

Global Dual Sourcing: Tailored Base Surge Allocation to Near and Offshore Production

Gad Allon and Jan A. Van Mieghem

Kellogg School of Management, Northwestern University, Evanston, IL 60208

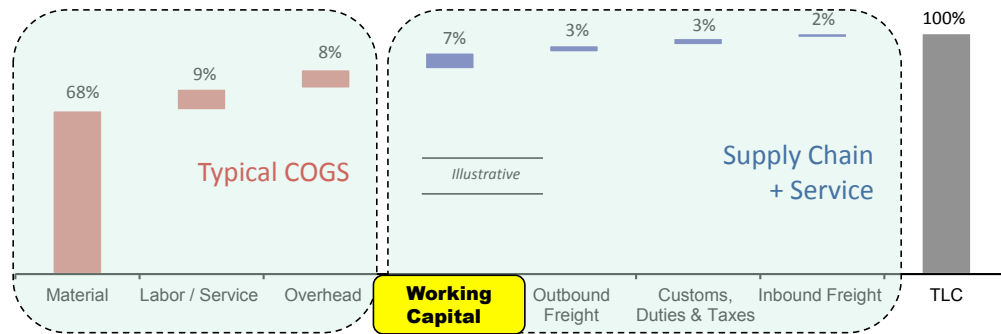
September 25, 2008; Revised Oct 8, 2008

When designing a sourcing strategy in practice, a key task is to determine the average order rates placed to each source because that affects costs and supplier management. We consider a firm that has access to a responsive near-shore source (e.g., Mexico) and a low-cost offshore source (e.g., China). The firm must determine an inventory sourcing policy to satisfy random demand over time. Unfortunately, the optimal policy is too complex to allow a direct answer to our key question. Therefore, we analyze a tailored base-surge (TBS) sourcing policy that is simple, used in practice, and captures the classic tradeoff between cost and responsiveness. The TBS policy replenishes at a constant rate from the offshore source and produces at the near shore plant only when inventory is below a target. The constant base allocation allows the offshore facility to focus on cost efficiency while the nearshore’s quick response capability is utilized only dynamically to guarantee high service. The research goals are to i) determine the allocation of random demand into base and surge capacity, ii) estimate corresponding working capital requirements, and iii) identify and value the key drivers of dual sourcing. Given that even this simple TBS policy is not amenable to exact analysis, we investigate a Brownian approximation that yields a simple “square-root” formula that is insightful to answer our questions and sufficiently accurate for practice, as is demonstrated with a validation study.

1. Introduction and Summary

A \$10 billion high-tech U.S. manufacturer of wireless transmission components was at a crossroads regarding its global network.¹ The company had two assembly plants, one in China and another in Mexico. While the Chinese facility enjoyed lower costs, ocean transportation made its order leadtimes 5 to 10 times as long as those from Mexico. With highly uncertain product demand—coefficients of variations of monthly demand for some products were as high as 1.25—sole sourcing was unattractive: Mexico was too expensive and China too unresponsive. The firm had to decide how it could best utilize these two sources by properly allocating product demand to them. In practice, specifying supply allocations is a key task of any sourcing strategy—be it global

¹ The sourcing strategy that motivated this paper is further described in Mini-Case 6 in Van Mieghem (2008).

Figure 1 The total landed cost is the cost to transform inputs at the source to outputs at destination.

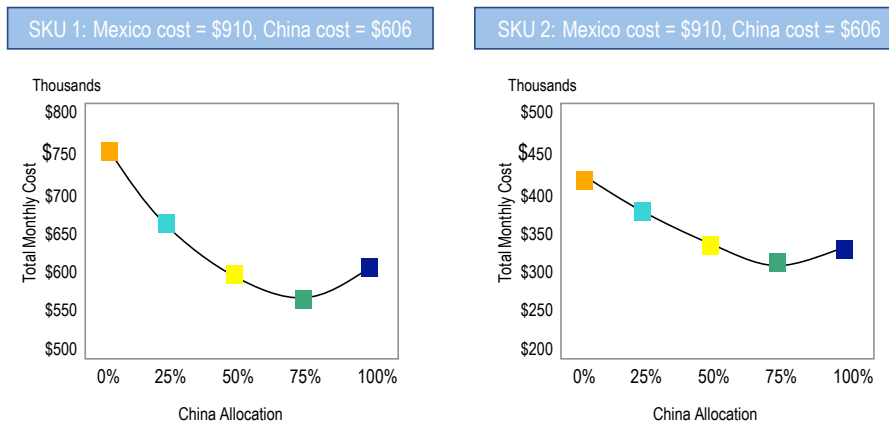
or domestic—because it affects costs and supplier management.² In this paper, we will refer to the average order rates as strategic allocation.

The manufacturer retained a management consultant company for advice. Their analysis focused on computing the total landed cost as a function of the allocation to China. The total landed cost represents the end-to-end cost to transform inputs at the source to outputs at destination (Van Mieghem 2008, p. 208). It captures not only the traditional cost of goods sold (material, labor, and overhead shown in Figure 1), but also accounts for supply chain costs such as transportation, customs, duties & taxes, as well as required working capital carrying costs. We will refer to all but working capital cost components as the “sourcing cost.” Computing the sourcing cost is tedious yet straightforward. In contrast, working capital greatly depends on leadtimes (which determine pipeline inventory) and service levels (which determine safety stock). While working capital is easily estimated for single sourcing using readily available standard inventory formulae, there are no such formulae for dual sourcing because the required inventory not only depends on the allocation to both sources but also on the replenishment policy. Therefore, as part of their analysis, the management consultants resorted to an extensive simulation study of the total landed cost.

The simulation study captured a variety of product parameters as well as five distinct China allocations (0%, 25%, 50%, 75% and 100%) using the firm’s replenishment policy, which we will refer to as a tailored base-surge (TBS) sourcing policy. This policy replenishes at a constant rate from

²The problem that this paper seeks to address is tightly linked to a practical setting with two global sources. While also relevant to domestic sourcing, the policy studied in this paper is most naturally applied and interpreted in a global setting.

Figure 2 The consultants' simulated total landed cost was minimized when allocating about 75% to China



China, yet orders from Mexico only when inventory is below a target. The presumption is that the low cost source cannot rapidly change volumes due to frictions such as long leadtimes or an inflexible level production process that is essential to achieve this cost advantage. The benefits of this policy are that it is simple to administer and it eliminates the need to explicitly account for the long lead time. In addition, the policy aligns the order dynamics with each source’s competencies: The slow source replenishes “base” demand and the fast source reacts to the remaining “surge” demand. As such, the TBS policy captures the classic tradeoff between cost and responsiveness: The constant base allocation allows China to operate under level production and thereby focus on cost efficiency, while Mexico’s quick response is utilized only dynamically to guarantee high service. The simulation (Fig. 2 shows two representative results) indicated that the total cost was convex, and, for the majority of parameter values, minimal when about 75% was sourced from China. The objective of this paper is to present an analytic model and formulae to predict the optimal allocation, understand its drivers, and tailor the sourcing strategy to the demand and supply characteristics. Somewhat remarkably, we also validate the robustness of the “three quarter” allocation rule of thumb which resulted from a rather coarse analysis.

We consider a model of a single-stage inventory system that replenishes from two supply sources using a TBS policy. The demand and supply processes can be general, correlated stationary stochastic processes. Even our simple TBS policy is not amenable to exact analysis. There are two options to proceed: (1) solve the exact problem numerically or via simulation, or (2) solve an approximate

problem analytically. Given that we seek simple formulae to determine the allocation and its key determinants, we develop a Brownian analytic model to approximate the inventory dynamics and cost under a TBS policy. We then optimize over the possible allocation and base-stock levels to arrive at the optimized TBS dual sourcing policy.

Our main results can be summarized as follows:

1. We present an analytic characterization of the optimized TBS dual sourcing policy, including its strategic allocation, expected total cost rate and base-stock level, as well as an analytic expression for the corresponding “overshoot” process. In addition, we provide a simple square-root formula to predict a near-optimal allocation.
2. The analytic characterizations, including the simple square-root formula, capture and quantify the classic trade-off between cost and responsiveness. They highlight the key drivers of the dual sourcing allocations: (i) the monetary ratio of the China cost advantage to the unit holding cost; (ii) average demand rate; (iii) the volatility of demand and China supply; and (iv) demand-supply correlations as well as serial time correlations. Our results not only confirm intuition but also provide new insight and permit easy quantification of the allocation and corresponding cost. For example, an increase in the monetary ratio (either due to a larger China cost advantage or a smaller holding cost) results, as expected, in a larger China allocation. Our formulae predict that this relationship is non-linear and follows a square-root. Similarly, an increase in demand volatility decreases the China allocation. Intuitively, this reduces the base demand while increasing the surge demand. Our formula quantifies what constitutes “base demand,” thereby providing the scientific underpinnings of the principle of strategic alignment when applied to dual sourcing. We also quantify and investigate the value of dual sourcing over single sourcing.
3. A numerical study shows that our analytic characterization and the simple square-root formula provide sufficiently accurate prescriptions relative to simulation-based optimization of the TBS policy as well as more complex policies. This study initially assumes parameter values traditionally used in the literature but then continues with applying the model to real data

from the motivating example. During the latter, we discuss how to calibrate model parameters in practice and validate the robustness of the “three quarter” allocation rule of thumb. This suggests that our results are readily applicable.

The remainder of this paper is structured as follows. The next section provides a review of the relevant literature and is followed by a discussion of the model. Section 4 specifies inventory dynamics under the TBS policy while section 5 presents the Brownian model approximation and analysis. Section 6 identifies key drivers and the value of dual sourcing. Section 7 discusses the impact of demand-supply correlations as well as serial time correlations. Section 8 reports the numerical validation study. Section 9 provides a conclusion and discussion of limitations. All proofs are relegated to the on-line Appendix.

