Dual Sourcing and Smoothing under Non-Stationary Demand Time Series: Re-shoring with SpeedFactories

Robert N. Boute
Research Center for Operations Management, KU Leuven, Belgium
Technology & Operations Management Area, Vlerick Business School, Belgium

Stephen M. Disney
Centre for Simulation, Analytics and Modelling, University of Exeter Business School, United Kingdom

Joren Gijsbrechts
Católica Lisbon School of Business & Economics, Portugal

Jan A. Van Mieghem
Kellogg School of Management, Northwestern University, United States

March 13, 2018; Revised March 09, 2019; May 4 and Nov 28, 2020

To appear in *Management Science*

We investigate near-shoring a small part of the global production to local *SpeedFactories* that serve only the variable demand. The short lead time of the responsive SpeedFactory reduces the risk of making large volumes in advance, yet it does not involve a complete re-shoring of demand. Using a break-even analysis we investigate the lead time, demand, and cost characteristics that make dual sourcing with a SpeedFactory desirable compared to complete off-shoring. Our analysis employs a linear generalization of the celebrated order-up-to inventory policy to settings where capacity costs exist. The policy allows for order smoothing to reduce capacity costs and performs well relative to the (unknown) optimal policy. We highlight the significant impact of auto-correlated and non-stationary demand series, which are prevalent in practice yet challenging to analyze, on the economic benefit of re-shoring. Methodologically, we adopt a linear policy and normally distributed demand and use $Z-$transforms to present exact analyses.

Key words: Inventory Management, Order Smoothing, Order-Up-To Policy, Auto-Regressive Demand, Integrated Moving Average Demand, Global Outsourcing, Dual Sourcing, $Z-$transform.

1. Introduction

Re-shoring initiatives are growing in popularity. Margolis [2017] reports of a UK company re-shoring 30% of its production back from China for economic reasons. Annual Chinese labor costs rose from £2.5K in 2007 to £8K in 2017, with some key personnel being paid £55K. Over the same timespan, Chinese factory leases have doubled to £4 per sq ft (London is £8 per sq ft; other areas of the UK are much cheaper). Cheaper, and easier to program, labor-reducing CNC machines facilitate more in-house production. Add transport and coordination costs and off-shoring economics appear even less attractive. With product demand being more unpredictable and speed-to-market becoming more important, this re-shoring trend may continue.

Recently, the concept of a *SpeedFactory*[1] has emerged. A SpeedFactory is a small, local, fast response plant that supplements a large off-shore plant [Shotter and Whipp 2016]. Adidas and Nike announced they are bringing back some manufacturing to Europe and North America, using automation to keep labor costs down and make products quickly. Nevertheless, they still make the large majority of their goods in Asia, an intriguing fact our model analysis can explain[2]. The need for such a near-shore SpeedFactory is driven by the demand for new products with shorter and more unpredictable life-cycles, partially due to the increased focus on e-commerce which makes speed key to the company's long-term growth. In the past, Nike would start manufacturing when it received an order it had to deliver in six months. Now Nike is moving towards quick response manufacturing matched to consumer need. This reduces the time from-manufacturing-to-market to 10 days or less. Similarly, Adidas opened a SpeedFactory in Germany and in Atlanta to do more of its own responsive manufacturing. Recently, Adidas has rescinded on their bold decision and moved their robotic SpeedFactory back to Asia [Ziady 2019].

We are also working with a high volume European apparel manufacturer. They currently outsource their printable t-shirt production to a Bangladesh supplier. The company is suffering from long replenishment lead-times and pressure for working capital reduction; they are investigating the possibility of a local SpeedFactory to help cope with the highly unpredictable demand and short life cycles. We later show how our analysis supports their decision.

Herein, we study when, and to what extent, such a SpeedFactory creates value by analyzing the combined impact of: product demand, supply chain costs, inventory service levels, replenishment smoothing rules, lead times, and off-shore and near-shore facilities' labor flexibility. We consider a family of linear order policies under a broad class of normally distributed demand patterns, including auto-correlated and non-stationary demand.

1.1. Background and motivation

We study a focal company that previously outsourced production to a low cost, global, off-shore, supplier with a long lead-time. The off-shore supplier charges a unit price to deliver the item to an inventory location from which customers are served. Upon arrival, product ownership is transferred to the focal company. The focal company controls this inventory (used to serve demand) by issuing replenishment orders to the off-shore supplier. This arrangement has worked well in the past, but recently off-shore labor costs have risen dramatically[3]. This preference is driving up factory labor

[1] The SpeedFactory moniker comes from Adidas.

[2] While Adidas sell 301M pairs of shoes each year, they plan to make only 1M (0.33%) shoes a year in their German SpeedFactory. Nike's near-shore, purpose-built footwear factory with Flex will deliver 3M pairs of shoes to North America in 2018. This is only 0.23% of the predicted 1.3B shoes to be shipped in 2018 [Bain 2017].

[3] In conversations we have held with Chinese factory owners, technician rates have increased three-fold in the last decade; regular worker rates have doubled in the same time frame. Migrant workers are increasingly averse to working long hours in distant factories. Many would rather work closer to their family in service industries.

costs, and increased demand unpredictability that puts pressure on inventories, is eroding many of the benefits of off-shoring.

SpeedFactories provide an accurate response to demand fluctuations. This local production facility, owned and operated by the focal company, has a nominal working week with a reserve capacity for meeting short-term demand peaks via over-time. In this re-shored scenario the focal company is responsible for material, labor, and over-time costs of local production. The short lead time of re-shored production allows for tighter control of inventory levels, off-setting the potentially higher local labor costs.

1.2. Related literature

Interest in re-shoring strategies, where companies reverse off-shoring decisions and bring manufacturing back home, has gained momentum recently. Suggested drivers of this global (re-)location include: the rising labor cost in low-cost countries, the increased fuel cost and associated transport costs, and the fast response time and leaner supply chain associated with locating manufacturing closer to the end customer [Ellram et al. 2013]. In their analysis of the manufacturing off-shoring and back-shoring activities of 1663 German manufacturing companies, Kinkel and Maloca [2009] show that 16-25% of off-shoring is re-shored within four years. Rather than off-shoring or re-shoring all manufacturing activities, the dual sourcing literature considers the combined use of off-shoring and near-shoring. The dual sourcing literature is rich (see Yao and Minner [2017] for a review); we restrict our literature review to the most relevant contributions.

Whittemore and Saunders [1977] show when the lead times of both sources differ by more than one period, the optimal policy depends on the entire ordering history and requires multi-dimensional dynamic programming. Therefore, the literature proposes various heuristic policies. Rosenshine and Obee [1976] study the standing order policy which orders at a constant rate from the regular global source and uses a base-stock policy for the emergency replenishment. Allon and Van Mieghem [2010] refer to a standing order policy as a tailored base-surge (TBS) policy, where the regular source supplies the *base* demand and the fast source supplies the remaining *surge* demand using a base-stock policy. They give guidance on the off-shore vs. near-shore production volume. Janakiraman et al. [2015] show the TBS policy is optimal when demand comes from a two-point distribution and when the probability of the smaller (base) demand is sufficiently large. They also show that the TBS performance, relative to the optimal policy, improves as the off-shore lead time increases. Xin and Goldberg [2018] prove that TBS is asymptotically optimal as the lead time difference $L \to \infty$.

Building upon Song and Zipkin [2009], Song et al. [2017] consider a dual-sourcing setting with endogenous stochastic lead times within a queuing theory framework under iid Poisson demand

with exponentially distributed lead times. Hua et al. [2015] consider a periodic review dual sourcing inventory system with general lead times and iid demand over a finite, and an infinite, time horizon, and characterize the structural properties of the optimal orders. All the aforementioned order policies are developed for linear inventory holding and backlog costs and linear order costs from both sources.

Boute and Van Mieghem [2015] introduce a linear control rule to determine off-shored and re-shored volumes. Their model captures each source's lead time, capacity cost, and flexibility to work over-time. Linear control rules have been investigated successfully by several researchers in various environments; see, for example Graves [1999], Zipkin [2000, p.393], Aviv [2007] and Huh and Nagarajan [2010] and the references therein.

1.3. Contribution

We contribute to the off-shoring, re-shoring, and dual sourcing literature as follows. We believe this is the first paper to provide a stochastic analysis of dual sourcing under correlated and non-stationary demand. We show that such demand can significantly increase the viability of dual sourcing relative to the traditionally-assumed iid demand in dual sourcing research. Our analysis employs a linear generalization of the celebrated order-up-to inventory policy to settings where capacity costs exist. The policy allows for order smoothing to reduce capacity costs and performs well relative to the (unknown) optimal policy. Using a break-even analysis we offer conservative guidance on when a SpeedFactory is attractive given specific product demand characteristics, supply chain costs, and replenishment lead-times. We illustrate our findings with an industrial case.

The strength of our linearized model is analytical tractability for continuous normally distributed demand. Yet we acknowledge this linearity is also the key restriction of our analysis. We therefore provide extensive robustness analysis to show that this linearization is reasonable.

