Supply Chain Choice on the Internet*

Serguei Netessine
University of Pennsylvania
Philadelphia, PA 19104, USA
netessine@wharton.upenn.edu

Nils Rudi
INSEAD
77305 Fontainebleau, France
nils.rudi@insead.edu

December 2001, revised January 2003, August 2004, July 2005 and December 2005
Forthcoming in *Management Science*

Abstract

Internet companies extensively use the practice of drop-shipping, where the wholesaler stocks and owns the inventory and ships products directly to customers at retailers' request. Under the drop-shipping arrangement, the supply chain benefits from risk pooling since the inventory for multiple retailers is stocked at the same location, the wholesaler's. Another more traditional channel alternative on the Internet is one in which retailers stock and own the inventory. These two supply chain structures, which predominate on the Internet, result in different inventory risk allocation, stocking decisions, and profits for channel members. Moreover, the two channel alternatives can be combined into a dual strategy whereby the retailer uses local inventory as a primary source and relies on drop-shipping as a backup. We model the dual strategy as a noncooperative game among the retailers and the wholesaler, analyze it, and obtain insights into the structural properties of the equilibrium solution to facilitate development of recommendations for practicing managers. Finally, we characterize situations in which each of three channels is preferable by specifying appropriate ranges of critical parameters, including demand variability, the number of retailers in the channel, wholesale prices and transportation costs.

*We would like to thank Gerard Cachon for enlightening discussions. Editors and two anonymous referees provided helpful suggestions that improved the paper. We gratefully acknowledge financial support from the Wharton E-Business Initiative.

1 Introduction and literature survey

Alliance Entertainment Corp., Ingram Entertainment and Baker & Taylor are major players in the business of distributing home entertainment products, including CDs, videos, games and books to Internet as well as to brick-and-mortar retailers. One of the services they offer is drop-shipping, or consumer direct fulfillment, the practice whereby the wholesaler stocks and owns the product and ships it directly to customers at the retailer's request. Under a drop-shipping arrangement, the retailer serves as a middleman who acquires customers and accepts orders, while the wholesaler owns and holds inventory and also fulfills orders. Drop-shipping offers two obvious benefits to the retailer: no investment into fulfillment capabilities is required, and the retailer takes no inventory risk (see the example of Spun.com [3]). Drop-shipping also benefits the wholesaler in that it allows her to charge a higher wholesale price and may provide access to a wider customer base (see Scheel [31], page 42). Finally, one would expect that the supply chain as a whole would benefit due to risk pooling and possible economies of scale in transportation in cases in which the wholesaler serves multiple retailers.

The invention of the Internet, a relatively cheap electronic medium, greatly reduced the transaction costs of drop-shipping by allowing a seamless integration of retailers and wholesalers. The Internet allows in-stock availability and prices to be communicated from the wholesaler to the retailer, and orders can be placed from the retailer to the wholesaler, all in real time. Before the Internet, such integration would have required implementing a rather expensive EDI system and hence was not economically viable in most channels. As a result of this integration of retailers and wholesaler, in a recent survey 30.6% of pure-play Internet retailers cited drop-shipping as their primary way of fulfilling orders. At the same time, 44.5% of Internet retailers relied primarily on stocking inventory internally (see [4]), while the rest either outsourced fulfillment to third parties or used the fulfillment capabilities of strategic partners.

Supply chain arrangements in which the retailers stock and own inventory have been extensively analyzed in the operations literature. But the practice of drop-shipping, although almost as popular among Internet retailers, has attracted only limited attention. The marketing literature offers some references to drop-shipping: although Scheel [31] provides extensive qualitative analysis of drop-shipping in the catalog business, he also notes that the practice is generally perceived as having somewhat limited potential. The goal of this paper is to analyze comparative advantages of inventory ownership in the traditional channel and risk pooling under drop-shipping.

1.1 Summary of results

We begin by presenting two alternative models: the traditional channel, in which each retailer stocks inventory locally, and the drop-shipping channel, in which the wholesaler stocks inventory centrally and ships directly to customers. We model both channels as two-stage supply chains with one wholesaler and multiple identical retailers. In both models the wholesaler charges the retailers a fixed wholesale price (not necessarily the same). Additionally, based on our conversations with Internet companies, we assume that different channels may have different transportation costs due to economies of scale and possibly other cost differences.