2. Literature Review

The dual sourcing literature dates back to Barankin (1961) who studies a single period model with emergency orders. The literature distinguishes between single and dual index policies, depending on whether one or two inventory positions are tracked. Another, somewhat independent distinction is between single and dual base stock policies, depending on the number of order-up-to levels used by the policy. Our TBS policy is a single index, single base-stock policy. The dual sourcing literature can also be divided into discrete and continuous review models.

Discrete review models include Fukuda (1964) who studies a dynamic inventory model with stochastic demand in which the deterministic leadtimes of both sources differ by exactly one period. He shows that single-index, dual base stock policies are optimal under mild conditions. Whittemore and Saunders (1977) extend Fukuda’s model to allow for arbitrary (yet still deterministic) leadtimes. They show that when leadtimes differ by more than one period the optimal policy is no longer a simple function of one or two inventory positions, but depends on the entire ordering history. The model in Rosenshine and Obee (1976) assumes a regular leadtime but immediate emergency replenishment. Their standing order policy, which was evaluated numerically, assumes a constant order rate from the regular source, a feature shared by our TBS policy. Tagaras and

Vlachos (2001) allow emergency replenishment within the regular review period. Veeraraghavan and Scheller-Wolf (2006) introduce a dual index policy for capacitated dual source models that can be computed using a simple simulation-based optimization procedure. They show that such dual index policy is nearly optimal when compared to state-dependent policies found via multidimensional dynamic programming. Scheller-Wolf et al. (2006) establish computationally that a single index policy can be highly effective, and even outperform a dual index policy. Sheopuri et al. (2007) generalize the dual index policy by considering two classes of policies that have an order-up-to structure for the emergency supplier. The authors show that the “Lost Sales inventory problem” is a special case of the dual sourcing problem. They use this property to suggest near-optimal policies within this class that often improve on the already excellent performance of dual index policies.³ One of their policy classes uses a single order-up-to level throughout and then allocates, in each period, demand to each supplier. The idea of determining the allocation is similar in spirit to our approach. However, to address our research question of strategic allocation, we first determine the average allocation throughout. This average allocation then determines a single order-up-to level which specifies dynamically when to source from the fast supplier. In addition, the papers above consider deterministic leadtimes; in contrast, one of the goals of our model is to explore the relationship between the optimal strategic allocation to each source and the volatility of the supply sources.

Continuous review models include Moinzadeh and Nahmias (1988) who consider two sources with deterministic leadtimes and fixed order costs. They extend the (Q, R) policy to two different lot sizes and two different reorder levels, and optimize over these four parameters. Assuming negligible fixed order costs, Moinzadeh and Schmidt (1991) consider a more sophisticated dual base-stock policy in which real-time supply information on the age of all outstanding orders and the inventory level is used. Song and Zipkin (2008) extend Moinzadeh and Schmidt (1991) by considering a system with multiple supply sources under stochastic demand and leadtimes. The authors develop

³ Dual index policies and their generalizations require the stationary distribution of an “overshoot” process which typically is obtained through simulation. We provide an analytic expression of an overshoot distribution that may be useful in the computation of the former policies.

performance evaluation tools for a family of policies that utilize real time supply information and under which the supply system becomes a network of queues with a routing mechanism called an overflow bypass. Bradley (2004), which is the closest to our model and inspired our analysis, considers a production-inventory problem where the inventory can be replenished from in-house production or through a subcontractor. The author constructs a Brownian approximation of the optimal control problem, assuming that the manufacturer uses a single-index, dual-base stock policy. By only using a single-base stock, our replenishment policy is simpler and provides greater tractability. This allows us to specify and investigate the optimal allocations explicitly. Zipkin (2000) highlights the connection between inventory and queueing theory and argues on p. 13 that “queueing theory remains our richest source of models for supply processes.”

The aforementioned papers focus on determining or optimizing the control parameters, or on evaluating the performance, of dual sourcing policies. While we also derive the optimal base-stock level of the TBS policy, our focus is on determining the optimal allocation of demand to either source. The latter is also the focus of the literature on “order splitting,” which studies inventory models with deterministic demand. Lau and Zhao (1994)’s paper belongs to this stream and studies a system with stochastic leadtimes and explores the impact of splitting rules on inventory costs and stockout risks.

Besides dual source inventory models, our work is also related to the literature on inventory models with returns. Under a TBS policy, the net demand experienced by the firm, after subtracting the base-demand replenishment, can be negative. Inventory models with returns, as studied by Fleischmann et al. (2002) and DeCroix et al. (2005), are characterized by the same feature. Fleischmann et al. (2002) study a Markovian model with fixed cost. The behavior of the inventory cost as a function of the return ratio is closely related to the behavior of the total cost in our model as a function of the allocation to China. DeCroix et al. (2005) study a more general serial system with returns and show that an echelon base-stock policy is optimal.

Our model explores the cost-responsiveness tradeoff when allocating supply to a responsive, yet expensive, source and a low-cost but remote source within an *existing* network. It does not explore

financial hedging or configuring global networks. For such models we refer the reader to Ding et al. (2007) and Lu and Van Mieghem (2008) and references therein.

3. Model

Consider a continuous-time model of a single-stage inventory system with two supply sources. The cumulative demand up to time t is a stationary stochastic process $D(t)$; demand in excess of available inventory is backlogged. Initially D is modeled as a counting (renewal) process whose iid interarrival times have mean $1/\lambda$ and coefficient of variation v_D ; later we will generalize to allow for correlated interarrival times. Similarly, to model production variability as well as congestion and disruption, the actual supply from either source is stochastic around its mean rate. To be precise, let $S_i(t)$ denote the cumulative quantity received from source i if it were continuously supplying during $[0, t]$. To align with our motivating example, we will use $i \in \{M, C\}$ for Mexico and China as concrete placeholders for the nearshore and offshore source, respectively. Initially, we again assume that $S_i(t)$ is a renewal process whose associated iid service times have mean $1/\mu_i$ and coefficient of variation v_i ; later we will generalize to allow for correlated intersupply as well as for crosscorrelations between intersupply and interdemand times. Thus, the parameter μ_M captures two effects: (1) it is the inventory inflow rate from Mexico when Mexico is supplying and (2) it is also a measure of the responsiveness of the fast source. The supply rates are constrained by the source capacities $\bar{\mu}_i$. We assume that both sources have sufficient supply so that single and dual sourcing are viable strategies; hence, $\bar{\mu}_i > \lambda$ for $i \in \{M, C\}$.

The unit order cost from source i is c_i . As discussed in the introduction, this sourcing cost includes all components of the total landed cost with the exception of working capital (i.e., inventory) cost. Source M is responsive but expensive, while source C is cheap but slow: $c_M > c_C$ but $\bar{\mu}_M > \bar{\mu}_C$.

Let the control $T_i(t)$ denote the actual cumulative amount of time that source i is supplying during $[0, t]$ so that $S_i(T_i(t))$ is the actual supply from source i during $[0, t]$. Let $I(t)$ denote the net-inventory process, i.e., the amount of inventory on hand minus the amount on backorder at time t . We then have the following dynamics:

$$I(t) = I(0) + S_C(T_C(t)) + S_M(T_M(t)) - D(t).$$

Let $I(\infty)$ denote the steady-state net-inventory process for a given control policy T . On-hand inventory I^+ incurs the familiar per-unit holding cost h per unit of time.⁴ Stockouts are backlogged and backorders I^- incur a per-unit backlogging penalty cost b per unit of time. In the usual way, the average inventory (or demand-supply mismatch) cost rate under this policy is $G = \mathbb{E}g(I(\infty))$, where $g(x) = hx^+ + bx^- = hx + (b+h)x^-$. Let ζ denote the critical fractile $b/(b+h)$ and $\bar{\zeta} = 1 - \zeta$.

The research question is to determine the allocation policy T that minimizes total cost C , the sum of expected mismatch and sourcing costs. We seek simple characterizations of how the sourcing volume λ should be allocated to the two sources. In other words, we want to characterize the “base demand” that should be allocated to China, and when tailored dual sourcing outperforms single sourcing.

Addressing these questions involves determining the optimal dynamic order policy, which is generally complex and not amenable to exact analysis. Therefore, in what follows, we first restrict attention to a particular allocation policy (the TBS policy) for which we provide some general results. To further quantify its performance, we then provide an analytic characterization using an approximate Brownian model of the TBS policy. In a third step, we present a simple square root formula that is a lower bound of the predicted optimal allocation in the Brownian model. Finally, our numerical study validates the accuracy of our approximation.

4. The Tailored Base Surge Allocation

The simplest tailored allocation policy orders a constant rate from the offshore source and orders only occasionally when needed from the nearshore source. Specifically, let μ_C denote the constant supply rate from China; clearly, $0 \leq \mu_C < \lambda$ to prevent unlimited inventory buildup. The policy orders from Mexico only when the on-hand inventory falls below a target level s . During that time, supply from Mexico is received at rate μ_M . Obviously, to keep up with demand, $\mu_C + \mu_M > \lambda$.

⁴ Here h is the average unit holding cost rate. Under the policy analyzed in this paper, its opportunity cost component is a weighted average $r((1 - \rho_C)c_M + \rho_C c_C)$, where ρ_C denotes the fraction sourced from China and r is the cost of capital. We shall see that ρ_C^* is close to 1, so that the opportunity holding cost $\simeq r c_C$ and we assume h is constant. As shown in Appendix B.14, incorporating the dependence of h on the allocation does not impact our main result (such as the square root formula) but it does significantly complicate exposition.

As stated in the Introduction, a tailored base surge (TBS) policy is used in practice because it is simple to administer and it allows the efficient source to operate under level production. It also is amenable to analysis and, hence, simple to tailor to particular demand-supply characteristics. The underlying assumption of the TBS is that the offshore source is not capable of implementing feedback control due to various frictions such as long transportation times or inflexible production.