2. A production-inventory model with two replenishment sources

We consider a discrete-time, discrete-review production-inventory model where stochastic customer demand, d_t, is satisfied from finished goods inventory (FGI) i_t, and excess demand is backlogged. In each period $t \in \mathbb{N}$, the firm can replenish its FGI from two sources, as illustrated in Figure 1: (a) A local (re-shored) wholly-owned SpeedFactory with the smallest possible lead-time, corresponding to an order received in time to fill next period's demand (following [Zipkin 2000, p404] we say that the risk period or total lead time is $L = 1$, this includes the one period review). (b) A global off-shore outsourced supplier with lead time $L > 1$. Denote the local and global replenishment orders by q_t and q_t^g, respectively. Period t starts by receiving orders q_{t-1} and q_{t-L}^g. Then demand d_t is

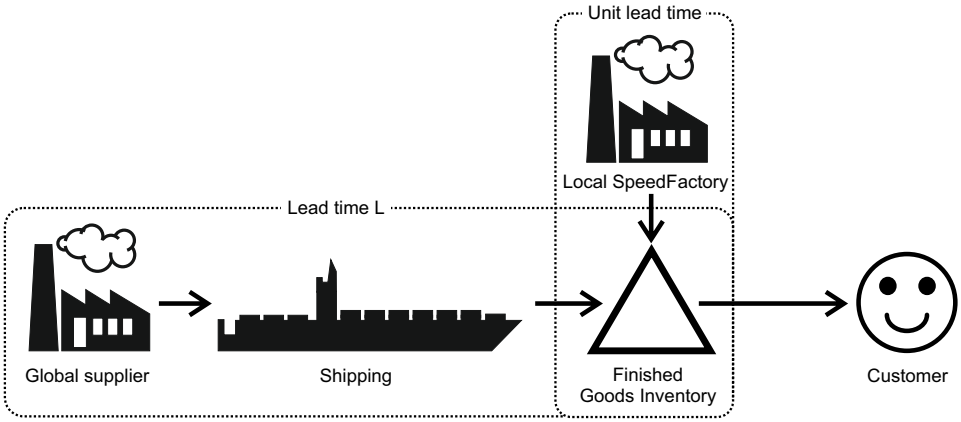


Figure 1 Our supply chain replenishes inventory from a global supplier and a local SpeedFactory.

observed and satisfied. Finally, orders q_t and q_t^g are placed based on the finished goods inventory i_t which evolves via

$$i_t = i_{t-1} - d_t + q_{t-L}^g + q_{t-1}. \tag{1}$$

In period t, inventory incurs the usual holding and backlog costs[4]:

$$C_t^i = h[i_t]^+ + b[-i_t]^+, \tag{2}$$

where h and b are the per period per unit holding and backlog cost, and $[x]^+ = \max[x, 0]$.

Inventory replenishment costs depend on the order quantity and on the source. The purchasing cost in period t from the global external supplier, C_t^p, is

$$C_t^p = pq_t^g, \tag{3}$$

where p is the total landed unit purchasing cost (p includes all the costs involved in producing and delivering the product). We assume the global supplier can outsource excess demand beyond their capacity to a third party within the same quality parameters, cost structure, and lead time. Hence, the global supplier effectively has no capacity limit.

In contrast, the internal SpeedFactory maintains an installed nominal production capacity of $k \geq 0$ units per period but can run overtime to produce $q_t > k$. The local production cost in period t, C_t^q, is[5]

$$C_t^q = uk + um[q_t - k]^+, \tag{4}$$

[4] Notation: $(\cdot)$ is used to clarify precedence, $[\cdot]$ for functions or operators, and $\{\cdot\}$ for lists.

[5] Per unit material and direct energy costs could also be included into the capacity cost function, but as this is entirely volume related, it will not affect the capacity decision, and any influence it would have in our break-even analysis later could be readily included in the variable p. To create a parsimonious model we have omitted this term.

where u is the per unit labor cost of producing an item during normal working hours (at full capacity)[6] and um is the per unit cost to produce excess demand in over-time[7,8]. If the local order is less than the installed capacity, $q_t < k$, workers stand idle, but still receive their full wage for the working week[9]. Otherwise, over-time cost of um is incurred which is at least as expensive as work in the nominal working week. The multiplier $m \geq 1$ can be thought of as a *flexibility coefficient*. When $m = 1$, total labor flexibility is available, all products are produced at a unit cost of u; as there are no guaranteed hours, there is also no idling. As $m > 1$ increases, labor is increasingly less flexible; when $m \to \infty$, labor is totally inflexible.

Program OPT Let Π denote the set of non-anticipatory order policies $\pi = \{(q_t, q_t^g) \in \mathbb{R}^2 : t \in \mathbb{N}_0\}$. The objective is to find an order policy $\pi \in \Pi$ and a SpeedFactory size $k \geq 0$ that minimizes the long-run average total cost per period $C^{\pi,k}$, where

$$C^{\pi,k} = \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} (C_t^i + C_t^p + C_t^q) \geq 0. \tag{5}$$

Let $C^{\mathrm{OPT}} = \inf_{\pi\in\Pi, k\geq 0} C^{\pi,k} \geq 0$ denote the optimal cost.

Unfortunately, the multidimensional dynamic program (DP) to solve Program OPT is not analytically tractable and the structure of the optimal policy remains unknown for general lead times. Therefore, we shall consider a family of *linear* dual sourcing order policies that are analytically tractable and are used in practice. When local supply is uncapacitated ($k = \infty$), Xin and Goldberg [2018] proved that our family of policies is asymptotically optimal with iid demand when $L \to \infty$, so they are expected to perform well in our setting where $L \gg 1$. These linear policies give an upper bound on C^{OPT} that we will quantify for settings where the DP can be solved numerically.

Note that the policies in OPT, including our linear policies, may place negative orders. To compare their performance with traditional inventory theory, which often assumes non-negative orders, we also consider the restricted program OPT_+:

[6] u includes depreciated overheads, indirect labor, and indirect energy allocated to each unit of nominal capacity.

[7] We note that over-time customs vary from country-to-country. In many Western countries working hours are limited by law and workers expect to be paid a premium for hours worked at night, on the weekend, and in excess of a standard working week. In other countries, working hours are not limited and workers do not receive an over-time premium. Even in the cases where premiums are paid, there are often other conventions to consider. For example, unions may insist that only full shifts of over-time are offered, allowing workers of recover their travel costs. Such highly contextual factors are not included in (4).

[8] We are aware of companies who have sold their old production equipment to local subcontractors when they upgraded their facilities. When these companies experience peak demand, a proportion of the excess demand is produced by these subcontractors. This set-up also matches our assumptions, allows for a linear analysis, and affords an alternative way to interpret (4). That is, in a 24/7 operation, $[q_t - k]^+$ is the subcontractors demand and um is the per unit subcontractor cost. Alternatively, excess demands at the local factory could also be produced by the global supplier and air-shipped to the FGI (in unit lead time). In this case the um incorporates the premium transportation costs.

[9] In a highly automated SpeedFactory, not only the workers, but also the machines stand idle when demand is less than the installed capacity; when machines incur linear depreciation costs, the equations remain.

Program OPT$_+$ Let Π_+ denote the restricted set of non-negative, non-anticipatory order policies $\pi = \{(q_t, q_t^g) \in \mathbb{R}_+^2 : t \in \mathbb{N}_0\}$ and $C^{\mathrm{OPT}_+} = \inf_{\pi \in \Pi_+, k \geq 0} C^{\pi,k}$. Clearly, $C^{\mathrm{OPT}} \leq C^{\mathrm{OPT}_+}$. We will define a "non-negative version" of our linear order policies whose cost provides an upper bound to C^{OPT_+}. In addition, by bounding the probability that our linear policies place negative local orders by $\epsilon > 0$, the linear policy is also conjectured to provide an upper bound to C^{OPT_+}. Proving this conjecture is exceedingly difficult and beyond the scope of this paper but an extensive numerical study that we present later supports the conjecture.

In addition to C^{OPT} and C^{OPT_+}, we will also compare the linear policy's cost to the two single-sourcing benchmarks: Let $\pi^l = \{(q_t, 0) \in \mathbb{R}_+^2 : t \in \mathbb{N}_0\}$ and $\pi^g = \{(0, q_t^g) \in \mathbb{R}_+^2 : t \in \mathbb{N}_0\}$ denote any single-sourcing local or global policy, respectively. These single-sourcing policies provide an upper bound:

$$0 \leq C^{\mathrm{OPT}} \leq C^{\mathrm{OPT}_+} \leq \min\{C^{\pi^l,k}, C^{\pi^g,k}\}. \tag{6}$$

It is well known that the optimal single-sourcing global replenishment policy is the celebrated Order-Up-To (OUT) policy. However, when the local plant is constrained with piecewise-linear convex production costs the OUT policy is not the optimal single-sourcing local policy. Boute and Van Mieghem [2015] show that the optimal single-sourcing local order policy in a setting under piece-wise linear convex local production costs with unit lead-time is characterized by a dual-base-stock policy that features a *region of inaction* (RoI). This "RoI policy" operates as follows: When the inventory exceeds the higher base-stock level y_1, no local order is placed. When the inventory is below y_1, we first use up the local production capacity k to replenish locally. If this raises the inventory to above the lower base stock y_2, we stick to this local order of k units and do not use over-time to replenish all observed demand; otherwise, we use over-time to raise the inventory to y_2. Thus there is a region of inaction where we order maximal k but less than the demand[10]. Our numerical robustness section will show that our linear policy performs close to the RoI policy to place local orders, q_t. Mathematically, the RoI$_+$ policy places local orders as:

$$q_t = \begin{cases} (y_1 - i_t)^+, & i_t > y_1 - k \\ k, & y_2 - k \leq i_t \leq y_1 - k \\ y_2 - i_t, & i_t < y_2 - k. \end{cases} \tag{7}$$

We denote the linearized version that drops the positive-part operator in equation (7) and allows negative orders by RoI. Unfortunately there are no simple solutions for the optimal base-stock levels and capacity level because the RoI policy is not a demand replacement policy. Numerically optimizing the four parameters is computationally demanding and non-obvious.