When faced with a choice between the traditional and drop-shipping channels, the retailer trades off a higher wholesale price against inventory risk, and the wholesaler trades off the lower margin against inventory risk. Moreover, transportation cost differences play a role. By combining the two strategies, the retailer can first satisfy demand with the internally stocked inventory (purchased at a lower wholesale price) and then request the wholesaler to drop-ship the rest. For example, BlueLight.com, a major Internet retailer, uses the dual strategy for CDs: it stocks certain quantities of the top 40 CDs to increase profit margins and relies on drop-shipping in case of a stock-out (see [1]) while utilizing pure drop-shipping for the remaining CD titles. Also, several major retailers have recently installed Internet kiosks inside physical stores (see [2]); if a customer does not find a product on the store shelves, she can order it in store over the Internet and the product will be shipped directly to her from a central warehouse. Such an arrangement resembles the dual strategy we analyze.

The dual strategy is somewhat similar to the combination of make-to-stock and assemble-to-order practices discussed in the operations literature (see Rudi and Zheng [29] and Cattani et al. [15]) with one major difference: under the make-to-stock/assemble-to-order practice, the same company decides both how much to stock and how much to assemble-to-order, while under the dual strategy considered in this paper, the retailers and the wholesaler manage their inventories competitively such that a game-theoretic situation arises. We model the dual strategy problem as a noncooperative game in which retailers compete with the wholesaler for demand and, moreover, they compete with each other for the wholesaler's inventory allocation. We analyze equilibria of this game, derive the structural properties of players' best-response functions and use them to show the uniqueness of the symmetric equilibrium. One of our results is that in equilibrium under the dual strategy it may be optimal for the wholesaler or for retailers to stock nothing, and thus the dual channel converts into a pure traditional or a pure drop-shipping channel. We are able to identify simple conditions that lead to such boundary equilibria. Using properties of supermodular games and derived properties of the best-response functions, we conduct sensitivity analysis of the equilibrium to the changes in the wholesale price, the drop-shipping markup and transportation cost parameters.

We further compare the traditional, drop-shipping and dual channels using numerical experiments. We find that, when the number of retailers is large, the benefits of risk pooling make either the drop-shipping or the dual channel more attractive than the traditional channel to both retailers and the wholesaler. This finding is consistent with the practical observation that drop-shipping is often used for books and electronic products, two segments in which there are many retailers, most served by a few large wholesalers. In the dual channel we find that changes in any cost-related parameters result in the reallocation of inventory between retailers and the wholesaler. For example, in this channel increasing the markup on drop-shipped products leads to higher retail inventory and lower wholesale inventory, whereas in the drop-shipping channel the same action would lead to higher wholesale inventory. Surprisingly, the impact of the transportation cost differential on profits in the dual channel is not always intuitive. In particular, as the transportation cost structure becomes more favorable for drop-shipping, the wholesaler's profit initially drops and the retailers' profit increases.

We find that the drop-shipping markup and the transportation cost differential are the main drivers of the choice of channel. We observe that both the drop-shipping and the dual channels have the potential to be Pareto-optimal choices, but the traditional channel does not. This finding indicates that most Internet retailers should strive to utilize drop-shipping at least to augment their internal inventory. However, in some industries wholesalers prefer the traditional channel and hence do not want to stock any inventory for drop-shipping, so this option is simply not available. In particular, this scenario arises when the drop-shipping markup is relatively small but transportation costs favor drop-shipping, so that the wholesaler does not make any money on shipping directly to customers even though the retailer prefers this option. In situations with a large set of problem parameters the dual channel is the Pareto-optimal choice, which is consistent with the proliferation of Internet kiosks and the increasing popularity of this arrangement. Due to the nonintuitive impact of transportation costs, the dual channel becomes the Pareto-optimal choice when the transportation costs are much higher in the drop-shipping channel than in the traditional channel. As a result, retailers carry a lot of inventory and rely on drop-shipping for only a small proportion of demand. Again, this finding is consistent with the existence of Internet kiosks in retail stores that already carry a lot of inventory. Overall, our results conform with the empirical findings of Randall et al. [27] that no channel option is uniformly preferred over the others and that channel choice depends on companies' environments.