Flow balance dictates that the long-run average supply from Mexico then is $\lambda - \mu_C$ and the long-run average sourcing cost rate is $c_M\lambda - \mu_C\Delta c$, where $\Delta c = c_M - c_C > 0$. Observe that by design, the replenishment leadtime from the slow source does not impact the TBS policy. However, we can easily account for any pipeline inventory holding costs.⁵

Under the TBS policy, continuous supply from China implies that the control $T_C(t) = t$ so that the model simplifies to a single-source inventory model with remaining demand $D(t) - S_C(t)$, which can be negative. This is mathematically equivalent to an inventory system with returns and it is well-established that a base-stock policy is optimal. A base-stock policy brings the inventory level after ordering as close to the base stock level s as possible (Porteus 2002, p. 67). In the typical base-stock dynamics, once inventory falls below s (after a potential transient initial regime), the inventory position stays at or below s and is a demand-replacement policy. This is not the case under a TBS policy because the slow source supply may occasionally exceed the actual demand resulting in excess inventory excursions above s . A similar “overshoot” phenomenon is observed in the dual index policies of Veeraraghavan and Scheller-Wolf (2006) and the generalizations by Sheopuri et al. (2007). This overshoot is a key disadvantage that is not present in Bradley (2004)’s dual-base stock policy.⁶

We will adopt a continuous-review base stock policy which requests supply at rate μ_M from the fast source whenever the net inventory falls below s . Let $Z = I - s$ denote the “excess inventory

⁵ Let L_M and L_C denote the average transportation times from Mexico and China, respectively, so that the associated in-transit holding costs is $L_M(\lambda - \mu_C)h + L_C\mu_C h$. The allocation only affects the terms in μ_C and that effect can be captured by inflating c_M by $h\Delta L$, where $\Delta L = L_C - L_M > 0$.

⁶ A dual base-stock policy has two parameters $s < s_0$. As with a TBS policy, the fast source shuts off when the net inventory I exceeds s . In addition, the slow source also shuts off when $I > s_0$.

process,” which is the inventory above the base stock. Under a TBS policy, the excess inventory dynamics simplify to:

$$Z(t) = Z(0) + S_M(T_M(t)) + S_C(t) - D(t),$$

where

$$T_M(t) = \int_0^t 1\{Z(u) < 0\} du.$$

Essentially, Z is a random walk stemming from the conventional order-up inventory dynamics with a superimposed GI/G/1 queue capturing the occasional excess inventory excursions. Let F denote the stationary distribution of Z (we will show that such limiting distribution does exist). The benefit of analyzing the excess inventory process Z is that it, and thus also its distribution F , is independent of the actual value of the base stock s .

The average steady-state total cost rate under a TBS policy with base-stock s is

$$\begin{aligned} C(s) &= \mathbb{E}g(Z(\infty) + s) + c_M \mu_M \mathbb{P}(Z(\infty) < 0) + c_C \mu_C \\ &= G(s) + c_M \mu_M \mathbb{P}(Z(\infty) < 0) + c_C \mu_C \\ &= G(s) + c_M \lambda - \mu_C \Delta c, \end{aligned}$$

given that a stationary solution requires stability so that $\mu_M \mathbb{P}(Z(\infty) < 0) = \lambda - \mu_C$. The inventory cost $G(s) = \mathbb{E}g(Z(\infty) + s) = h\mathbb{E}(Z(\infty) + s) + (b + h)\mathbb{E}(Z(\infty) + s)^-$ and integration by parts of the last term yields

$$G(s) = hs + h \int_{-\infty}^{+\infty} x dF(x) + (b + h) \int_{-\infty}^{-s} F(x) dx.$$

PROPOSITION 1. *The inventory cost $G(s)$ is convex and the optimal base-stock s^* is found as a fractile of the steady-state excess inventory distribution: If F is continuous, then $F(-s^*) = \bar{\zeta}$.*

This type of newsvendor solution has appeared in previous analyses of inventory shortfall as discussed by Bradley and Glynn (2002). Ultimately, however, we want to understand how the total cost $C(s^*)$ depends on the sourcing rates. This requires understanding how the sourcing rates impact the stationary distribution F . Given that our system involves GI/G/1 queue dynamics, its stationary distribution cannot be solved analytically in general. Therefore, to get analytic insight, we will characterize a Brownian approximation.

5. A Brownian Model of the TBS Policy

Brownian models are tractable approximations of complex stochastic models that have been successful in providing insight into inventory and queuing systems, often with surprising accuracy. In essence, Brownian models approximate the asymptotic behavior of the underlying stochastic processes by a long-term deterministic trend (drift) component and a superimposed stochastic variability (Brownian) component. Appendix A shows how this approximation can be justified with rigorous limit theorems; in the remainder, however, we will simply use these approximations. For example, the Brownian approximation of a renewal process such as the demand process in our model has drift $\delta = \lambda$ and variance $\sigma^2 = \lambda v_D^2$. Applying this approximation to the two supply processes, assuming they are independent renewal processes, directly yields the following dual-drift Brownian model:

If $Z < 0$, both sources supply and the excess inventory dynamics of Z are those of the renewal process $S_M + S_C - D$. Its Brownian approximation Z^* thus has drift (infinitesimal mean) δ_1 and infinitesimal variance σ_1^2 , where

$$\delta_1 = \mu_M + \mu_C - \lambda > 0 \text{ and } \sigma_1^2 = \lambda v_D^2 + \mu_M v_M^2 + \mu_C v_C^2.$$

If $Z \geq 0$, only the slow source supplies and Z follows the renewal process $S_C - D$, whose Brownian approximation Z^* has negative drift $-\delta_2$ and variance σ_2^2 , where

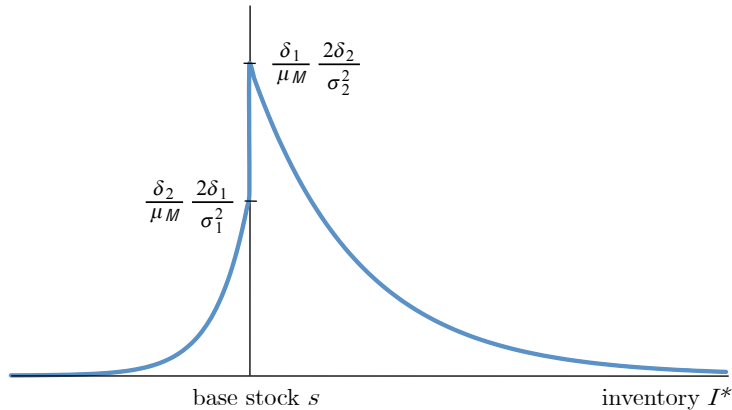
$$-\delta_2 = \mu_C - \lambda < 0 \text{ and } \sigma_2^2 = \lambda v_D^2 + \mu_C v_C^2.$$

Our numerical study below will show that this Brownian approximation is sufficiently accurate for strategic allocation. In addition, while we have assumed independent renewal processes, the Brownian approximation remains valid for any stationary sequence of production times and interdemand times, provided the variance is adjusted accordingly, as we will show in section 7.

5.1. Steady state distribution

The Brownian approximation Z^* of the inventory process is a piecewise constant diffusion process whose steady-state limit exists and follows directly from Browne and Whitt (1995).

Figure 3 The stationary inventory has a bi-exponential distribution centered around the base stock level s .



PROPOSITION 2. The steady-state limit $Z^*(\infty)$ has probability density and distribution function

$$f(x) = \begin{cases} \frac{\delta_2}{\mu_M} \frac{2\delta_1}{\sigma_1^2} \exp\left(\frac{2\delta_1}{\sigma_1^2}x\right), & x < 0, \\ \frac{\delta_1}{\mu_M} \frac{2\delta_2}{\sigma_2^2} \exp\left(-\frac{2\delta_2}{\sigma_2^2}x\right), & x \geq 0, \end{cases} \quad F(x) = \begin{cases} \frac{\delta_2}{\mu_M} \exp\left(\frac{2\delta_1}{\sigma_1^2}x\right), & x < 0, \\ 1 - \frac{\delta_1}{\mu_M} \exp\left(-\frac{2\delta_2}{\sigma_2^2}x\right) & x \geq 0. \end{cases} \quad (1)$$

Notice that this proposition provides an analytic characterization of the overshoot process that may be useful in the computation of dual index policies and their generalizations.

PROPOSITION 3. For a fixed base stock s , the distribution function F increases in μ_M and, if $\lambda(v_D^2 + v_C^2) \geq \mu_M(v_C^2 - v_M^2)$, also in μ_C .

Typically, higher supply rates stochastically increase inventory. However, if the offshore variability is high ($v_C \gg v_M$), increasing the offshore allocation may reduce inventory availability.