[10] The marginal over-time cost exceeds the marginal benefit of raising inventory (reducing backlogging relative to holding costs). It is better to wait and replenish in the future at regular cost versus now at an over-time cost.

3. TBS policies with local linear POUT control under normal demand

The linear policies we consider are TBS policies that place a constant global order under iid and auto-correlated demand (and, as we shall show, a dynamic order under non-stationary demand) and use the local SpeedFactory to respond to demand variation and control inventory. TBS policies are used in practice and convenient as they directly facilitate specifying offshoring contracts in terms of *strategic allocations*: Let $\gamma \in [0,1]$ denote the decision variable that we refer to as the *strategic SpeedFactory allocation*, which is the long-run fraction of the expected demand per period μ to be produced locally. Then, the global constant order when using a TBS policy is

$$q_t^g|_{\text{TBS}} = \mu(1-\gamma). \tag{8}$$

With an off-shore constant order generated by (8), the local SpeedFactory is the only means to control the FGI that is subject to stochastic demand $d_t - \mu(1-\gamma)$. Analytically, the TBS policy reduces the dual sourcing problem to a much simpler single sourcing problem where it only remains to specify the local order policy.

To allow for an analytic study of the appropriate SpeedFactory allocation $\gamma^\star$ and capacity $k^\star$, we consider a TBS policy where local orders are controlled by a generalization of the OUT policy, called the *Proportional Order-Up-To* (POUT) policy [Disney et al. 2016]:

$$q_t|_{\text{TBS-POUT}} = \mu\gamma + (1-\alpha)\left(i^\star - i_t\right), \tag{9}$$

where $i^\star$ is the desired safety stock level and $\alpha \in (-1,1]$ is a smoothing parameter. Without smoothing ($\alpha = 0$), the POUT policy reduces to the familiar OUT policy which attempts to restore inventory deviations in a single replenishment decision. When $0 < \alpha < 1$, only a fraction $(1-\alpha)$ of the actual inventory deviation is corrected in each decision, deferring the remaining correction to be recovered in future decisions. This creates a smoothing policy where the variability of the orders can be less than the variability of the demand [Deziel and Eilon 1967]. From a control theory perspective, the feedback control is proportional and thus α is also known as a *proportional feedback controller*. With total smoothing in a TBS policy, $\alpha = 1$, a constant, or level, production policy is obtained; a negative smoothing parameter indicates order variance amplification (bullwhip). The key benefit from the POUT policy with $0 < \alpha \leq 1$ is a reduction of the production order variability, which reduces factory capacity requirements, at the expense of slower control, and thus higher variance, of inventory. Boute and Van Mieghem [2015] show that the POUT policy closely tracks the optimal RoI policy in the presence of capacity costs; our robustness section also confirms this.

With the linear POUT order rule, the inventory in each period is a linear superposition of demands. Given that a linear superposition of normal distributions remains normally distributed,

the inventory, and thus also the local order (9), in each period is normally distributed if d_t is normally distributed. If the limiting inventory and local order distributions exists, they are normally distributed with variances denoted by σ_i^2 and σ_q^2, which are to be determined. We can then evaluate and optimize system performance analytically. Let $\{\phi[\cdot], \Phi^{-1}[\cdot]\}$ denote the probability density function and the inverse of the cumulative distribution function of the standard normal distribution respectively. (All proofs are housed in Appendix A).

Proposition 1 *Let $d_t = N(\mu, \sigma)$ be identically, normally distributed for each t. When dual sourcing using a linear TBS-POUT policy, the optimal safety stock level and inventory cost are,*

$$i^\star = \sigma_i z_i, \quad z_i = \Phi^{-1}\left[b/(b+h)\right], \quad \textit{and} \quad C_i^\star = \sigma_i(h+b)\phi\left[z_i\right] \geq 0. \tag{10}$$

In the linearized model, the optimal local SpeedFactory capacity and cost are,

$$k^\star = \mu\gamma + \sigma_q z_q, \quad z_q = \Phi^{-1}\left[(m-1)/m\right], \quad \textit{and} \quad C_q^\star = u\mu\gamma + um\sigma_q\phi[z_q] \geq 0. \tag{11}$$

The purchasing cost from the off-shore supply under TBS is,

$$C_p^\star = p(1-\gamma)\mu \geq 0. \tag{12}$$

The associated total minimal long-run average cost is

$$C^{\text{TBS-POUT},k^\star} = C_i^\star + C_q^\star + C_p^\star = \sigma_i(h+b)\phi\left[z_i\right] + \sigma_q um\phi\left[z_q\right] + ((u-p)\gamma + p)\mu \geq 0. \quad \square \tag{13}$$

Note, Proposition 1 allows demands to be correlated over time. Our linear analysis also allows capacity $k^\star < 0$ for extremely small values of $\gamma < -\frac{\sigma_q}{\mu}\Phi^{-1}\left[\frac{m-1}{m}\right]$ (in the numerical setting used throughout the paper, $k^\star < 0$ when $\gamma < 0.023$). Imposing a positive $[k^\star]^+$ for these instances results, however, in non-normal orders, such that we cannot find the critical fractile. To preserve linearity, we alternatively increase γ to ensure $k^\star \geq 0$. Increasing γ also reduces the probability that a local production order is negative and it limits the proportion of production being produced in overtime. By this, increasing γ ensures the accuracy of our linear approximation to the true non-linear problem. We explore this in the robustness analysis in section 4.3. Assuming linearity, it only remains to determine (1) the inventory and local order variances and (2) the smoothing parameter (which all depend on the correlation structure). The significant benefit of our linear control approach is that these variances can be computed in closed-form using Z-transforms as explained in Appendix B. When the normal demands are independent and identically distributed (iid), the variances under the linear TBS-POUT policy are:

$$\sigma_i^2|_{\text{iid}} = \sigma^2\left(\frac{1}{1-\alpha^2}\right) \quad \text{and} \quad \sigma_q^2|_{\text{iid}} = \sigma^2\left(\frac{1-\alpha}{1+\alpha}\right). \tag{14}$$

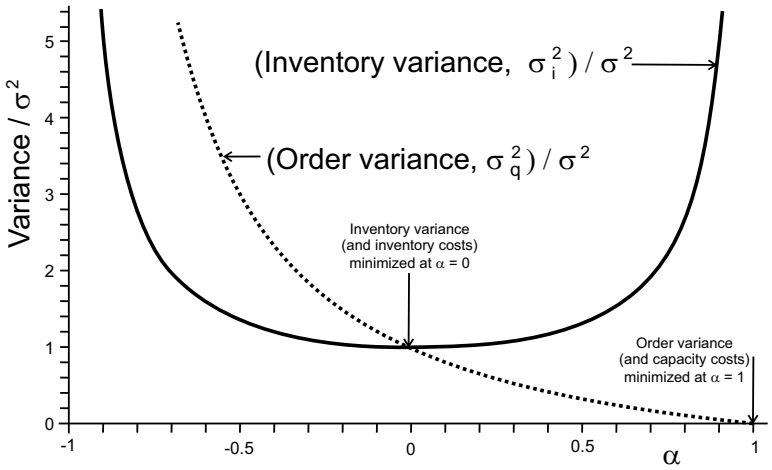


Figure 2 By tuning the smoothing parameter α, the POUT policy can trade-off the local order variance (which drives the local SpeedFactory's production cost) and the inventory variance.

In the absence of smoothing ($\alpha = 0$), POUT reduces to the classic, demand-replacing OUT policy and both variances equal the demand variance. Incorporating smoothing ($0 \le \alpha \le 1$) adds a control variable to manage the trade-off between inventory and local order variability that is shown in Figure 2. The local order variance is decreasing convex in the smoothing parameter α with a minimum of $\sigma_q^2 = 0$ under complete smoothing at $\alpha = 1$. The inventory variance is convex in α with a minimum of σ_i^2 at $\alpha = 0$ (recall that OUT minimizes inventory costs) and asymptotes to ∞ at $\alpha = \{-1, 1\}$. Under a TBS-POUT policy, both variances are independent of $\{L, \gamma, \mu\}$ and therefore so is the smoothing feedback parameter $\alpha^\star$ that minimizes the sum of inventory and local capacity costs.