1.2 Relation to the literature

Our paper addresses two key issues: 1) inventory ownership and stocking decision rights and 2) risk pooling. From the inventory ownership/stocking decision rights point of view, the traditional channel in our paper is similar to that in most of the supply chain literature since the retailer both owns inventory and decides on stocking quantity (see, for example, Zipkin [38]). In contrast, in the drop-shipping channel both inventory ownership and stocking decisions rest with the wholesaler. The only related arrangement we are aware of is Vendor Managed Inventory (see, e.g., Bernstein et al. [8]), in which the wholesaler is endowed with stocking decision rights but inventory is still owned by the retailer. A combination of Vendor Managed Inventory with consignment transfers both inventory ownership and stocking decision rights to the wholesaler, a scenario somewhat similar to our drop-shipping model. The important differences are that under consignment, inventory is stored at each retailer's location so there are no benefits to risk pooling. Moreover, the structure of transportation costs is different.

In the dual channel model we consider, each echelon of the supply chain stocks inventory, a setting that parallels the multi-echelon literature. Typically, though, in this stream of literature all echelons are managed centrally (whereas in our paper retailers and the wholesaler are separate entities) and, moreover, sales in each period are limited by inventory at the lowest (retail) echelon (whereas in our model the wholesaler's inventory is used as a backup). Cachon and Zipkin [14] and Lee and Whang [22] relax the former assumption by considering a supply chain with one retailer and each echelon locally managed. Bernstein and Federgruen [9], Cachon [11], and Chen et al. [16] further consider decentralized multi-echelon supply chains with multiple retailers. Lawson and Porteus [21] relax the second assumption by adding an option to expedite the product between stages but still assume that inventory is managed centrally. We are not aware of any papers that relax both of these assumptions. The dual strategy is also similar to the practice of subcontracting in case demand exceeds capacity (see Van Mieghem [36] and Plambeck and Taylor [25]). In another related paper, Cachon [12] models advance purchase discounts, a practice whereby the retailer can prebook inventory before the season at a lower per-unit price. His work is similar to ours in that both the retailer and the wholesaler are allowed to hold inventory. Cachon focuses on identifying Pareto-optimal price-only contracts and studies supply chain efficiency under such contracts. However, since there is only one retailer, the benefits of risk pooling are not reflected, and issues of inventory allocation among retailers do not arise.

Traditionally research on risk pooling focused on comparing the costs and benefits of building one central, versus many regional, warehouses for a single company (see Eppen [19]). Van Mieghem [35] and Rudi and Zheng [29] consider dual strategies under risk pooling by combining risk pooling at an extra cost per unit with the cheaper local inventory. All of these papers focus on the performance of a single firm and therefore do not compare the benefits of risk pooling for both the retailers and the wholesaler, whereas we do. Anupindi and Bassok [6] compare centralized and decentralized inventory management in a model with two retailers and one supplier while focusing on the effect of market search by customers in case of a stockout at one retailer. They compare the decentralized model with the model in which retailers form an alliance to centralize their inventories. Their work is similar to ours in that they consider the impact of risk pooling not only on downstream retailers, but also on the upstream supplier. The difference, however, is that in the case of centralized inventory management, Anupindi and Bassok [6] assume that the retailers still make stocking decisions, while in our drop-shipping model the wholesaler makes the stocking decisions, resulting in different channel dynamics. Furthermore, Anupindi and Bassok do not study the dual strategy that is the main focus of this paper.

Another relevant stream of literature considers transshipment, i.e., the practice whereby a retail location that is out of stock of certain goods can receive them from another location with excess inventory. The most relevant papers in this stream are Rudi et al. [30] and especially Dong and Rudi [18] where both the retailers' and the wholesaler's profits are considered. Seifert and Thonemann [32] and Seifert et al. [33] model single-directional transshipments from physical to Internet retailers. Anupindi et al. [7] consider a very general decentralized transshipment model in which multiple retailers not only stock inventory internally but also stock it at multiple jointly owned warehouse locations (similar to Anupindi and Bassok [6]). They focus on inventory allocation mechanisms and use cooperative game theory to define appropriate allocations and separate ownership of stock using the notion of "claims." The major difference between these papers and our model is that under transshipment the wholesaler does not take inventory risk and cannot satisfy customer demand, so the distinct features of drop-shipping do not apply.

A number of marketing papers explicitly focus on the issue of channel choice but not the issues of inventory ownership and inventory risk that are central to this paper. A notable exception is the marketing-operations interface paper by Porteus and Whang [26] where the authors use a principal-agent framework to model a single firm with multiple marketing managers responsible for promoting their products and a production manager responsible for mounting an effort to increase the available manufacturing capacity. The centralized decision-maker is endowed with the power to set up an efficient compensation scheme. Although this work includes some essential elements of our model, like demand uncertainty and inventory risk, it is clearly very different in motivation, modeling assumptions and goals pursued.