Notice that the density can be written as $f(x) = p_1 f_1(x)$ for $x < 0$ and $p_2 f_2(x)$ for $x \geq 0$ where f_i are exponential density functions (integrating to 1) and $p_1 + p_2 = 1$, as shown in Fig. 3. Conceptually, this follows from the properties of truncated reversible Markov processes. The densities f_i are the scaled versions of the densities that are obtained from a reflected diffusion with a single constant drift and variance over the unrestricted state-space. This means that the terms $\frac{\sigma_i^2}{2\delta_i}$ denote conditional average excess inventory:

$$\frac{-\sigma_1^2}{2\delta_1} = \mathbb{E}[Z^*(\infty)|Z^*(\infty) < 0] \quad \text{and} \quad \frac{\sigma_2^2}{2\delta_2} = \mathbb{E}[Z^*(\infty)|Z^*(\infty) > 0],$$

and the Brownian estimate $\mathbb{E}I^*$ of the average steady-state inventory is

$$\mathbb{E}I^* = s + \mathbb{E}Z^*(\infty) = s - \frac{\delta_2}{\mu_M} \frac{\sigma_1^2}{2\delta_1} + \frac{\delta_1}{\mu_M} \frac{\sigma_2^2}{2\delta_2}.$$

The steady-state distribution can be calculated exactly if demand and supply are independent Poisson processes. Appendix B.4 shows that the exact distribution is bi-geometric, the discrete counterpart of the continuous-state bi-exponential distribution. In addition:

PROPOSITION 4. *Consider Poisson demand and independent exponentially distributed intersupply times. Then the Brownian estimate $\mathbb{E}I^*$ matches the exact average steady-state net-inventory up to the expected discretization error: $\mathbb{E}[I - I^*] = -\frac{1}{2}$.*

Given that a continuous approximation on an integer-valued state space includes an expected discretization error, this means that the Brownian estimate is as good as one can hope for. The important point to note is that this error is independent of the rate parameter values within their valid ranges $0 \leq \mu_C < \lambda < \mu_M + \mu_C$. In other words, in this setting, the Brownian estimate of the expected excess inventory is for all practical purposes exact (even when the so-called heavy-traffic conditions $\lambda \simeq \mu_M + \mu_C$ and $\lambda \simeq \mu_C$ that are derived in Appendix A are not satisfied).

5.2. Optimal base-stock s^* and associated inventory cost $G(s^*)$

The explicit characterization (1) of the distribution F allows the specification of the optimal base-stock $s^* = -F^{-1}(\bar{\zeta})$ and associated inventory cost for fixed allocations μ . Given the specific bi-exponential structure of F , we distinguish two operating regimes: $s^* \geq 0$ versus $s^* < 0$. Following Bradley (2004), we say that the control policy is preventive if the fast source is engaged while inventory is on hand ($s^* \geq 0$) and reactive when the fast source only supplies backorders ($s^* < 0$).

PROPOSITION 5. *Consider fixed supply rates μ_M and μ_C . If $\bar{\zeta} \leq \frac{\lambda - \mu_C}{\mu_M}$, then the optimal base-stock is positive so that the fast source supplies to stock (“preventive mode”) with*

$$s^* = -\frac{\sigma_1^2}{2\delta_1} \ln \frac{\mu_M}{\delta_2} \bar{\zeta} \geq 0 \quad \text{and} \quad G(s^*) = hs^* + h\frac{\sigma_1^2}{2\mu_M} + h\frac{\delta_1\sigma_2^2}{2\delta_2\mu_M}.$$

Otherwise the optimal base-stock is negative and the fast source only engages to cover backlog (“reactive mode”) with

$$s^* = \frac{\sigma_2^2}{2\delta_2} \ln \frac{\mu_M}{\delta_1} \zeta < 0 \quad \text{and} \quad G(s^*) = -bs^* + b\frac{\delta_2\sigma_1^2}{2\delta_1\mu_M} + b\frac{\sigma_2^2}{2\mu_M}.$$

The optimal base-stock s^* is decreasing in μ_M and μ_C . In addition, for a constant allocation, the absolute value of the optimal base-stock and the total cost are increasing in volatility v_D, v_C and v_M . Operating in preventive (reactive) mode is similar to operating the Mexico source in make-to-stock (make-to-order) fashion. The optimal regime is preventive when relative holding costs h/b and the contingent supply μ_M are small, otherwise it is better to move to a make-to-order model in which we operate in reactive mode and use the ample capacity of the fast source to cover backlogs.

5.3. Optimal contingent “surge” rate μ_M^*

As expected, a higher contingent surge rate μ_M is always desirable:

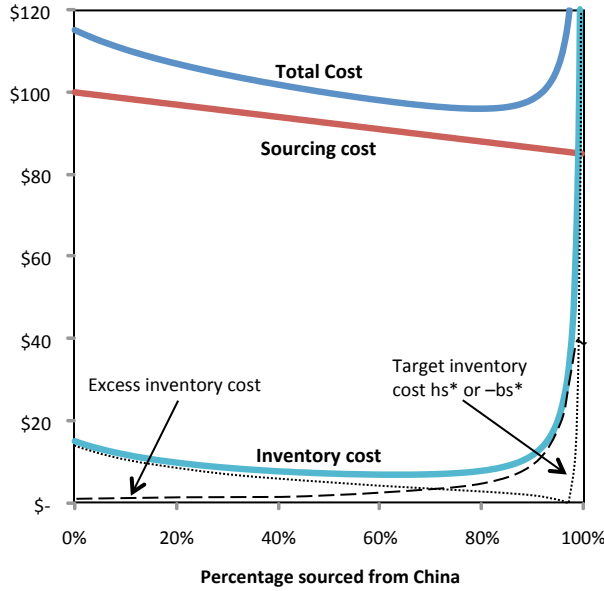
PROPOSITION 6. *The inventory cost $G(s^*|\mu_M, \mu_C)$, and thus the total cost, is strictly decreasing in μ_M so that the optimal contingent surge rate $\mu_M^* = \bar{\mu}_M$.*

Higher supply rates allow us to reduce the base-stock and thus reduce holding costs. The effect on total cost, however, might seem counterintuitive: given that sourcing from Mexico is more expensive, why would a higher contingent supply rate μ_M always be better? Notice that a higher contingent supply rate μ_M does not imply higher allocation, which depends also on the probability $\mathbb{P}(Z(\infty) < 0)$ of a shortfall. All else being equal, increasing μ_M must reduce $\mathbb{P}(Z(\infty) < 0)$ while keeping the product constant to retain flow balance. Conceptually, a higher contingent supply rate has only positive impact because it improves responsiveness and reduces backlogs, while average sourcing costs remain unaffected.

Having characterized the optimal base-stock and corresponding inventory cost for the TBS policy as well as the optimal contingent supply rate, we can now solve for the optimal China allocation.

5.4. Optimal “base” rate μ_C^*

To determine the optimal China allocation, we first establish that the inventory cost $G(s^*|\mu_M, \mu_C)$, and thus the total cost C , is convex in μ_C . Prior to stating the formal result, consider Fig. 4 for a representative case. The figure shows that the two components of the inventory cost—the base stock cost hs^* or bs^* , and the excess cost—are not convex by themselves, but their sum is. This typically holds and the following proposition gives some simple sufficient conditions but the Appendix shows that the necessary conditions are much less stringent yet more complex.

Figure 4 The total cost and its components as a function of the China allocation. (Parameters: $h = 1$, $b = 50$, $c_M = 100$, $c_C = 85$, $\lambda = 1$, $\mu_M = 1.3$, all coefficients of variation = 1.)

PROPOSITION 7. Let $\max(1 + \frac{v_C^2 + v_D^2 + v_M^2}{2v_D^2 + 2v_C^2 - v_M^2}, 1 + \frac{v_D^2}{v_D^2 + v_C^2})\lambda \leq \bar{\mu}_M$. If $v_M < v_C$ let also $\bar{\mu}_M \leq \lambda \frac{(v_D^2 + v_C^2)}{v_C^2 - v_M^2}$ and $\frac{v_C^2 - 2v_M^2}{2} \leq v_D^2$. Then the inventory cost $G(s^* | \mu_M, \mu_C)$ is convex in μ_C .

(The last condition on v_D guarantees that $\max(\frac{3(v_D^2 + v_D^2)}{2v_D^2 + 2v_C^2 - v_M^2}, 1 + \frac{v_D^2}{v_D^2 + v_C^2}) < \frac{(v_D^2 + v_C^2)}{v_C^2 - v_M^2}$, so that there is a viable interval for $\bar{\mu}_M$.) Basically, the total cost is guaranteed to be convex if $\bar{\mu}_M$ exceeds λ and, if $v_M < v_C$, the demand volatility v_D^2 exceeds $.5v_C^2 - v_M^2$. The optimal China allocation μ_C^* must satisfy the first order conditions $\frac{\partial}{\partial \mu_C} G(s^*) = \Delta c > 0$, which can be simplified:

PROPOSITION 8. Under the assumption of Proposition 7, the optimal base rate $\mu_C^* = \max(0, \hat{\mu}_C)$ where $\hat{\mu}_C < \lambda$ is the unique solution to

$$2 \frac{\Delta c}{h} = \frac{\lambda(v_D^2 + v_C^2) + \mu_M(v_M^2 - v_C^2)}{(\mu_M + \mu_C - \lambda)^2} \ln \frac{\bar{\zeta}\mu_M}{\lambda - \mu_C} + \frac{\lambda(v_C^2 + v_D^2)}{(\lambda - \mu_C)^2} - \frac{(\lambda v_D^2 + \mu_M v_M^2 + \mu_C v_C^2)}{(\lambda - \mu_C)(\mu_M + \mu_C - \lambda)}, \quad (2)$$

if $s^* > 0$, and otherwise:

$$2 \frac{\Delta c}{b} = -\frac{\lambda(v_D^2 + v_C^2)}{(\lambda - \mu_C)^2} \ln \frac{\zeta\mu_M}{\mu_M + \mu_C - \lambda} + \frac{\lambda v_D^2 + \mu_C v_C^2}{(\lambda - \mu_C)(\mu_M + \mu_C - \lambda)} - \frac{\lambda(v_C^2 + v_D^2) + \mu_M(v_M^2 - v_C^2)}{(\mu_M + \mu_C - \lambda)^2}. \quad (3)$$

These transcendental equations are easily solved numerically and will be used later to derive the comparative statics. We seek, however, simple analytic expressions. Assuming that $\rho_C = \mu_C/\lambda$ is

close to 1 (which will be validated in our numerical study) and $\mu_M = \bar{\mu}_M > \lambda$, we can consider only the dominant term in the optimality equations. Doing so yields a simple square-root formula that is a lower bound and a good estimate:

PROPOSITION 9. *Suppose $v_C^2 + v_D^2 \ll 2\lambda h^{-1}\Delta c$, then an estimate of the optimal base fraction $\rho_C^* = \mu_C^*/\lambda$ is*

$$\rho_C^* \simeq \rho_C^{est} = 1 - \sqrt{\frac{v_C^2 + v_D^2}{2\lambda h^{-1}\Delta c}}.$$

In preventive mode, if $\lambda(v_D^2 + v_C^2) + \mu_M(v_M^2 - v_C^2) \geq 0$, then the square-root estimate is a lower bound: $\rho_C^{est} \leq \rho_C^$.*

The condition $v_C^2 + v_D^2 \ll 2\lambda h^{-1}\Delta c$ guarantees that the Brownian approximation is accurate. However, our numerical validation study will show that the square root formula yields a valid estimate even when ρ_C^{est} is as low as 60%. Also note that in reactive mode, as well as in preventive mode when $\lambda(v_D^2 + v_C^2) + \mu_M(v_M^2 - v_C^2) < 0$, the neglected second-order terms in the first-order conditions have opposite signs. This suggests that the accuracy of the square root formula is higher in those cases, mitigating the fact that it cannot be bounded then.