For any capacity k and (possibly correlated) normal demands, the cost (13) can be written as

$$C^{\text{TBS-POUT},k} = \varphi(\sigma_i + \lambda(\sigma_q - \sigma_i)) + A, \tag{15}$$

where $\lambda \in [0, 1]$ indicates the capacity intensity and $\varphi \ge 0$ the scaling factor,

$$\lambda = um\phi[z_q]/\varphi \text{ and } \varphi = (h+b)\phi[z_i] + um\phi[z_q], \tag{16}$$

and A are other terms in the cost function that are not influenced by α and thus can be ignored. The optimal smoothing parameter, α, can then be found:

Proposition 2 *Let d_t be identically, normally distributed for each t. For a TBS policy with local linear POUT control and given capacity intensity $\lambda \in [0, 1]$, the optimal smoothing level is*

$$\alpha^\star = f^{-1}[\lambda] \text{ where } f[\alpha] = \frac{\sigma_q}{\sigma_q - \sigma_i \left(\frac{d\sigma_q^2}{d\alpha} \Big/ \frac{d\sigma_i^2}{d\alpha} \right)}. \tag{17}$$

The first order condition (17) is always easily solved graphically or numerically as f is an inverse function that defines for each α the corresponding unique *optimal* λ. For normal iid demands, the condition can be solved analytically. Substituting the standard deviations (14) and their derivatives in (17) reveals that $\alpha^* = \lambda$ is a unique solution for $C^{\text{TBS-POUT},k^\star}$ and is unimodal in α:

Corollary 1 *With normal iid demand, the optimal smoothing for a TBS policy with local linear POUT control and given capacity intensity* $\lambda \in [0,1]$ *is* $\alpha^\star = \lambda$.

The corollary states that when labor is more rigid, capacity becomes more constrained and one should smooth more, exactly because it reduces production variance.

Having fully specified the minimal cost when dual sourcing using a linear TBS-POUT policy that allocates fraction γ of the mean demand μ for production to the near-shore SpeedFactory, the last optimization considers the optimal allocation $\gamma^\star$. Equation (13) directly shows that the dual sourcing cost $C^{\text{TBS-POUT},k^\star}$ is linear in γ with coefficient $(u-p)\mu$. Aside from the knife-edge case at $u = p$ (in which case the cost is independent of γ), the optimal γ is a boundary solution:

Proposition 3 (Bang-bang) *Let* $d_t = N(\mu,\sigma)$ *be identically, normally distributed for each* t. *When dual sourcing using a TBS policy with local linear POUT control, the optimal strategic allocation* $\gamma^\star = 0$ *if* $u > p$ *and* $\gamma^\star = 1$ *if* $u < p$.

When the local cost u during normal working hours is below the offshore price p the optimal strategic allocation $\gamma^\star = 1$ reduces to complete reshoring; i.e., single sourcing from the local Speedfactory. The more interesting and relevant case is $u > p$ for which $\gamma^\star = 0$. This means that the local SpeedFactory is allocated a normal demand with mean 0 and standard deviation σ: it produces *on average* zero volume but serves *all the variability*. This implies that 50% of its orders are negative and in the next section we investigate how to adjust this result when orders must be non-negative. Yet the key insight will be robust: dual sourcing with a SpeedFactory that only serves the variable demand (hence with a capacity of the scale of σ) can be attractive even when $u > p$.

Recall, however, a linear TBS-POUT policy with $\gamma^\star = 0$ still dual sources and *differs* from single sourcing from the global source. Indeed, while TBS-POUT can replicate single local sourcing, it cannot replicate single offshoring because it requires a SpeedFactory serving variable demand. To analyze whether dual sourcing with a TBS-POUT policy is viable, we compare its cost to the optimal single-sourcing cost from the uncapacitated off-shore supplier. Here a traditional OUT policy is optimal and the associated inventory and purchasing cost serves as a benchmark:

$$C^s = \sigma_{i,s}(h+b)\phi\left[z_i\right] + p\mu, \tag{18}$$

with $\sigma_{i,s} = \sigma\sqrt{L}$ the standard deviation of the inventory under single sourcing with iid demand.

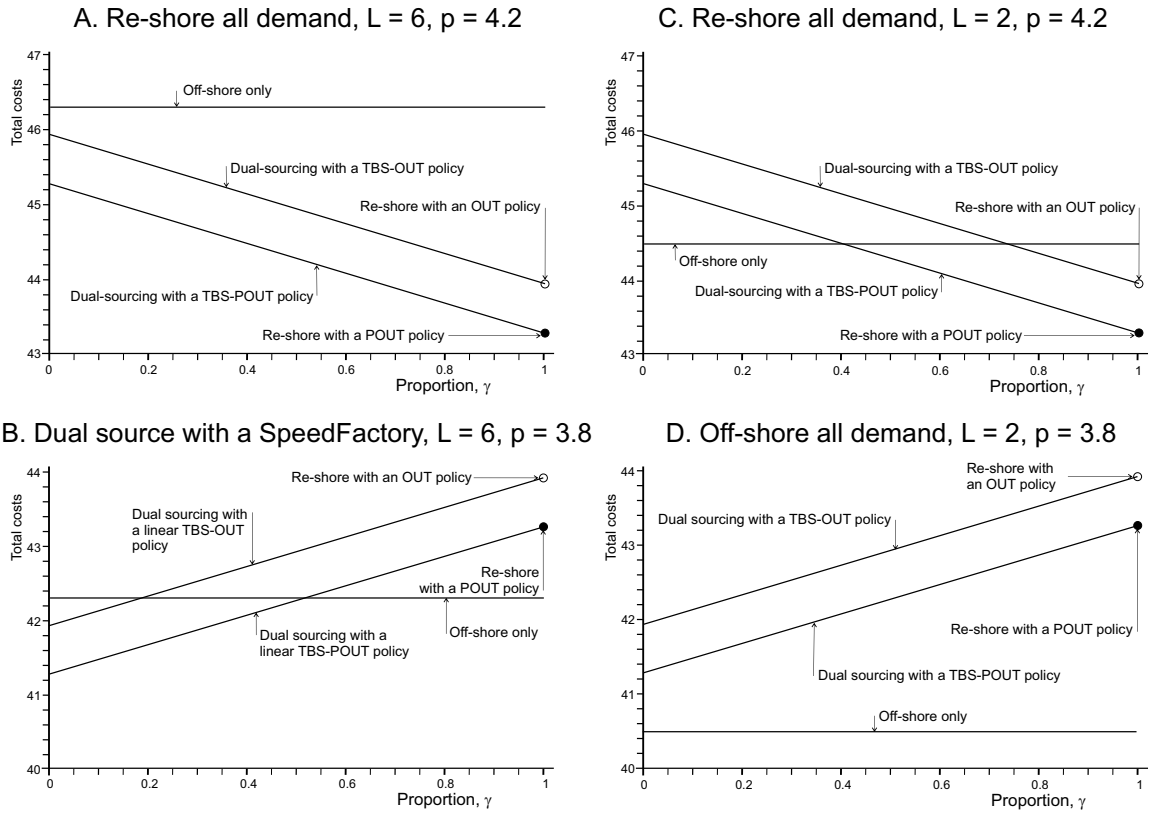


Figure 3 Comparing costs when single-sourcing versus dual sourcing with linear TBS-POUT and TBS-OUT policies under normal iid demand. Parameters: $\alpha = \alpha^\star, h = 1, b = 9, u = 4, m = 1.5, \sigma = 1, \mu = 10$.

Figure 3 compares the optimal dual sourcing cost under the linear TBS-OUT and TBS-POUT policies for two off-shore purchase prices and two off-shore lead times. Given that TBS-POUT can replicate TBS-OUT (when setting smoothing $\alpha = 0$), TBS-POUT dominates TBS-OUT. Figure 3 demonstrates that smoothing local SpeedFactory orders with TBS-POUT reduces local order variability (and capacity costs) below those of the TBS-OUT policy, enhancing the value of a SpeedFactory. While the lead time greatly influences the single-offshore cost C^s (the horizontal line), by definition the TBS costs (the sloped lines) are independent of the leadtime. The local unit production cost $u = 4$. For the parameters used in Figure 3, we can observe: When the outsourcing price is $p = 3.8$ in the bottom panels, complete off-shoring with short off-shore lead time dominates both dual sourcing with TBS-POUT and complete re-shoring (panel D). Yet with a long lead time, dual sourcing with TBS-POUT allocating only the variable demand to a SpeedFactory dominates (panel B). The top panels, A and C, assume a higher off-shoring purchasing price $p = 4.2$. With longer lead time ($L = 6$ in panel A), dual sourcing with TBS-POUT always dominates complete off-shoring. Given linearity, complete re-shoring is optimal: $\gamma^\star = 1$. With shorter off-shore lead time ($L = 2$ in panel C), full off-shoring is more attractive, yet is still dominated by complete re-shoring.

In the next section, we analyze the robustness of our key insight that dual sourcing with a SpeedFactory that only serves the variable demand can be attractive even when $u > p$.

4. Robustness Analysis of the linear TBS-POUT Policy

In this section we investigate three robustness questions:

1. How reasonable is the linear TBS-POUT policy? We present optimality gaps of TBS-POUT with Program OPT for a wide testbed of parameters.
2. What is the impact of demand distributions that are not normal?
3. What is the impact of a non-negativity constraint on local orders?

4.1. TBS-POUT performance and optimality gaps

The purpose of this section is to ascertain the performance of the TBS-POUT policy relative to the optimal policy computed numerically by solving its dynamic program. Due to the curse of dimensionality, we can only solve Program OPT for lead times up to $L = 7$ for a discrete demand distribution with limited support. For a first performance analysis, we consider all lead times between $L = 2$ and $L = 7$ for six symmetric beta-binomial demand distributions that vary by coefficients of variation (CoV), as shown in Table 1. Symmetric beta-binomial distributions are characterized by only one parameter and cover a range of distributions, including a bell-shaped (CoV$= 0.5$), uniform (CoV ≈ 0.7) and two-point (CoV$= 1$) distribution. The optimal dual sourcing policy and cost C^{OPT}, as well as the optimal local single sourcing policy and cost $C^{\pi^l,k^\star}$ were computed exactly via an linear programming (LP) formulation of the underlying dynamic program, which is a stationary Markov decision problem (MDP). For $L = 7$, the LP has 3,250,000 state-action pairs and took 6.5hrs CPU time. For $L = 2$, only 1,040 state-action pairs are solved in under 1min. The TBS-POUT cost was evaluated by solving its underlying Markov Chain; the reported results thus have no simulation errors. The TBS-POUT order quantity was rounded to integers using the *round half away from zero* convention. The three optimal parameters $\{k^\star, \gamma^\star, i^\star\}$ were identified by a grid search on the integers (α was enumerated in increments of 0.01).