The rest of this paper is organized as follows. In Section 2 we present and analyze the traditional, drop-shipping, and dual strategy models. Section 3 contains a detailed analysis and comparison of the three channels. Section 4 discusses the impact of relaxing some of our assumptions and concludes with a summary of managerial insights and implications for practicing managers. All proofs are relegated to Appendix.

2 Supply chain models

We assume that there is a single product (or, equivalently, that there are multiple products but the problem is separable and solved for each product independently). The wholesaler either distributes the product to the retailers in the traditional channel (Model T) or ships directly to the customers at each retailer's request in the drop-shipping channel (Model D). The two structures are combined under the dual strategy (Model C) with retailers' inventory used as a primary source of stock. We use a single-period model that captures the trade-offs related to inventory risk while remaining sufficiently transparent to analysis and managerial insights. It is possible to extend our results into a multi-period model (see Netessine and Rudi [24]) using rather standard tools at the cost of additional notation that makes the insights less transparent. We ignore the possible salvage value of leftover inventory which can easily be incorporated at the cost of introducing an extra parameter, but without any additional insights. There are n identical retailers, indexed by $i = 1, ..., n$. Retailer i faces demand d_i, and all demands d_i are random variables with identical distributions and a symmetric correlation structure; we therefore drop subscript i whenever appropriate. The symmetry assumption provides a basis for comparison among the models while still preserving the essence of the problem. We define D to be the total demand for the product, i.e., $D = \sum_i d_i$, and we denote by $f_X(\cdot)$ the density function of the random variable X. We denote retailer i's profit by π_i and the wholesaler's profit by Π. Optimal stocking quantities and profits are marked with the superscript *, while best responses are marked with superscript Br and bold variables indicate vectors.

Each retailer sells the product at an exogenously given unit price r' (which includes the shipping charge, if any), and the wholesaler buys the product from the upstream manufacturer at a fixed unit cost c. At the beginning of the period, each retailer stocks q_{Ti} units of the product in the traditional model, the wholesaler centrally stocks Q_D units of the product in the drop-shipping model, and both parties stock q_{Ci} and Q_C, respectively, under the dual strategy. Retailers and the wholesaler are all independently managed. Retailers pay a wholesale price w for each unit purchased to stock. Furthermore, retailers may request that the wholesaler drop-ship products. Since under a drop-shipping arrangement the wholesaler takes on the task of doing fulfillment and bears the inventory risk, a reasonable drop-shipping contract will have a higher wholesale price than when retailers buy the product for their own stock. In practice, the difference in wholesale prices is typically 10-20% (see Scheel [31], page 42). Hence, we denote the wholesale price under drop-shipping by $w + \delta'$ where δ' is the additional *drop-shipping markup*. We assume that the wholesaler exists in a competitive environment and thus does not possess price-setting power (e.g., books or CDs can be purchased from several major distributors). This assumption will be discussed in Section 4. Our analysis also ignores fixed costs, but we provide a discussion of their potential impact on the choice of channel in Section 4.

As described in the introduction, the structure of transportation costs widely defined (e.g., including picking and packing) can vary according to channels. The total price quoted to the customer typically includes a separate shipping charge. However, the amount a customer pays for shipping may or may not exceed the actual shipping cost incurred. For example, in the traditional channel the quantity q_T is shipped in bulk from the wholesaler to the retailer and later the quantity sold is shipped to customers (typically in single units). We will denote by t_I the inbound unit transportation cost, and by t_T the outbound unit transportation cost, both incurred by the retailer in the traditional channel. In contrast, in the drop-shipping channel, the wholesaler ships directly to customers. In this case, although the actual transportation costs are incurred by the wholesaler, the retailer receives the payment from the customer and transfers a part of this payment to the wholesaler. We will let t_D be the outbound transportation cost incurred by the wholesaler. A part of this transportation cost is passed on to the retailer through the drop-shipping markup. For convenience, we define $r = r' - t_T$, the net unit-revenue in the traditional channel, and $\delta = \delta' - t_D$ as the net drop-shipping markup and $\tau = t_T - t_D$, the transportation cost difference of the two channels. The combination of parameters τ, t_I and δ allows us to account for a variety of practical situations that can arise (e.g., both or neither channel members can make/lose money on transportation charges, and they can split costs/benefits arbitrarily). To avoid trivial situations, we assume that $c < w \leq w + \delta < r + \tau$.