6. Drivers and Value of Dual Sourcing

6.1. Key drivers of strategic allocation

The square-root formula of Proposition 9 directly provides the following key drivers, insights and quantification on strategic allocation. First, the key monetary trade-off in determining the China allocation is $\Delta c/h$, which can be expressed as follows. Recall that the unit holding cost $h = (\text{cost of capital } r + \text{physical holding cost } p)c_C$, so that the key trade-off simplifies to

$$\begin{aligned} \frac{\Delta c}{h} &= \frac{\Delta c_{source} - h\Delta L}{h} = \frac{\Delta c/c_C}{r+p} - \Delta L \\ &= \frac{\text{relative cost advantage}}{\text{cost of capital} + \text{physical holding cost}} - \text{transportation time difference.} \end{aligned}$$

Note that this equation is in time units and that it captures the combined impact of monetary cost concerns as well as responsiveness. This is exactly the tradeoff at the essence of this model. It shows that the China allocation is high when (i) China has a high relative cost advantage (as

expected); (ii) the cost of capital and the physical holding cost are low (meaning small opportunity costs as well as low risk of obsolescence); and (iii) relatively small transportation time difference between China and Mexico. Not only does this confirm intuition, the equation also quantifies the factors and their interaction.

Second, the allocation depends on product volume and thus on its stage in the product life cycle: As the volume grows, the China allocation should increase according to the square root formula. Later, during the decline phase, that allocation should decrease thereby reflecting a shift in the relative importance from cost to responsiveness.

Third, the allocation depends mostly on the China supply volatility and the demand volatility. Our approximation depends equally on both but is rather insensitive to the Mexico supply volatility. As expected, as China becomes a less reliable source, its allocation is reduced. Interestingly, as the demand volatility increases, the allocation to China is reduced as well. Both effects reflect the fact that China is the less flexible source.

The combined impact of these three key drivers is summarized through the ratio

$$\frac{h \frac{v_C^2 + v_D^2}{2}}{\lambda \Delta c} = \frac{\text{holding cost of safety stock}}{\text{sourcing cost savings}},$$

which captures the natural trade-off in dual sourcing and quantifies it: as the ratio increases, the China allocation reduces.

6.2. Second-order drivers of strategic allocation

Factors such as Mexico's supply rate and volatility are not present in the square-root formula and thus can only have a modest impact on the strategic allocation. Nevertheless, to understand their impact, we can compute their comparative statics from the first-order conditions in Proposition 8.

Differentiating equation (2) with respect to v_M^2 yields:

$$\frac{d\mu_C}{d(v_M^2)} = -\frac{\frac{\partial^2 G}{\partial(v_M^2)\partial\mu_C}}{\frac{\partial^2 G}{\partial\mu_C^2}} = \frac{\frac{\mu_M}{\delta_2\delta_1} - \frac{\mu_M}{\delta_1^2} \ln\left(\frac{\mu_M}{\delta_2}\zeta\right)}{\frac{\partial^2 G}{\partial\mu_C^2}}.$$

The terms can be signed using propositions 7 (convexity of G) and 5 (sign of \ln term). It follows that the optimal allocation to China increases with the Mexican supply volatility. This relationship

exhibits operational hedging: the dependence on a less dependable Mexico source is mitigated by an increased allocation to the China source.

Similarly, the impact of an increase in Mexico's contingent supply rate $d\mu_C^*/d\mu_M$ is negative, as shown in the Appendix. The explanation here is a little more subtle and goes as follows. We know that, ceteris paribus, a higher contingent supply rate μ_M reduces the shortfall frequency $\mathbb{P}(Z(\infty) < 0) = (\lambda - \mu_C)/\mu_M$, which allows a reduction in the base stock and reduces inventory costs (Prop. 6). When we simultaneously adjust the China supply, however, the effect on total cost depends on the rate at which the shortfall frequency decreases in μ_M . If it decreases slowly so that the average flow from Mexico $\mu_M \mathbb{P}(Z(\infty) < 0)$ increases in μ_M , then an increase in μ_M must be compensated by a decrease in μ_C to retain flow balance. In contrast, if the shortfall frequency decreases faster such that $\mu_M \mathbb{P}(Z(\infty) < 0)$ decreases in μ_M , then we must compensate by increasing μ_C . Our analysis shows that the former effect is true. The following corollary is even more interesting:

COROLLARY 1. *As the contingent Mexican capacity $\bar{\mu}_M$ increases, the strategic China allocation is monotone decreasing with asymptote:*

$$\lim_{\bar{\mu}_M \rightarrow \infty} \mu_C^* = \lambda \left(1 - \sqrt{\frac{v_D^2 + v_C^2}{\lambda 2b^{-1} \Delta c} \ln \zeta^{-1}} \right)^+.$$

With unlimited Mexican available capacity, the firm will operate in reactive mode and the asymptote follows directly from taking the limit $\mu_M \rightarrow \infty$ in the first order condition (3). (This square-root asymptote is actually a more refined allocation estimate of the one in Proposition 9 where $\ln(1+h/b)$ was approximated by h/b .) The important implication is that the China allocation has a strictly positive floor under moderate volatility and cost advantage. In other words, even with unlimited Mexican available capacity, dual sourcing remains optimal in our model. We now turn our attention to assessing this incremental value of dual sourcing over single sourcing from Mexico.

6.3. Value of dual sourcing

The value of dual sourcing is the reduction in total cost relative to single sourcing. The evaluation of traditional single sourcing using a base-stock policy is a special case of our results above and can be found by setting $\mu_C = 0$. For example, under single sourcing from Mexico, the Brownian estimate of the stationary inventory shortfall Z^* is negative and exponentially distributed:

$$F_M(x) = \exp\left(\frac{2(\mu_M - \lambda)}{\sigma_M^2}x\right) 1_{\{x < 0\}},$$

where $\sigma_M^2 = \lambda v_D^2 + \mu_M v_M^2$. It directly follows that the corresponding average inventory is

$$\mathbb{E}I_M = s + \mathbb{E}Z_M^*(\infty) = s - \frac{\sigma_M^2}{2(\mu_M - \lambda)}.$$

Let h_M denote the unit holding cost rate when single sourcing from Mexico. The optimal base-stock s_M^* then is:

$$s_M^* = -\frac{\sigma_M^2}{2(\mu_M - \lambda)} \ln \frac{\mu_M}{\lambda} \frac{h_M}{b + h_M}.$$

As before, single sourcing operates in preventive mode if $s_M^* \geq 0$, and in reactive mode otherwise.

The Appendix shows that the associated optimal inventory cost $G_M(s_M^*)$ equals

$$G_M(s_M^*) = \begin{cases} h s_M^* & \text{if } \frac{h_M}{b+h_M} \leq \frac{\lambda}{\mu_M} \text{ (preventive mode),} \\ -b s_M^* + b \frac{\sigma_M^2}{2(\mu_M - \lambda)} & \text{if } \frac{h_M}{b+h_M} > \frac{\lambda}{\mu_M} \text{ (reactive mode).} \end{cases}$$

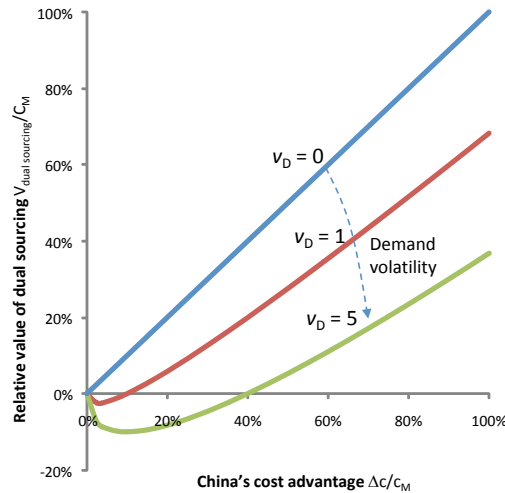
The associated incremental value of dual sourcing relative to single sourcing from Mexico is

$$\begin{aligned} V_{\text{dual}} &= C_{\text{single}} - C_{\text{dual}} = [G(s_M^*) + c_M \lambda] - [G(s^*) + c_M \lambda - \mu_C^* \Delta c] \\ &= G(s_M^*) - G(s^*) + \mu_C^* \Delta c. \end{aligned}$$

Dual sourcing always gains in sourcing cost savings ($\mu_C^* \Delta c > 0$), yet the inventory cost could increase. Our results can quantify this value of dual sourcing. To study its drivers, we focus on a relevant special case where the dominant volatility stems from demand and Mexican capacity is unlimited. In that case, we know that both the single sourcing from Mexico and dual sourcing policies operate in reactive mode so that the value of dual sourcing becomes:

$$V_{\text{dual}} = -b s_M^* + b \frac{\sigma_M^2}{2(\mu_M - \lambda)} - \left(-b s^* + b \frac{\delta_2 \sigma_1^2}{2 \delta_1 \mu_M} + b \frac{\sigma_2^2}{2 \mu_M} \right) + \mu_C^* \Delta c.$$

Figure 5 The relative value of dual sourcing as a function of the China cost advantage.