Table 1 shows that the optimality gap reduces in the off-shore lead time L and the coefficient of variation in demand. By construction, TBS-POUT is independent of L, while C^{OPT} is non-decreasing in L because a shorter L dominates a longer L. Hence, the optimality gap is always non-increasing in the lead time. Recall that CoV=0.5 is bell-shaped and closest to a normal distribution.

Table 1 also highlights the optimal single sourcing at a costs

To investigate the parameter robustness of the results in Table 1, our second performance analysis fixed $L = 4$ and the mean of the beta-binomial distribution $\mu = 2$ and calculated the optimality gap for a testbed of 486 parameter combinations: CoV$\in \{0.5, 0.6, 0.7., 0.8, 0.9, 1\}$, $h = 1, b \in$

Coefficient of variation (CoV)		0.5	0.6	0.7	0.8	0.9	1
Demand distribution	$\mathbb{P}[D=0]$	0.0625	0.1206	0.1942	0.2824	0.3844	0.5
	$\mathbb{P}[D=1]$	0.25	0.2375	0.2032	0.1506	0.0823	0
	$\mathbb{P}[D=2]$	0.375	0.2838	0.2052	0.134	0.0666	0
	$\mathbb{P}[D=3]$	0.25	0.2375	0.2032	0.1506	0.0823	0
	$\mathbb{P}[D=4]$	0.0625	0.1206	0.1942	0.2824	0.3844	0.5
Optimal cost C^{OPT}	$L=2$	9.99	10.42	10.98	11.40	11.60	11.60
	$L=3$	10.28	10.82	11.39	11.93	12.26	12.40
	$L=4$	10.40	10.97	11.58	12.10	12.45	12.80
	$L=5$	10.47	11.07	11.69	12.16	12.51	12.83
	$L=6$	10.51	11.12	11.75	12.19	12.55	12.86
	$L=7$	10.53	11.14	11.78	12.21	12.56	12.92
Optimized $C^{\pi^l,k^\star}$		11.13	11.71	12.37	12.58	12.85	13.17
Optimized $C^{\text{TBS-POUT},k^\star}$		10.79	11.42	12.18	12.69	13.29	13.80
Optimality gap $\frac{C^{\text{TBS-POUT},k^\star}}{C^{\mathrm{OPT}}}-1$	$L=2$	7.96%	9.60%	10.97%	11.38%	14.54%	18.97%
	$L=3$	4.93%	5.57%	6.97%	6.45%	8.38%	11.29%
	$L=4$	3.70%	4.11%	5.23%	4.94%	6.69%	7.81%
	$L=5$	2.97%	3.17%	4.22%	4.40%	6.21%	7.55%
	$L=6$	2.58%	2.70%	3.70%	4.10%	5.88%	7.27%
	$L=7$	2.42%	2.51%	3.45%	3.99%	5.77%	6.81%

Table 1 Total costs and optimality gaps for optimized TBS-POUT policies for independent beta-binomial distributed demand with 5 support values and mean $\mu=2$, and parameters $\{h=1, b=9, p=3.8, u=4, m=1.5\}$.

Coefficient of variation (CoV)	0.5	0.6	0.7	0.8	0.9	1
Optimality gap, $\frac{C^{\text{TBS-POUT},k^\star}}{C^{\mathrm{OPT}}}-1$						
Minimum (0%)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
First Quartile (25%)	2.62%	2.63%	2.03%	1.43%	2.27%	2.48%
Median (50%)	4.54%	5.68%	6.77%	6.91%	6.54%	7.04%
Third Quartile (75%)	5.56%	6.99%	8.06%	9.26%	10.49%	12.91%
Maximum (100%)	7.28%	9.01%	12.57%	14.10%	16.34%	18.93%

Table 2 Distribution of 486 optimality gaps for optimized TBS-POUT policies for independent beta-binomial distributed demand with 5 support values and mean $\mu=2$ and $L=4$ over the parameter testbed: $h=1, b\in\{9,19,99\}, \lambda\in\{0.25,0.5,0.75\}, \frac{p}{u}\in\{0.7,0.8,0.9\}$ and $m\in\{1.1,1.5,2\}$.

$\{9,19,99\}, \lambda\in\{0.25,0.5,0.75\}, \frac{p}{u}\in\{0.7,0.8,0.9\}$ and $m\in\{1.1,1.5,2\}$. Table 2 provides confidence that the results in Table 1 are representative for a wide range of parameter values.

Given that the optimality gaps above consider a discrete demand distribution with only 5 support points, our third performance analysis compares TBS-POUT to Program OPT for a discretized normal demand distribution with same mean $\mu=2$ as in Tables 1 and 2. To be able to compute OPT, we fixed minimal $L=2$ (which gives a conservative analysis given that optimality gaps decrease in L) and COV$=0.5$. We discretized the normal distributions using 40 bins (about the maximum number of bins for which we could solve Program OPT) over the truncated support $\mu\pm4\sigma=[-2,6]$. For the parameters of Table 1 with $m=1.1$ (so that a SpeedFactory is viable, as will be discussed in Section 5), we found that the optimality gap of TBS-POUT was 1.94%. We

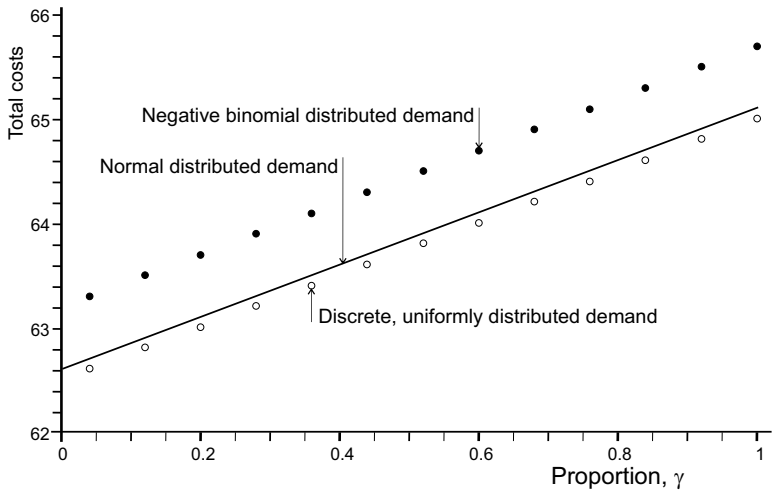


Figure 4 Dual sourcing with a linear TBS-POUT policy when the demand distribution is normal, uniform, or negative binomial; all with same $\mu = 12.5, \sigma = 4.61$. Parameters: $h = 1, b = 9, u = 4, p = 3.8, m = 1.5$. Policy controls for non-normal distributions are optimized numerically. Recall, TBS costs are independent of L.

also calculated the cost of the best TBS-RoI policy, which had an optimality gap of 1.59%. This suggests that the TBS-POUT performance is reasonable, both with respect to OPT and TBS-RoI, and that the optimality gaps in Table 1 for a 5-point discrete beta-binomial distribution are conservative relative to discrete distributions that are closer to a normal distribution.

4.2. Impact of non-normal demand distributions

Our analytic analysis of the linear TBS-POUT policy assumes a normal demand distribution. To investigate the robustness of our analytic insights we compared the performance of TBS-POUT for three distributions that have equal $\mu = 12.5$ and $\sigma = 4.61$ (indeed, our analysis can be viewed as a two-moment approximation of the actual distribution). Figure 4 replicates Figure 3, panel B for a negative binomial demand distribution (also used by Axsäter [2013]) with parameters $p = 0.412$ and $r = 17.9$, a discrete uniform distribution over $[5, 20]$, and a continuous normal distribution $N(12.5, 4.61)$. The costs for the normal demand were obtained analytically. The stationary costs for the negative binomial and discrete uniform demand were obtained by solving their Markov chains and policy parameters were numerically optimized.

