix and \( \alpha \) is a vector, both defining sensitivities of the utility function related to the consumptions. Furthermore, \( C_N \) defines the self-price sensitivity and degree of product differentiation of the corresponding linear-demand system; \( C_N \) and \( \alpha \) jointly determine the market potential.

The surplus of the representative consumer is \( U(q) - p^t q \). The optimal amount of each product that the representative consumer needs to buy is the solution of \( \max_q (U(q) - p^t q) = \max_q (\alpha^t q - \frac{1}{2} q^t C_N q - p^t q) \). The first-order conditions for the maximization problem are \( C_N q = \alpha - p \) and lead directly to the linear demand function for differentiated products, \( q = C_N^{-1}(\alpha - p) \).

Now we introduce one useful—although non-traditional—notation: \( (y)_{N-1} \) will mean the \((N - 1)\)-vector consisting of the first \((N - 1)\) components and omitting the last component \( y_N \). Similarly \( C_{N-1} \) will refer to just the first \((N - 1)\) rows and columns of \( C_N \).

Suppose, without loss of generality, product \( N \) is out of the consideration set, then \( q_{\Omega \setminus N} \) for products \( i = 1...N - 1 \) would be given by the first order conditions of representative-consumer surplus associated with \( C_{N-1} \). Thus \( C_{N-1} q_{\Omega \setminus N} = (\alpha - p)_{N-1} \) and \( q_{\Omega \setminus N} = C_{N-1}^{-1}(\alpha - p)_{N-1} \). Notice that the same price vector \( p \) is used for the case when the consideration set is \( \Omega \setminus N \). As suggested by Assumption 1 above, retailers do not adjust prices within the single selling period. Then we have the following theorem.

**Theorem 1** \( \gamma_{N1} = (-1)^{2+N} \det(C_{N1}^N) / \det(C_{NN}^N) \) where \( \det(C_{ij}^N) \) is the \((i, j)\) minor (the determinant of the matrix that is formed by omitting the \(i^{th}\) row and \(j^{th}\) column) of \( C_N \).

**Proof.** Note that from linear algebra, the inverse matrix \( C_N^{-1} = \text{adj}(C_N) / \det(C_N) \), where the adjoint matrix \( \text{adj}(C_N) = [(-1)^{i+j} \det(C_{ij}^N)]^t = [(-1)^{i+j} \det(C_{ji}^N)] \), \( \det(C_{ij}^N) \) is the \((i, j)\) minor of \( C_N \), and \( \det(C_N) \) is the determinant of \( C_N \). We shall derive \((q_{\Omega \setminus N} - q_1)/q_N \) using expressions related to the \( C_N \) matrix. When product \( N \) is not available, the first-order condition becomes \( q_{\Omega \setminus N} = C_{N-1}^{-1}(\alpha - p)_{N-1} \). Then for any \( i \neq N \), we have: \( \alpha_i - p_i = \)
\[\sum_{j=1}^{N-1} c_{ij} q_j q_{[\Omega,N]} = \sum_{j=1}^{N} c_{ij} q_j = \sum_{j=1}^{N-1} c_{ij} q_j + c_{iN} q_N.\]

Recall that when all products are available, we have \(\alpha_i - p_i = \sum_{j=1}^{N} c_{ij} q_j = \sum_{j=1}^{N-1} c_{ij} q_j + c_{iN} q_N.\)

So \(\sum_{j=1}^{N-1} c_{ij} \frac{(q_j q_{[\Omega,N]-q_j})}{q_N} = c_{iN}\) or \(C_{N-1}^{-1}(c_{i1}, \ldots, c_{iN-1})^t,\) which gives

\[
\begin{bmatrix}
(q_1 q_{[\Omega,N]-q_1})/q_N \\
(q_2 q_{[\Omega,N]-q_2})/q_N \\
\vdots \\
(q_{N-1} q_{[\Omega,N]-q_{N-1}})/q_N
\end{bmatrix}^t = C_{N-1}^{-1}(c_{i1}, c_{i2}, \ldots, c_{iN-1})^t,
\]

where \(C_{N-1}^{-1} = [(-1)^{i+j} \det(C_{N-1}^{ji})]/\det(C_{N-1}).\)

In particular

\[
\frac{(q_1 q_{[\Omega,N]-q_1})}{q_N} = \sum_{i=1}^{N-1} \{c_{iN}(-1)^{i+j} \det(C_{N-1}^{ji})\}/\det(C_{N-1}).
\]

(2)

Now the \((N,1)\) minor in \(C_N\) is

\[
\det(C_N^{N1}) = \det
\begin{bmatrix}
  c_{11} & \ldots & c_{1N-1} & c_{1N} \\
  c_{21} & \ldots & c_{2N-1} & c_{2N} \\
  \vdots & \vdots & \vdots & \vdots \\
  c_{N-1,1} & \ldots & c_{N-1,N-1} & c_{N-1,N}
\end{bmatrix}.
\]