Using $v_C = v_M = 0$, we get that $\sigma_M^2 = \sigma_1^2 = \sigma_2^2 = \lambda v_D^2$ and letting $\mu_M \rightarrow \infty$ yields:

$$\begin{aligned}
 V_{\text{dual}} &= \frac{b\sigma_2^2}{2} \frac{1}{(\lambda - \mu_C^*)} \ln \zeta + \mu_C^* \Delta c \\
 &= \frac{b\lambda v_D^2}{2} \frac{1}{\sqrt{\frac{v_D^2}{2b^{-1}\Delta c} \lambda \ln \zeta^{-1}}} \ln \zeta + \lambda \left(1 - \sqrt{\frac{v_D^2}{\lambda 2b^{-1}\Delta c} \ln \zeta^{-1}} \right) \Delta c \\
 &= \lambda \Delta c - \sqrt{2\lambda b \Delta c \ln \zeta^{-1}} v_D \simeq \lambda \Delta c - \sqrt{2\lambda h \Delta c} v_D
 \end{aligned}$$

where the last step used the fact that h/b is small so that $\ln(1 + h/b) \simeq h/b$. Note that, when $\sigma_M^2 = \lambda v_D^2$ and $\mu_M \rightarrow \infty$, the total cost under single sourcing from Mexico simplifies to only the sourcing cost $c_M \lambda$, so that the relative value of dual sourcing in this case is:

$$\frac{V_{\text{dual}}}{C_{\text{single}}} \simeq \frac{\Delta c}{c_M} - \sqrt{2 \frac{h}{\lambda c_M} \frac{\Delta c}{c_M}} v_D$$

With deterministic supply and ample Mexican capacity, the relative value of dual sourcing is bounded by the relative sourcing cost savings, and is reduced by increased working capital requirements. The latter increase as demand volatility and holding costs, as well as the relative China advantage, increase. If the demand volatility or the unit holding cost is too high, dual sourcing becomes suboptimal, as shown in Figure 5.

The TBS policy assumes that feedback control on China is not feasible; which precludes the comparison of dual sourcing with single sourcing from China. If one allows feedback control, this comparison follows the same lines as our comparison with single sourcing from Mexico.

7. Serial and Cross Correlated Demand and Supply

So far we have confined the analysis to settings in which the demand and supply processes were tractable, independent renewal processes. The strength of the Brownian approximation, however, is not only analytic tractability but also generality: it can handle complex correlated processes, provided the variance terms are adjusted appropriately. Extending Bradley and Glynn (2002), Proposition 10 (relegated to the Appendix) shows the asymptotic time-average variance of a general stationary demand & supply process $\{(D(t), S_C(t), S_M(t)) : t \geq 0\}$ with interarrival times $\{(U_i, V_i^C, V_i^M) : i \in \mathbb{N}\}$. We use this proposition to discuss three applications that highlight specific correlation structures observed in practice.

First, assume that demand has autocorrelation function $\text{corr}(U_1, U_{k+1}) = \theta^k$ with $|\theta| < 1$ while the supply processes are independent renewal processes. Then $\text{covar}(U_1, U_{k+1}) = \theta^k \text{var}U_1$ and

$$\begin{aligned}\sigma_2^2 &= \lambda^3 \text{var}U_1 \left(1 + 2 \sum_{k=1}^{\infty} \theta^k \right) + \mu_C^3 \text{var}V_1^C \\ &= \lambda v_D^2 \left(\frac{1+\theta}{1-\theta} \right) + \mu_C v_C^2, \\ \sigma_1^2 &= \lambda v_D^2 \left(\frac{1+\theta}{1-\theta} \right) + \mu_C v_C^2 + \mu_M v_M^2.\end{aligned}$$

Hence, relative to our earlier setting of independent renewal processes, demand that is serially correlated over time has the effect of adjusting v_D^2 . Positive time correlations increase volatility and thus reduce the China allocation and the value of dual sourcing. In contrast, negative time correlations are mean reversing and increase the China allocation and value of dual sourcing.

Second, assume the demand and China supply are correlated renewal processes with correlation coefficient ϕ while the Mexico supply process is an independent renewal process. Then

$$\begin{aligned}\sigma_2^2 &= \lambda^3 \text{var}U_1 + \mu_C^3 \text{var}V_1^C - 2\lambda\mu_C \min(\lambda, \mu_C) \text{covar}(U_1, V_1^C) \\ &= \lambda v_D^2 + \mu_C v_C^2 \left(1 - 2\phi \frac{v_D}{v_C} \right) \\ \sigma_1^2 &= \lambda v_D^2 + \mu_C v_C^2 \left(1 - 2\phi \frac{v_D}{v_C} \right) + \mu_M v_M^2\end{aligned}$$

Hence, relative to independent renewal processes, cross correlated demand and China supply has the effect of adjusting the China supply volatility v_C^2 . Positive cross correlations $\phi > 0$ could represent

a situation where economic cycles impact both demand and China supply productivity. This would decrease China volatility and thus increase the China allocation and the value of dual sourcing. In contrast, negative cross correlations may arise due to congested transportation and import/customs processes, which would decrease the value of dual sourcing. To our knowledge, there is no empirical evidence as to which effect dominates.

Third, assume that both supplies are correlated renewal processes with correlation coefficient ϕ while demand is an independent renewal process. Given that $\bar{\mu}_M > \mu_C$, Proposition 10 yields:

$$\begin{aligned}\sigma_2^2 &= \lambda v_D^2 + \mu_C v_C^2 \\ \sigma_1^2 &= \lambda v_D^2 + \mu_C v_C^2 + \mu_M^3 \text{var} V_1^M + 2\mu_C^2 \mu_M \text{covar}(V_1^C, V_1^M) \\ &= \lambda v_D^2 + \mu_C v_C^2 \left(1 + 2\phi \frac{v_M}{v_C}\right) + \mu_M v_M^2\end{aligned}$$

Thus, relative to independent renewal processes, correlated supply has the effect of adjusting the China supply volatility v_C^2 but only when both sources are active. Given that only σ_1^2 is affected, its impact is more subtle because we cannot just adjust one volatility parameter as in the two earlier applications. Rather, we must recompute the first order conditions for the optimal China allocation. In the on-line Appendix, we show that the square root formula is not impacted so that the effect of correlated supply processes is of second order in our model. It is interesting that supply correlation—which has been advocated as an important reason to diversify the supply base—has little impact on sourcing allocation, and hence on the value of dual sourcing in our model.

8. Numerical Validation Study

We conduct a numerical study to illustrate and validate some of the key results discussed above. The goal of this validation study is to answer three questions: (i) How well does the TBS policy perform relative to the dual-base stock policy in minimizing total cost? We use the dual-base stock policy as a proxy of the optimal policy given that Bradley (2005) shows that a dual-base stock policy is optimal when all interarrival and intersupply times are exponentially distributed. (ii) How well does the Brownian approximation perform relative to simulation-based optimization

in predicting optimal allocation? (iii) How well does the square root formula (Prop. 9) perform relative to the exact analysis of the Brownian approximation (using Prop. 8), both in terms of cost minimization and allocation prediction?

We shall address these three questions by comparing four parameter cases:

1. The “base case” uses parameters similar to those studied in Bradley (2004) and Moinzadeh and Nahmias (1988): inter-demand times are independent and identically normally distributed with mean of $\lambda = 1$ and coefficient of variation $v_D = 1$; inter-supply times from Mexico are independent and identically normally distributed with $\mu_M = \bar{\mu}_M = 1.5$ and $v_M = 1$; and inter-supply times from China are independent and identically normally distributed with μ_C (the decision variable) and $v_C = 0.5$. (When simulating, negative sampled interarrival times were truncated to 0.⁷) The holding and backlogging costs are $h = \$1$ per period of time per unit and $b = \$50$ per period of time per unit.
2. The “More volatile China” case uses the same parameters as in the base case with only one modification: the volatility of the inter-supply times form China is raised to $v_C = 1$.
3. The “More expensive holding” case also differs from the base case in one parameter: the holding cost is increased to $h = \$2.5$ per unit per period of time.
4. The “Volatile China, Expensive holding” case combines cases 2 and 3: it uses the base-case parameters with two modifications: $v_C = 1$ and $h = \$2.5$ per unit per period of time.

Comparing the TBS with dual-base stock policies Our validation study starts with investigating how well the optimized TBS policy performs relative to the more complex dual base stock policy. For various values of China’s relative cost advantage $\Delta c/c_M$ and China’s allocation μ_C , we simulated total cost under a TBS policy with various base-stock levels. A numerical search then found the cost-minimizing allocation and base-stock level and the corresponding optimal cost under TBS. For the dual base stock policy, we first used the TBS-cost-minimizing allocation as the contingent supply rate from China. Then, we simulated the total cost for a grid of possible base

⁷ We computed the coefficient of variation of the truncated sample and found it was very close to that of the non-truncated distribution.