We observe that: (1) the curves are linear suggesting that Proposition 3 is robust, (2) higher moments that differ from those of the normal distribution of course change the cost, yet the impact is small ($\leq 1\%$ for the negative binomial demand and $\leq 0.5\%$ for the discrete uniform demand, (3) the optimal safety stock $i^\star$ and SpeedFactory capacity $k^\star$ for the non-normal demands are within one unit away from the normal demand prescription (not shown). The numerically optimized α is $\alpha^\star = 0.51$ for negative binomial demand and discrete uniform demand while the analytic prescription $\alpha^\star = 0.554$ for normal demand. This suggests that our analytic expressions

CoV	$C^{\text{TBS-POUT},k^\star}$	$k^\star$	$\alpha^\star$	q_g	$i^\star$	$\mathbb{P}[q_t<0]$
0.5	10.79	0	0.54	2	2	0.14
0.6	11.42	0	0.62	2	2	0.18
0.7	12.18	0	0.59	2	2	0.22
0.8	12.69	0	0.58	2	3	0.25
0.9	13.29	0	0.55	2	3	0.27
1	13.80	0	0.33	1	3	0.17
CoV	$C^{\text{TBS-POUT}_+,k^\star}$	$k^\star$	$\alpha^\star$	q_g	$i^\star$	$\mathbb{P}[q_t=0]$
0.5	10.97	1	0.50	1	1	0.20
0.6	11.63	1	0.55	2	1	0.19
0.7	12.17	0	0.11	2	1	0.47
0.8	12.38	0	0.11	2	1	0.52
0.9	12.66	0	0.09	2	1	0.55
1	12.99	0	0.13	2	1	0.58

Table 3 (TOP) Probability that $\mathbb{P}[q_t<0]$ for the TBS-POUT policy with the Beta-Binomial demand and parameters of Table 1. (BOTTOM) Performance of truncated TBS-POUT$_+$ policy. Parameters: $h=1, b=9, u=4, p=3.8, m=1.5$. Policy controls numerically optimized.

derived for the normal distribution (Proposition 1) are a reasonable starting point for non-normal distributions.

4.3. Impact of non-negativity constraint on local orders

Our analytic analysis of the linear TBS-POUT policy allows negative local orders. Here we investigate the impact of restricting local orders to be non-negative.

Table 3 compares the optimized $C^{\text{TBS-POUT}}$ to the optimized $C^{\text{TBS-POUT}_+}$ for the beta-binomial demand distribution and parameters used in Table 1. TBS-POUT$_+$ is the modified TBS-POUT policy that truncates negative TBS-POUT orders to zero. The cost difference is between 0.12% to 5.86%. We also compared C^{OPT} to C^{OPT_+}, which depends on the lead time L. For the lead times reported in Table 1, $C^{\text{OPT}_+}/C^{\text{OPT}}-1$ had a median of 0.000%, average of 0.030%, and maximum of 0.225%. This suggest that, for discrete beta-binomial demand, restricting to non-negative orders has minimal impact on the optimal cost and rather moderate impact on the cost under TBS-POUT.

For normal demand, the probability that TBS-POUT places a negative order $\mathbb{P}[q_t<0]=\Phi[-\mu\gamma/\sigma_q]$ is smaller for larger demands (high μ) and more smoothing (small σ_q). For practically non-negative normal demand distributions, the probability of a negative local production order can be made arbitrarily small (i.e., $\mathbb{P}[q_t<0]<\eta$) by increasing the allocation γ so that the SpeedFactory is allocated demand $N(\gamma\mu,\sigma)$ such that the local order q_t is positive with probability greater than $1-\eta$. This accomplishes two feats: First, the SpeedFactory continues to serve the variable demand but also produces on average a small fraction γ of the average demand. Second, its optimal capacity $k^\star=\mu\gamma+\sigma_q z_q$ can be made positive and sufficiently large to recover factory set-up and overhead costs (not considered here) by increasing γ. In other words, by slightly increasing the SpeedFactory size our key insight remains.

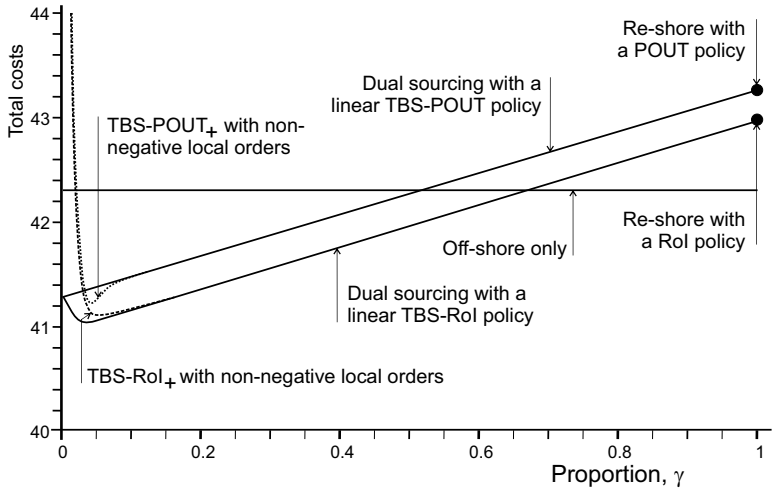


Figure 5 The linear and non-linear TBS-POUT policies track the linear and non-linear TBS-RoI policies as an upper bound. Figure drawn for parameters: $h=1, b=9, u=4, m=1.5, p=3.8, L=6, \sigma=1, \mu=1$ **with the TBS-POUT$_+$ policy parameters** $\{\alpha^\star, k^\star, i^\star\}$ **and the TBS-RoI/TBS-RoI$_+$ policy parameters** $\{k, y_1, y_2\}$ **numerically optimized.**

Figure 5 shows the results of an exact numerical exploration of the impact of non-negative local production orders for the optimized TBS-POUT and TBS-RoI policies[11], replicating the parameter values of Panel B in Figure 3. Here we have discretized the normally distributed demand, $N(10,1)$, into 100 bins over a support of $\mu \pm 4\sigma = [6, 14]$ to get exact results via the dynamic program. We observe: a) TBS-POUT$_+$ and TBS-POUT are both upper bounds on the cost of TBS-RoI and TBS-RoI$_+$. b) Unless $\gamma \to 0$, the costs of the restricted policies TBS-POUT$_+$ and TBS-RoI$_+$ concur with those of the linear policies TBS-POUT and TBS-RoI. c) At $\gamma = 0$ the cost of TBS-POUT is equivalent to the cost of TBS-RoI. d) The costs of the POUT policies track the costs of the RoI policies within 1%. Furthermore, our key insight remains: dual sourcing with a small local SpeedFactory, able to serve the variation in demand, is able to outperform full global outsourcing, even when the local production cost is higher then the global purchase cost.

5. Determinants of SpeedFactory viability under normal iid demand

Having demonstrated its robustness, we now adopt the linear TBS-POUT policy under normal iid demand to investigate when a SpeedFactory outperforms full off-shoring using a break-even analysis.[12] The difference between (18) and (13) shows that a SpeedFactory is economical if

$$C^{\text{TBS-POUT},k^\star} - C^s = (h+b)(\sigma_i - \sigma_{i,s})\phi[z_i] + um\sigma_q\phi[z_q] + \mu\gamma(u-p) < 0, \tag{19}$$

[11] Unfortunately, the solution space is too large for an equivalent MDP analysis of the OPT policies.

[12] Note that an analytical comparison against optimal single local sourcing, or even against optimal dual sourcing with no SpeedFactory ($k=0$) is elusive for normally distributed demands. These optimal policies can be found using dynamic programming, but only for very limited demand support which does not produce fine comparative results.

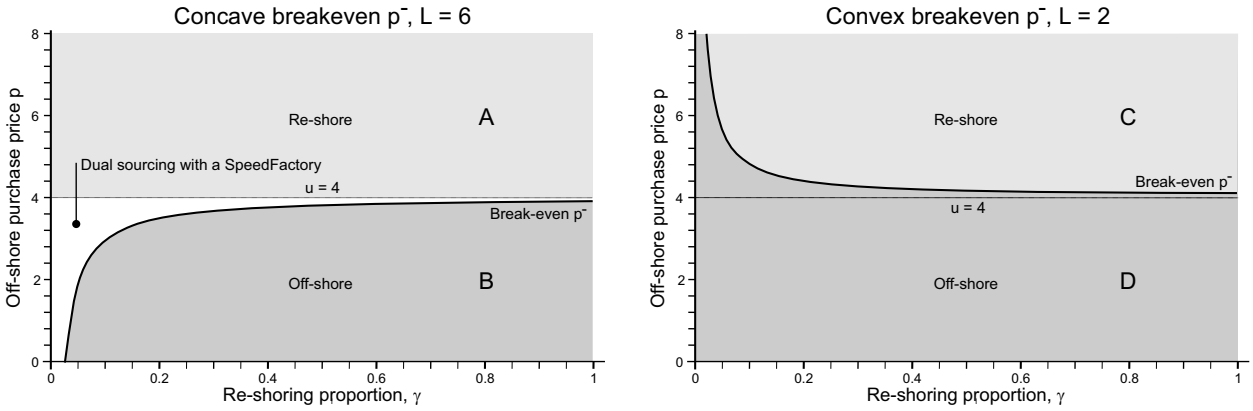


Figure 6 A SpeedFactory controlled with TBS-POUT dominates full offshoring when the off-shore purchasing price p exceeds the break-even price and demand is normal iid. Figure drawn for parameters: $h = 1, b = 9, u = 4, m = 1.5, \sigma = 1, \mu = 10, \alpha^\star = \lambda$.

or, equivalently, when the off-shore purchasing price p exceeds the break-even price p^-:

$$p > p^- = \frac{h+b}{\mu\gamma}(\sigma_i - \sigma_{i,s})\phi[z_i] + \frac{um}{\mu\gamma}\sigma_q\phi[z_q] + u. \tag{20}$$

If $p > p^-$ it is economical to re-shore at least a proportion of demand. The amount to re-shore depends on the concavity of the break-even price. To see this, consider the strategy space plotted in Figure 6 for an off-shore lead-time of $L = 6$ on the left-hand concave plot and $L = 2$ for the right-hand convex plot. A concave p^- implies that the break-even p^- is *below* the regular local unit production cost u. If $p > u$, then complete re-shoring is optimal; this corresponds to panel A in Figure 3. If $p < u$, then panel B in Figure 3 applies and there is a small strategy zone $p^- < p < u$ where dual sourcing with a SpeedFactory serving the variable demand yields lower total cost even though local production costs are higher than the off-shore purchasing price. The reason is the short SpeedFactory lead time allows tight FGI control and results in inventory cost savings that outweigh the increased local production costs. In contrast, a convex p^- implies that the break-even p^- exceeds the regular local unit production cost u. If $p < p^-$, complete off-shoring is optimal, corresponding to panel D in Figure 3. Thus, the off-shore purchased price p has to rise even further before re-shoring is profitable. But when it does become profitable to re-shore, all demand should be re-shored, corresponding to panel C in Figure 3.