Expanding this by the last column we get:

\[
\det(C_N^{N1}) = (-1)^{1+N-1} c_{1N} \det
\begin{bmatrix}
  c_{22} & \ldots & c_{2N-1} \\
  \vdots & \vdots & \vdots \\
  c_{N-1,2} & \ldots & c_{N-1,N-1}
\end{bmatrix} + \ldots + \det
\begin{bmatrix}
  c_{12} & \ldots & c_{1N-1} \\
  \vdots & \vdots & \vdots \\
  c_{N-1,2} & \ldots & c_{N-1,N-1}
\end{bmatrix} + \ldots + (-1)^{i+N-1} c_{iN} \det
\begin{bmatrix}
  c_{i+1,2} & \ldots & c_{i+1,N-1} \\
  \vdots & \vdots & \vdots \\
  c_{N-1,2} & \ldots & c_{N-1,N-1}
\end{bmatrix} + \ldots
\]

\[
\]
.. + (-1)^{N-1+N-1}c_{N-1}
\det \begin{bmatrix}
  c_1 & 2 & \cdots & c_1 & N-1 \\
  c_2 & 2 & \cdots & c_2 & N-1 \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  c_{N-2} & 2 & \cdots & c_{N-2} & N-1
\end{bmatrix}.

Observe, for example from the second matrix above, that the \( N-2 \times N-2 \) matrix is also the submatrix if we remove the \( i \)th row and 1st column from the matrix \( C_{N-1} \).

So we can write
\[
\text{det}(C_N^{N1}) = \sum_{i=1}^{N-1} (-1)^{i+N-1}c_{iN} \text{det}(C_{N-1}^{i1})
\] (3)

By substituting (3) into (2) and (1), we have:
\[
\gamma_{N1} = (q_{1(N1)} - q_1)/q_N = \sum_{i=1}^{N-1} \{c_{iN}(-1)^1 \text{det}(C_{N-1}^{i1})\}/\text{det}(C_{N-1})
\]
\[
= (-1)^{2-N} \text{det}(C_N^{N1})/\text{det}(C_{N-1}).
\]

Finally, recall that \( \text{det}(C_{N-1}) = \text{det}(C_N^{NN}) \), then
\[
\gamma_{N1} = (-1)^{2+N} \text{det}(C_N^{N1})/\text{det}(C_N^{NN}).
\]

An interesting observation from Theorem 1 is that the spill rate is not a direct function of the retail prices charged by the newsvendors: as long as the newsvendors keep the same prices for the direct and indirect demand, the spill customers will have foreseen this and will accept the price when substituting. With the same sets of parameters (\( C_N \) and \( \alpha \)), a change in prices (\( p \)) will lead to a change in the direct demand (\( q \)) as predicted by the demand function \( q = C_{N-1}^{-1}(\alpha - p) \), and consequently might lead to a change in the amount of unsatisfied consumers for a product as well. However, the proportion of consumers of a product who would like to switch to another product after a stockout will remain unchanged.

The common way of estimating \( C_N \) and \( \alpha \) of the representative consumer is through estimating the parameters in the corresponding demand function, \( q = C_N^{-1}\alpha - C_N^{-1}p \), which have well-established economic interpretations. Extensive research has been done on estimating demand parameters in economics (e.g., [22]). Now consider a duopolistic industry with a representative consumer having \( U(q) = q_1^\alpha + q_2^\alpha - \frac{1}{2}(\beta_1 q_1^2 + 2\delta q_1 q_2 + \beta_2 q_2^2) \). Thus
\[ C_N = \begin{bmatrix} \beta_1 & \delta \\ \delta & \beta_2 \end{bmatrix}. \]

Then the inverse demand system is given by
\[ p_1 = \alpha_1 - \beta_1 q_1 - \delta q_2 \quad \text{and} \quad p_2 = \alpha_2 - \beta_2 q_2 - \delta q_1. \]

Using Theorem 1, we can easily calculate the spill rate \( \gamma_{21} = \delta/\beta_1 \) and \( \gamma_{12} = \delta/\beta_2 \).

A commonly used linear-demand form in marketing (e.g., [23]) is
\[ q_i = a_i - b_i p_i + \sum_{j \neq i} \theta_{ij} (p_j - p_i), \]
where \( a_i, b_i \) and \( \theta_{ij} \) respectively represent the market potential, the self-price sensitivity and the “degree of competition” or “index of differentiation/substitutability” between products \( i \) and \( j \). The advantage of this form is that it separates the self-price effect from the cross-price effect of products. Note that there is an one-to-one mapping between parameters in this particular linear-demand form and parameters in \( C_N q = \alpha - p \). After estimating parameters such as \( a_i, b_i \) and \( \theta_{ij} \), we can solve for \( C_N \) and \( \alpha \) through a sequence of linear equations. Now, how do we represent the spill rate in terms of the parameters in this commonly used linear-demand form?

**Proposition 1** Under the assumption that the demand function has the form of
\[ q_i = a_i - b_i p_i + \sum_{j \neq i} \theta_{ij} (p_j - p_i), \]
then \( \gamma_{N1} = \theta_{1N}/(b_N + \sum_{k \neq N} \theta_{Nk}) \).