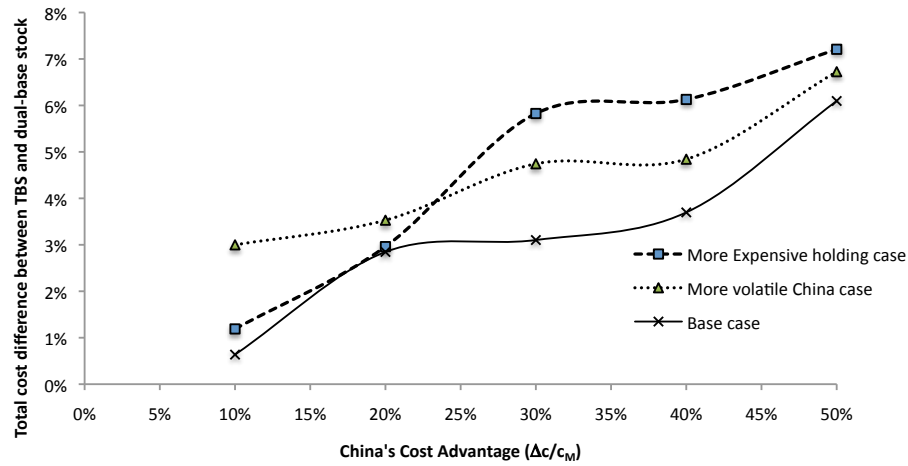
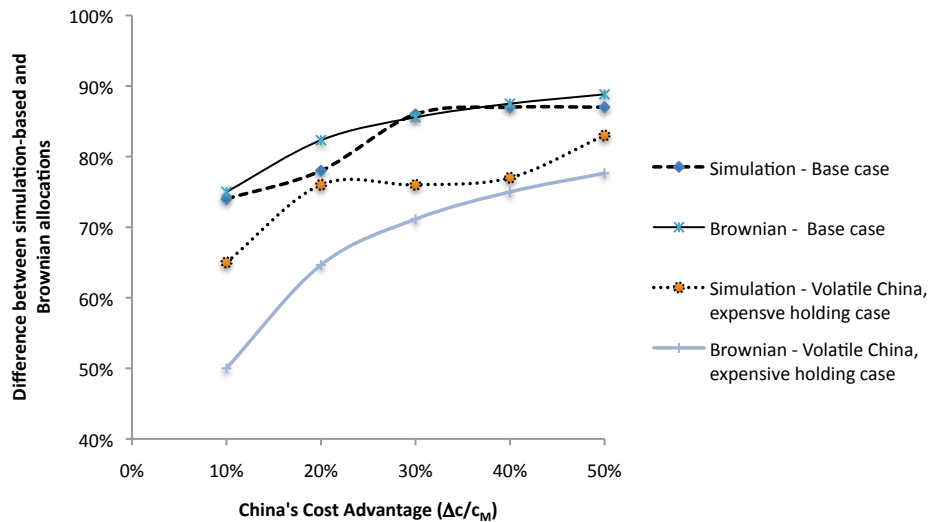
stock pairs and obtained both optimal base stock levels using a numerical search over this grid. An extensive numerical study in which we also optimized over the contingent China supply rate $\mu_C < \lambda$ under a dual base stock policy found no significant improvement, suggesting that the dual base stock performance is relatively insensitive to the contingent China supply rate. This also suggests that the optimal China supply rate for our single base-stock policy remains nearly optimal for the dual base-stock policy, echoing the finding in Scheller-Wolf et al. (2006) that compares single-index with dual-index policies.

Figure 6 depicts the percentage cost improvement as a function of China's relative cost advantage under the more complex (optimized) dual base stock policy relative to the (optimized) TBS for three parameter cases. (The sample error was less than 5%.) In all these settings the total cost difference was, as expected, increasing in China's cost advantage. The reason behind this monotonicity is that the China supply increases as the China cost advantage increases. This is accompanied with a rising risk of excess inventory (overshoots above the base-stock level). In those conditions, a second base-stock level, above which the firm can "shut-down" supply from China, is increasingly beneficial. Similarly, the dual base stock policy becomes more attractive when the holding cost or the volatility increase.

While the total cost difference increases, it never exceeds 7.2% in our numerical study. This suggests that the use of the optimized TBS policy in practice (instead of a more complex policy) can be justified as long as the China's cost advantage is modest. In the motivating example of Figure 2 where the TBS was used, the China cost advantage was less than 30%.

Comparing the Brownian approximation to simulation-based optimization The second question addressed by our validation study is: Under the TBS policy, how close is our predicted allocation (using the exact analysis of the Brownian approximation in Prop. 8) relative to the optimal allocation obtained by simulation-based optimization?

Figure 7 shows the allocation levels using both methods for two cases. In the base case, the allocations under both methods are practically identical. Even in the unfavorable case of high volatility and high holding cost, the Brownian approximation still performs well as long as China's

Figure 6 Comparing the total cost under TBS and under dual base stock policies.**Figure 7** Optimal Allocation: Simulation-Based Optimization vs. Brownian Approximation

cost advantage is not too small. The latter is not surprising given that, with small cost advantage and high volatility and holding costs, the China allocation becomes so small that it falls outside the heavy-traffic regime where the Brownian approximation is assured to be good.

In addition, our numerical study shows that the (non-reported) total cost difference between the allocation obtained using the simulation-based optimization and the Brownian approximation never goes above 2%, even when the allocations are quite different.

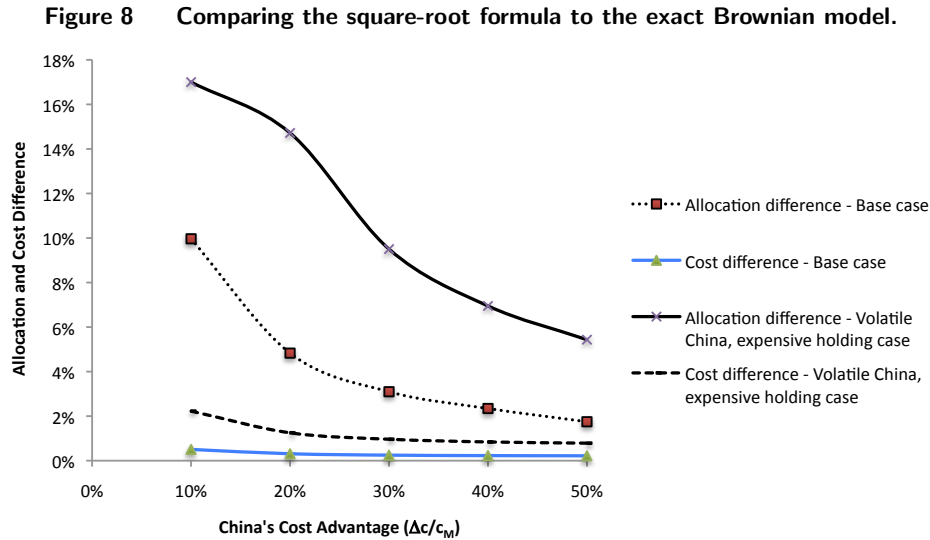
Comparing the square-root formula to the Brownian approximation The third question we address is: How well does the square-root formula (Prop. 9) perform relative to the exact

analysis of the Brownian model (using Prop. 8), both in terms of total cost minimization and allocation prediction? Figure 8 shows the difference between the allocation predicted by the square root formula and the allocation that solves the non-linear first-order condition of Prop. 8. It also shows the corresponding difference in the Brownian total cost estimate evaluated at these two allocations. The figure shows these results as a function of the China cost advantage in the base case and in the “Volatile China, Expensive holding” case.

One can observe that the square-root allocation is fairly close to the exact Brownian allocation, as long as the China cost advantage is not too small. Indeed, in our numerical study, the allocation difference is always below 10% in the base case and below 5% as long as the cost advantage exceeds 10%. For the more volatile case, such an accuracy level is attained only when the cost advantage is greater. The reason for this discrepancy is the same one mentioned above: the approximation works best when the fraction of orders coming from China is larger than 70% which requires a significant Chinese cost advantage in a volatile setting (as shown in Figure 7). It is important to note that even when the allocations are not close, their corresponding total cost difference remains small due to the flatness of the inventory cost curve around its minimum (as shown in Figure 4). The main implication is that a simple allocation, such as the 3/4 rule suggested by the consultants in our motivating example, can be a reasonable rule of thumb: In the absence of specific data or the ability to conduct a more thorough analysis, it results in an allocation that is quite robust in terms of total cost implications.

Practice-based validation study The numerical study reported so far assumed parameter values that have traditionally been used in the literature and our goal so far was to assess the quality of our analytic approximations and of the TBS policy. Now, in an attempt to test the robustness of these results, we next calibrate the parameters using the actual data observed in practice in our motivating example. To this end, we calibrate the holding cost and the demand volatility using real data.

The monthly demand experienced by the firm varied between 5,000 and 67,000 units while the sourcing cost c_C varied between a few hundred to a couple thousand dollars. In addition,

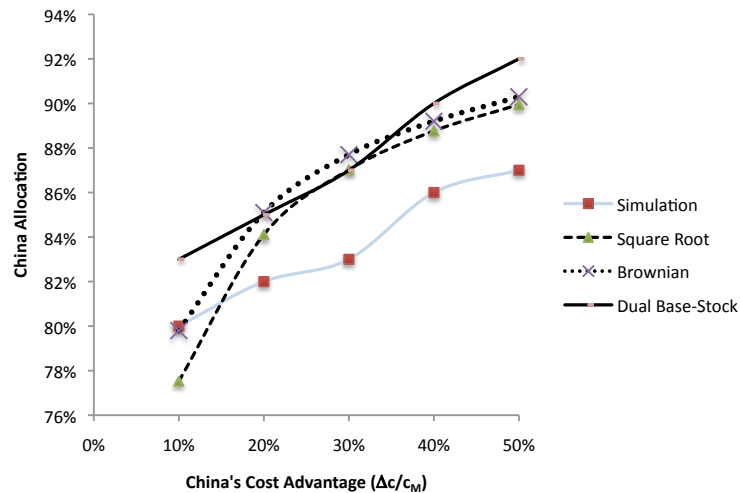


the annual holding cost h which equals (interest rate + physical holding cost) c_C was estimated at about $.6c_C$. This yields the following order-of-magnitude values for our parameters: the scaled holding cost $h/\lambda \approx (.6 * \$1000/12)/5000 \approx \$0.01$ per unit. Next, the monthly demands in the actual observed data exhibited a coefficient of variation between 0.05 to 1.25. These values were converted into the coefficient of variation of inter-arrival times using renewal theory: $v_D = \lambda \times$ (the coefficient of variation of demand rate) = $5000 \times (0.05 \text{ to } 1.25) = 15 \text{ to } 88$.⁸ To simulate these arrival processes, we sampled interarrival times that were independent and identically gamma distributed.