The concavity of the break-even price is key to the strategy prescription and depends on the model parameters in a complex matter that can be decomposed into three dimensionless factors:

Proposition 4 (p^- concavity) *Let $d_t = N(\mu, \sigma)$ be identically, normally distributed for each t. When dual sourcing using a linear TBS-POUT policy, the break-even price p^- is concave in γ if*

$$\frac{h+b}{um} \cdot \frac{\sigma_{i,s} - \sigma_i}{\sigma_q} \cdot \frac{\phi[z_i]}{\phi[z_q]} > 1. \tag{21}$$

σ	C^s	γ	$k^\star$	$C^{\text{TBS-POUT}}$	$\frac{C^s}{C^{\text{TBS-POUT}}}-1$
1	42.3	0.2	1.77	41.68	1.74%
2	46.6	0.2	1.54	44.95	3.54%
		0.3	2.54	45.15	3.11%
		0.4	3.54	45.35	2.68%
3	50.9	0.3	2.31	48.43	4.85%
		0.4	3.31	48.63	4.46%
		0.5	4.31	48.83	4.07%
4	55.22	0.4	3.08	51.91	5.96%
		0.5	4.08	52.11	5.60%
		0.6	5.08	52.31	5.24%
		0.7	6.08	52.51	4.87%

Table 4 Economic value of SpeedFactories under TBS-POUT and iid normal demand for different demand variabilities with $\{h=1, b=9, p=3.8, u=4, m=1.5, L=6, \mu=10, \alpha^\star=0.554186\}$.

Then dual sourcing dominates single offshoring when $p^- < p < u$ *(local production costs* exceed *offshore purchasing prices). Otherwise* p^- *is convex and complete re-shoring is optimal when* $p > p^-$.

Condition (21) is the product of three dimensionless factors:

1. The financial factor is the ratio of the inventory costs to local production costs. As inventory costs increase, or local production costs decrease, SpeedFactories are more viable.

2. The variability factor is the ratio of the inventory variability reduction when using TBS-POUT versus using single global sourcing to the SpeedFactory order variability. The smoothing of the POUT policy increases the numerator and decreases the denominator relative to a standard OUT policy, and thus enhances the viability of SpeedFactories. The variability factor also captures the lead time effect via $\sigma_{i,s}$; $\sigma_{i,s}=\sqrt{L}\sigma$ under iid normal demand. Given that σ_i and σ_q are independent of leadtime under a TBS policy, this means that SpeedFactories are more viable as the off-shore leadtime L increases.

3. The service/flexibility factor shows that higher inventory availability targets and higher labor flexibility lead to local SpeedFactories.

Table 4 illustrates the economic benefit of dual sourcing with a SpeedFactory compared to full off-shoring, using (19). When demand variability increases, the performance of dual sourcing with a SpeedFactory improves compared to single sourcing from an off-shore supplier.

Note that (21) is a conservative condition as the (unknown) optimal dual sourcing policy dominates the TBS-POUT policy and thus makes a SpeedFactory even more valuable. The analytic analysis in the remainder of this paper similarly presents a conservative analysis.

6. SpeedFactories are more viable with non-iid demand

Ali et al. [2012] studied a dataset of weekly demand over a 2-year period for 1798 stock keeping units (SKUs) from a European retailer. They identified the first order autoregressive, AR(1),

demand processes in 544 time series (30.3%), the first order integrated moving average, IMA(0,1,1), processes[13,14] in 426 time series (23.7%), and iid demand processes in 293 time series (16.3%). As these were the three most popular ARIMA processes, and accounted for 70% of the retailers SKU's, we explore the economic value of Speedfactories under AR(1) and IMA(0,1,1) demand.

6.1. TBS-POUT and SpeedFactory viability under correlated demand: AR(1)

The autoregressive mean-centered AR(1) demand process is

$$d_t = \rho(d_{t-1} - \mu) + \mu + \epsilon_t. \tag{22}$$

When $|\rho| < 1$, the effect of the initial shock, $\rho^t \epsilon_0$, tends to zero as $t \to \infty$, indicating the demand process is stationary. Let $\hat{d}^{\star}_{t+n,t}$ denote the minimum mean squared error (MMSE) point forecast made at time t of the demand in period $t+n$. For AR(1), Box et al. [2008] show that

$$\hat{d}^{\star}_{t+1,t} = \rho(d_t - \mu) + \mu. \tag{23}$$

The TBS-POUT policy for AR(1) demand issues a *constant* order to the off-shore supplier via (8) and a local SpeedFactory order via

$$q_t|_{\mathrm{TBS,AR}} = \hat{d}^{\star}_{t+1,t} - (1-\gamma)\mu + \alpha(i^{\star} - i_t). \tag{24}$$

The total cost difference ((19)) between complete off-shoring and the dual sourced SpeedFactory under TBS-POUT still holds under AR(1) demand, we only need to calculate the three variances (single-sourcing inventory variance $\sigma^2_{i,s}$, and dual sourcing inventory and production variances σ^2_i and σ^2_q) that arise with AR(1) demand. Using Z-transforms, Online Appendix A shows:

$$\sigma^2_{i,s}|_{\mathrm{AR}} = \frac{\sigma^2}{(1-\rho)^2}\left(\frac{\rho(1-\rho^L)(2+\rho-\rho^{L+1})}{1-\rho^2} + L\right), \tag{25}$$

$$\sigma^2_q|_{\mathrm{AR}} = \sigma^2 \frac{(\alpha^2+\alpha-2)\,\rho - 2\alpha\rho^2 + \alpha + 2\rho^3 - 1}{(\alpha+1)\,(\rho^2-1)\,(1-\alpha\rho)}, \tag{26}$$

$$\sigma^2_{i|\mathrm{AR}} = \frac{\sigma^2}{1-\alpha^2}. \tag{27}$$

When (26) and (27) are used inside (5) we obtain the dual sourcing costs under AR(1) demand using a TBS-POUT policy. It only remains to optimize the smoothing parameter α:

[13] Hanssens [1998] studies the factory orders for a computer peripheral finding them to be IMA(0,1,1) with $\beta = 0.603$.

[14] Cui et al. [2015] finds that 6 out of 14 time series of demand in a consumer packaged goods company were IMA(0,1,1); a further two time series were identified with an IMA(0,1,0) structure.

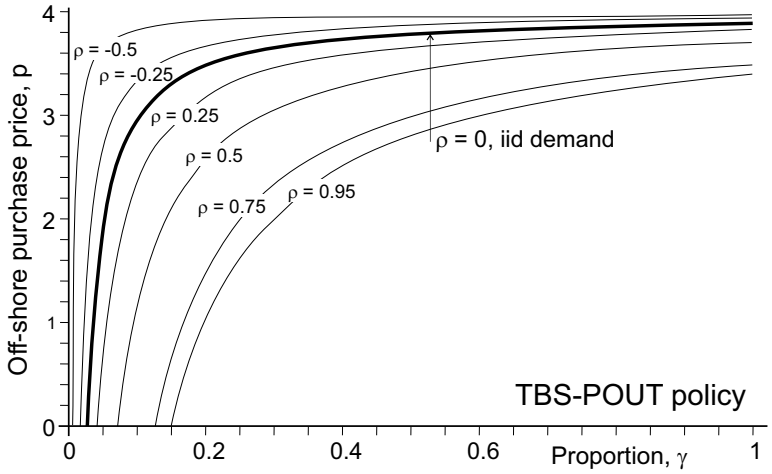


Figure 7 Compared to iid demand ($\rho = 0$, bold line), positively auto-correlated demand significantly increases the viability region where dual sourcing using SpeedFactories under TBS-POUT policies dominates single sourcing. Parameters: $L = 6, h = 1, b = 9, u = 4, m = 1.5, \sigma = 1, \mu = 10$.

Lemma 1 *For TBS-POUT with AR(1) demand, the optimal smoothing* $\alpha^\star = f_{AR}^{-1}[\lambda]$ *where*

$$f_{AR}[\alpha] = \frac{\sigma_{q|AR}}{\sigma_{q|AR} + \sigma_{i|AR} \frac{\left(\alpha^2-1\right)^2\left(\rho\left(2\alpha\rho+\rho-\alpha^2-1\right)-1\right)}{\alpha(\alpha+1)^2(\alpha\rho-1)^2}}. \ \square \tag{28}$$

Lemma 1 specifies $\alpha^\star$ in terms of the inverse function and provides a direct graphical solution: put λ on the vertical axis and find corresponding $\alpha^\star = f_{\mathrm{AR}}^{-1}[\lambda]$ on the horizontal axis. Alternatively, $f_{\mathrm{AR}}[\alpha^\star] = \lambda$ is easily solved numerically for $\alpha^\star \in (-1, 1]$ given $\lambda \in [0, 1]$ and $\rho \in (-1, 1)$, as shown in Online Appendix B, which also presents analytic bounds on $\alpha^\star$.