**Proof.** Note that \( q = C_N^{-1}(\alpha - p) = [(-1)^{i+j} \det(C_{ij}^N)]/\det(C_N)(\alpha - p) \). Comparing this with \( q_i = a_i - (b_i + \sum_{j \neq i} \theta_{ij} p_i + \sum_{j \neq i} \theta_{ij} p_j) \), we should have \( \theta_{ij} = (-1)i+j+1 \det(C_{ij}^N)/\det(C_N) \) and \( (b_i + \sum_{j \neq i} \theta_{ij}) = \det(C_{ij}^N)/\det(C_N) \).

Then we get \( \theta_{1N}/(b_N + \sum_{k \neq N} \theta_{Nk}) = (-1)^{2+N} \det(C_N^{N1})/\det(C_N^{NN}) = \gamma_{N1}, \) as \( (-1)^{2+N} = (-1)^{2-N} \).

**Proposition 2** Assume \( q_i = a_i - b_i p_i + \sum_{j \neq i} \theta_{ij} (p_j - p_i) \). The spill rate \( \gamma_{Ni} \) increases with the degree of substitutability between products \( i \) and \( N \) (\( \theta_{iN} \) and \( \theta_{iN} = \theta_{Ni} \)), and decreases with the degree of substitutability between product \( N \) and any other product \( k \) (\( \theta_{Nk} \) for \( k \neq i \)), and also decreases with the self-price sensitivity of product \( N \) (\( b_N \)).
This is straightforward to derive and also intuitive. When products $i$ and $N$ become closer substitutes, then more unsatisfied consumers of product $N$ will switch to $i$. However, when product $N$ becomes a closer substitute of another product $k \neq i$, then less unsatisfied consumers of product $N$ will switch to product $i$. Also, when product $N$’s customers become more sensitive to prices (a larger $b_N$), then less unsatisfied customers of product $N$ switches to product $i$. In the symmetric case, $\gamma = \theta / (b + (N - 1)\theta)$. A larger $N$, i.e., a larger consideration set, also makes the spill rate smaller.

4 An Example Application and Future Research

There are a number of situations to which our spill result applies. Let us demonstrate with the archetypical model of $N$ newsvendors, each stocking an amount $S_i$ of a substitutable product $i$ at cost $c_i$ and offering it for sale at $p_i$. The total profit is given by $E\pi_i = -c_i S_i + p_i E[\min\{S_i, \tilde{q}_i\}] + p_i E[\min\{(S_i - \tilde{q}_i)^+, \sum_{j \neq i} \theta_{ij}(\tilde{q}_j - S_j)^+ / (b_j + \sum_{k \neq j} \theta_{jk})\}]$, which includes the cost, the revenue from the direct sales and the revenue from indirect sales. It can also be written as

$$E\pi_i = \pi_i^d - c_i s_i + p_i E[\min\{s_i, D_i^e\}]$$

(4)

where $\pi_i^d = (p_i - c_i)(a_i - b_i p_i + \sum_{j \neq i} \theta_{ij}(p_j - p_i))$ is the deterministic portion of profit, $s_i$ is the safety stock given by $s_i = S_i - q_i$, and $D_i^e = x_i + \sum_{j \neq i} \theta_{ij}(x_j - s_j)^+ / (b_j + \sum_{k \neq j} \theta_{jk})$ is the effective demand for newsvendor $i$. See [16] for the existence and uniqueness of a pure-strategy Nash equilibrium of the game.

**Proposition 3** At a symmetric equilibrium (with $c_i = c$, $a_i = a$, $b_i = b$, and $\theta_{ij} = \theta$, $s_i = s$ and $p_i = p$), the only scenario that cannot occur is $ds/d\theta < 0, dp/d\theta > 0$.

The proof can be easily obtained by taking derivatives of the first-order conditions of (4) with respect to $\theta$. A stronger degree of competition will not simultaneously lead to an increased retail price and a decreased safety stock, a situation that definitely hurts consumers.
Here the degree of product differentiation, \( \theta \), plays a dual role in describing consumer behavior: a larger \( \theta \) means not only a stronger degree of competition but also a larger spill rate.

Potential future research in this area is rich. One direction is to explore the spill rate when the expected demand function has a non-linear form; for example, a constant-price-elasticity demand model or an attraction-demand model. Another direction is to see what happens with a multiplicative-stochastic-demand system. For the application side, it would be interesting to explore how product differentiation affects a single newsvendor’s assortment decision when there is both price and inventory competition among products. The explicit modeling of spill rate relative to important economic parameters associated with the demand system will hopefully bring new insights to the joint assortment and inventory management literature.

Acknowledgements

The authors thank the Area Editor and the anonymous Associate Editor for their constructive comments and suggestions, which have significantly improved this paper.

This research is supported by the Natural Sciences and Engineering Research Council of Canada, No. 4743 and No. 312572-05.

References