We used these parameters to obtain both the allocation and the corresponding total cost using (1) simulation-based optimization, (2) the Brownian approximation, and (3) the square-root approximation all assuming a TBS policy, and (4) simulation-based optimization assuming a dual based stock policy. Figure 9 depicts the allocation from the four methods mentioned again as a function of the China cost advantage. Observe that the allocations predicted by the four methods are very close, demonstrating that (i) the 3/4 allocation rule and (ii) the accuracy of the approximations derived in this paper are both quite robust under practical settings.

⁸ The fact that the interarrival coefficients of variations are significantly higher than the monthly coefficient of variation is due to the fact that the latter exhibits strong aggregation effects. In addition, actual order patterns are staggered or batched. For example, an order for 1,000 units results in one high interarrival time followed by 999 interarrival times of 0.

Figure 9 Comparing the optimal allocation using parameters consistent with practice.



9. Summary and Discussion

The dual sourcing literature has traditionally focused on determining sophisticated dynamic policies that approach optimal performance. The main research question of this paper, however, is more strategic in nature: to determine optimal average sourcing allocation. We were able to answer this question by assuming a simpler policy that is used in practice. This tailored base-surge (TBS) policy echoes a fundamental tenet in strategy: it aligns the ordering patterns with the core competencies of the suppliers. The constant base allocation allows China to focus on cost efficiency while Mexico’s quick response is utilized only dynamically to guarantee high service. Our model is a first attempt to provide some theory and quantification of this intuition of tailoring the sourcing strategy.

The model provides the following insights and quantification on strategic allocation. First, we present an analytic characterization of the TBS dual sourcing policy that culminates in a simple square-root formula. This formula specifies the near-optimal strategic allocation that separates stochastic demand into “base” and “surge.” Second, we determine the target inventory level and the corresponding cost under this near-optimal allocation. Our formulas allow an estimation of working capital requirements under dual sourcing, which have been lacking in the literature. Third, we identify and value the key drivers of dual sourcing. The square-root formula suggests a classification into first and second order drivers and we discuss each one in detail. In particular, we discuss the key role of supply and demand volatilities in dual sourcing. Our mode of analysis allows us

to go beyond the typical assumptions of independence and also discuss the impact of serial time correlations as well as intra demand-supply correlations.

A numerical study demonstrates that the results are robust and validates practice. We demonstrate robustness by showing that the TBS policy is near optimal in terms of total cost minimization; that the Brownian approximation results in reasonably accurate estimations of allocation and cost; and that the square root formula results in reasonably accurate predictions of minimal cost. The numerical study also validates the suggested practice of allocating 3/4 to the slow source as a starting point. With more specific data, the 3/4 allocation can and should be further tailored to the specific demand and supply characteristics using our results.

As with every model, ours has limitations. We do not explicitly model scale economies in ordering or production. Our results, however, show the presence of scale economies (our expressions are non-linear in the demand rate λ) due to statistical economies of scale. Our policy assumes that feedback control on China is not feasible; which precludes the comparison of dual sourcing with single sourcing from China. If one allows feedback control, this comparison follows the same lines as our comparison with single sourcing from Mexico. Finally, we have focused on a single product and a single market setting under centralized control. Future work should extend to multi-product, multi-market settings under decentralized control.

Acknowledgments: We are grateful to Cort Jacoby, Ruchir Nanda, and Brian Bodendein from Deloitte Consulting and to Achal Bassamboo, Marty Lariviere, Hyoduk Shin, and Kellogg seminar participants for detailed suggestions that improved the content of this paper.

Appendix

PROPOSITION 10. *Let $\{(D(t), S_C(t), S_M(t)) : t \geq 0\}$ be a general stationary process with interarrival times $\{(U_i, V_i^C, V_i^M) : i \in \mathbb{N}\}$. Then, its asymptotic infinitesimal variances are:*

$$\begin{aligned} \sigma_2^2 = & \lambda^3 \left(\text{var}U_1 + 2 \sum_{k=1}^{\infty} \text{covar}(U_1, U_{k+1}) \right) + \mu_C^3 \left(\text{var}V_1^C + 2 \sum_{k=1}^{\infty} \text{covar}(V_1^C, V_{k+1}^C) \right) \\ & - 2\lambda\mu_C \left(\min(\lambda, \mu_C) \text{covar}(U_1, V_1^C) + \mu_C \sum_{i=1}^{\infty} \text{covar}(U_{i+1}, V_1^C) + \lambda \sum_{j=1}^{\infty} \text{covar}(U_1, V_{j+1}^C) \right) \end{aligned}$$

$$\begin{aligned} \sigma_1^2 = & \sigma_2^2 + \mu_M^3 \left(\text{var}V_1^M + 2 \sum_{k=1}^{\infty} \text{covar}(V_1^M, V_{k+1}^M) \right) \\ & - 2\lambda\mu_M \left(\min(\lambda, \mu_M) \text{covar}(U_1, V_1^M) + \mu_M \sum_{i=1}^{\infty} \text{covar}(U_{i+1}, V_1^M) + \lambda \sum_{j=1}^{\infty} \text{covar}(U_1, V_{j+1}^M) \right) \\ & + 2\mu_C\mu_M \left(\min(\mu_C, \mu_M) \text{covar}(V_1^C, V_1^M) + \mu_C \sum_{i=1}^{\infty} \text{covar}(V_{i+1}^M, V_1^C) + \mu_M \sum_{j=1}^{\infty} \text{covar}(V_1^M, V_{j+1}^C) \right). \end{aligned}$$

References

- Barankin, E. W. 1961. A delivery-lag inventory model with an emergency provision. *Naval Research Logistics Quarterly* **8**(3) 285–311.
- Bradley, J. R. 2004. A Brownian approximation of a production-inventory system with a manufacturer that subcontracts. *Operations Research* **52**(5) 765–784.
- Bradley, J. R. 2005. Optimal control of a dual service rate M/M/1 production-inventory model. *European Journal of Operational Research* **161** 812–837.
- Bradley, J. R., P. W. Glynn. 2002. Managing capacity and inventory jointly in manufacturing systems. *Management Science* **48**(2) 273–288.
- Browne, S., W. Whitt. 1995. Piecewise-linear diffusion processes. J. H. Dshalalow, ed., *Advances in Queueing: Theory, Methods, and Open Problems*. CRC Press, Boca Raton, FL, 463–480.
- DeCroix, G., J. S. Song, P. Zipkin. 2005. A series system with returns: Stationary analysis. *Operations Research* **53**(2) 350–362.
- Ding, Q., L. Dong, P. Kouvelis. 2007. On the integration of production and financial hedging decisions in global markets. *Operations Research* **55**(3) 470–489.
- Fleischmann, M., R. Kuik, R. Dekker. 2002. Controlling inventories with stochastic item returns: a basic model. *European Journal of Operational Research* **138** 63–75.
- Fukuda, Y. 1964. Optimal policies for the inventory problem with negotiable leadtime. *Management Science* **10**(4) 690–708.
- Lau, H. S., L. G. Zhao. 1994. Dual sourcing cost-optimization with unrestricted lead-time distributions and order-split proportions. *IIE Transactions* **26**(5) 66–75.
- Lu, X. L., J. A. Van Mieghem. 2008. Multimarket facility network design with offshoring applications. *Manufacturing & Service Operations Management* Published online in Articles in Advance, April 17, 2008, DOI: 10.1287/msom.1070.0198.

- Moinzadeh, K., S. Nahmias. 1988. A continuous review model for an inventory system with two supply modes. *Management Science* **34**(6) 761–773.
- Moinzadeh, K., C.P. Schmidt. 1991. An (S-1, S) inventory system with emergency orders. *Operations Research* **39**(2) 308–321.
- Porteus, E. L. 2002. *Foundations of Stochastic Inventory Theory*. Stanford University Press, Stanford, CA.
- Rosenshine, M., D. Obee. 1976. Analysis of a standing order inventory system with emergency orders. *Operations Research* **24**(6) 1143–1155.
- Scheller-Wolf, A., S. K. Veeraraghavan, G-J van Houtum. 2006. Effective dual sourcing with a single index policy. *working paper, Wharton at the University of Pennsylvania, Philadelphia, PA*.
- Sheopuri, A., G. Janakiraman, S. Seshadri. 2007. New policies for the stochastic inventory control problem with two supply sources. *Working paper, Stern School of Business, New York University, NY, NY*.
- Song, J.S., P. Zipkin. 2008. Inventories with multiple supply sources and network of queues with overflow bypasses. *Management Science* Forthcoming.
- Tagaras, G., D. Vlachos. 2001. A periodic review inventory system with emergency replenishments. *Management Science* **47**(3) 415–429.
- Van Mieghem, J. A. 2008. *Operations Strategy: Principles and Practice*. Dynamic Ideas, Belmont, MA.
- Veeraraghavan, S. K., A. Scheller-Wolf. 2006. Now or later: Simple policy for effective dual sourcing in capacitated systems. *Operations Research* Forthcoming.
- Whittmore, A. S., S. C. Saunders. 1977. Optimal inventory under stochastic demand with two supply options. *SIAM Journal of Applied Math.* **32**(2) 293–305.
- Zipkin, Paul H. 2000. *Foundations of Inventory Management*. McGraw-Hill, New York, NY.