One of the key assumptions we make is that demand for each retailer is exogenously determined. The consequence of this assumption is that adding a retailer to the channel increases the total demand that the channel faces and does not cannibalize demand for the other retailers. Naturally, this assumption drives some of our results. A different approach might be to include the effect of cannibalization: while adding a retailer to the channel should increase total demand, demand for each specific retailer might go down. We believe, however, that our assumption is a reasonable one for many of the situations we have in mind: imagine that there are numerous supply chains of all three types (traditional, drop-shipping, and dual), and each retailer belongs to one of them. Then, if a retailer desires to switch between channels, he will bring his customer base without affecting demand for other retailers either in the channel that he leaves or in the channel that he joins. Hence, when we analyze the impact of the number of retailers on a channel's profits, the reader should not think of an addition of a retailer as a new entry into the market; it is merely a shift from one channel to another. This assumption is especially appropriate if retailers are very small compared to the total market size.

In our model T there is no explicit justification for the existence of the wholesaler, just as there is no justification for the existence of the retailer in model D, i.e., it is not clear why the retailer or the wholesaler gets a cut of the channel's profit since the only function each performs is the collection of the profit margin. In practice, there are many other reasons for the existence of the retailers and the wholesaler in a particular channel that we chose to ignore to keep our focus on inventory risk. Some reasons are access to a unique customer base, aggregation of different product lines, contractual obligations, etc. (see Randall et al. [28]).

2.1 The traditional channel

In the traditional channel, the retailers' and the wholesaler's profits are:

$$\begin{aligned} \pi_T &= rE\min(d, q_T) - (w + t_I)\, q_T, \\ \Pi_T &= (w - c)nq_T. \end{aligned}$$

Retailers make stocking decisions individually, each maximizing its expected profit, leading to the well-known newsvendor optimality condition $\Pr(d < q_T^*) = (r - w - t_I)/r$.

2.2 The drop-shipping channel

In the drop-shipping channel there is an issue of allocating inventory among retailers. Various allocation mechanisms are possible (Cachon and Lariviere [13] and Anupindi et al. [7] focus on this issue). We assume that the allocation rule is efficient (i.e., inventory is always allocated if there is demand for it) and increasing (i.e., a retailer never receives less by requesting more). Cachon and Lariviere [13] provide extensive discussion of and the motivation behind these assumptions. Let $A_i(\mathbf{q}_C, Q_C, \mathbf{d})$ be the inventory allocation to the retailer i through drop-shipping. Then the expected profits of an arbitrary retailer and the wholesaler in the drop-shipping channel can be expressed as

$$\begin{aligned} \pi_{Di} &= (r + \tau - w - \delta) E A_i(\mathbf{q}_C, Q_C, \mathbf{d}), \\ \Pi_D &= (w + \delta) E \min(D, Q_D) - c Q_D. \end{aligned}$$

In the drop-shipping channel, the inventory decision is made by the wholesaler, maximizing her expected profit and leading to the newsvendor optimality condition with respect to total demand and a set of cost/revenue parameters that is different from those of the traditional channel $\Pr(D < Q_D^*) = (w + \delta - c)/(w + \delta)$.

2.3 The dual channel: combining drop-shipping and stocking decisions

A company does not necessarily have to choose between traditional and drop-shipping channels. The two pure strategies can also be combined to reap the benefits of both channels. Under such an arrangement, the retailer would stock a certain amount of inventory internally and use drop-shipping only in cases where demand could not be completely satisfied from his own inventory. In what follows, we begin by formulating the game and introducing an appropriate inventory allocation rule. We then demonstrate the existence of equilibrium in this game and analyze the best response functions. This analysis reveals that an equilibrium solution can be interior or boundary. We then conduct a parametric sensitivity analysis of the interior equilibrium solution with respect to problem parameters. This is done both to gain insights into the distribution of inventory between the retail and the wholesale echelons and to obtain intermediate results that help to demonstrate the uniqueness of the symmetric equilibrium in this game, as shown in the last result in this section.

Each retailer independently and simultaneously sets the stocking level q_{Ci} by purchasing inventory from the wholesaler at unit price w before the uncertainty in demand is resolved. Once demand is known, each retailer requests drop-shipping at a unit price $w + \delta$ for any demand in excess of q_{Ci}. In Section 4.2