Using (25)-(27) inside (19), numerically optimizing $\alpha^\star$ using Lemma 1 provides the break-even price, p^-, for which dual sourcing with a SpeedFactory and TBS-POUT policy dominates single sourcing from offshore supplier, see Fig. 7. As demand becomes more positively correlated, the break-even curves fall and the strategy space where a SpeedFactory dominates (i.e., the area between the concave breakeven curve and $p = u = 4$) grows significantly: if $\rho = 0.95$, the strategy space is about four times the size of the iid strategy space. (Recall that this region is conservative and a lower bound as the optimal dual sourcing policy outperforms TBS-POUT.) Furthermore, for strongly negative auto-correlation $\rho < \overline{\rho} = -0.6180$, full off-shoring is likely to become dominant. Note, as the break-even curves have the same structure under AR(1) demand as under iid demand, Proposition 2 and Proposition 3 continue to hold.

Not only do SpeedFactories dominate as autocorrelation increases, their value relative to full off-shoring also increases, as shown in Table 5. Depending on how much of the demand $0 < \gamma < 1$ is produced locally, the SpeedFactory solution exhibits better performance than full off-shoring as the demand becomes more correlated.

ρ	C^s	$C^{\text{TBS-POUT}}$	$\frac{C^s}{C^{\text{TBS-POUT}}}-1$ improvement range
-0.5	41.09	$40.93+2\gamma$	-4.5% to 0.4%
-0.25	41.57	$40.94+2\gamma$	-3.3% to 1.52%
0	42.30	$41.28+2\gamma$	-2.3% to 2.42%
0.25	43.44	$41.81+2\gamma$	-0.9% to 3.74%
0.5	45.33	$42.53+2\gamma$	1.76% to 6.17%
0.75	48.68	$43.68+2\gamma$	6.15% to 10.26%
0.95	53.22	$47.28+2\gamma$	7.41% to 11.17%

Table 5 Comparison of dual sourcing with TBS-POUT against full off-shoring under AR(1) demand. Note, the % cost improvement range is determined by setting $\gamma=0$ for the upper bound and $\gamma=1$ for the lower bound. Parameters: $p=3.8, L=6, h=1, b=9, u=4, m=1.5, \sigma=1, \mu=10$.

6.2. SpeedFactory viability under non-stationary demand: IMA(0,1,1)

Dekimpe and Hanssens [1995] discuss the short- and long-term effects of sales promotions. If a sales promotion causes customers to switch brands only temporarily, the increased demand has a stationary effect; i.e., the demand level soon reverts back to fluctuating around its pre-promotion mean. However, if the sales promotion causes a permanent change in the customer's brand choice, demand becomes non-stationary. Dekimpe and Hanssens [1995] found the majority of 400 time series (54%) to be non-stationary. Interestingly, Dekimpe and Hanssens [1995] find 29% of European sales time series were non-stationary (matching the European results from Ali et al. [2012]), but that 72% of US and Canadian sales were non-stationary. The simplest non-stationary demand process (and the most common in the study of EU retail demand by Ali et al. [2012]) is the Integrative Moving Average process of first order, IMA(0,1,1):

$$d_t = \begin{cases} \mu+\epsilon_0, & \text{if } t=0, \\ d_{t-1}-(1-\beta)\epsilon_{t-1}+\epsilon_t, & \text{otherwise.} \end{cases} \tag{29}$$

Here, μ is an initial condition that does not have any long-term impact. For invertability reasons (so that β can be estimated from past data), the moving average parameter β is restricted to $\beta\in[0,2)$. Graves [1999] showed that exponential smoothing,

$$\hat{d}^{\star}_{t+1,t} = \begin{cases} \mu, & \text{if } t<0, \\ \beta d_t+(1-\beta)\hat{d}^{\star}_{t,t-1}, & \text{otherwise,} \end{cases} \tag{30}$$

provides the one-period-ahead minimum mean squared error point forecast of the IMA(0,1,1) process because $d_{t+1}-\hat{d}^{\star}_{t+1,t}=\epsilon_{t+1}$ and it can be shown that $\forall i>1$, $\hat{d}^{\star}_{t+i,t}=\hat{d}^{\star}_{t+1,t}$ as $\forall i>1$, $\mathbb{E}[\epsilon_{t+i,t}]=0$.

Constant off-shore orders under non-stationary IMA(0,1,1) demand result in non-stationary local orders with infinite variance, which leads to infinite local production costs. Therefore, with IMA(0,1,1) demand, we set the global off-shore order to the expected demand in the period after the off-shore lead time minus a strategic fraction $\gamma\in[0,1]$ of the initial demand that is produced in the local factory so that the off-shore supplier gets a *dynamic* order:

$$q^g_t|_{\text{DYN}} = \hat{d}^{\star}_{t+L,t}-\mu\gamma. \tag{31}$$

This results in a stationary demand at the local SpeedFactory with finite order variance. The POUT policy that specifies local orders to control the local inventory is also generalized to

$$q_t|_{\mathrm{DYN}} = \hat{d}^{\star}_{t+1,t} - (\hat{d}^{\star}_{t+1,t-L+1} - \mu\gamma) + (1-\alpha)\,(i^{\star} - i_t)\,. \tag{32}$$

Notice, $\hat{d}^{\star}_{t+1,t}$ and $\hat{d}^{\star}_{t+1,t-L+1}$ are two different point forecasts, produced at different times, but both predicting the demand in period $t+1$. Both are dependent upon the demand process; with iid demand, both forecasts coincide, but with general time series they may differ. Optimization over α ensures the POUT performance is never worse than the OUT performance (where $\alpha = 0$).

Using Z-transforms, Online Appendix A derives the single-sourcing inventory variance $\sigma^2_{i,s}$, and dual sourcing inventory and production variances σ^2_i and σ^2_q resulting from the DYN-POUT policy $\pi = \{(q^g_t|_{\mathrm{DYN}}, q_t|_{\mathrm{DYN}}) \in \mathbb{R}^2 : t \in \mathbb{N}_0\}$ under IMA(0,1,1) demand:

$$\sigma^2_{i,s}|_{\mathrm{IMA}} = \sigma^2 L\big(1 + \beta(L-1) + \beta^2(L-1)(2L-1)/6\big), \tag{33}$$

$$\sigma^2_q|_{\mathrm{IMA}} = \sigma^2\left(\frac{1-\alpha}{1+\alpha} + 2\beta(1-\alpha^{L-1}) + \beta^2(L-1)\right) \text{ and } \sigma^2_i|_{\mathrm{IMA}} = \frac{\sigma^2}{1-\alpha^2}. \tag{34}$$

With the off-shore supplier absorbing the non-stationary demand variation, both the local replenishment orders and the FGI levels are stationary under IMA demand. This leads to an optimal FGI safety stock of $i^{\star} = \sigma_i z_i$ and an optimal installed SpeedFactory capacity base of $k^{\star} = \mu\gamma + \sigma_q z_q$. At the same time, the off-shore supplier's production will now always be less variable than the demand, indicating dual sourcing may be advantageous for suppliers who are willing to sacrifice a small amount of total volume to gain access to a smoother demand.

To address the non-stationarity, we write the total costs under complete off-shoring as

$$C^s = C^{\star}_{i,s} + p \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d_t\right], \tag{35}$$

and the total costs under dual sourcing using the DYN-POUT policy as,

$$C^{\mathrm{DYN},k^{\star}} = C^{\star}_i + C^{\star}_q + p \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} (\hat{d}^{\star}_{t+n.t} - \mu\gamma)\right], \tag{36}$$

where $\{C^{\star}_i, C^{\star}_q\}$ is as defined in (10) and (11), respectively[15]. As $\mathbb{E}[\hat{d}^{\star}_{t+n,t}] = \mathbb{E}[d_t]$, the difference between (35) and (36) can be re-arranged to find the break-even purchase price, p^{-}, indicating when dual sourcing with a SpeedFactory and DYN-POUT policy may be preferred over full off-shoring. The result is (20) after the variances are suitably updated using (33) and (34). It remains to optimize the smoothing parameter of the DYN-POUT policy under IMA demand:

[15] Note, the inventory costs are independent of $\mu\gamma$.

β	C^s	$C^{\text{DYN-POUT}}$	Absolute Improvement Lower Bound	Relative Improvement [Lower, Upper] Bound
0	$4.30+C^\star_{p,s}$	$3.28+2\gamma+C^\star_{p,d}$	$1.02-2\gamma$	$[0\%, (1.02-2\gamma)/(4.30) < 24\%]$
0.25	$7.22+C^\star_{p,s}$	$4.45+2\gamma+C^\star_{p,d}$	$2.78-2\gamma$	$[0\%, (2.78-2\gamma)/(7.22) < 38\%]$
0.5	$10.35+C^\star_{p,s}$	$5.48+2\gamma+C^\star_{p,d}$	$4.87-2\gamma$	$[0\%, (4.87-2\gamma)/(10.34) < 47\%]$
0.75	$13.53+C^\star_{p,s}$	$6.64+2\gamma+C^\star_{p,d}$	$6.89-2\gamma$	$[0\%, (6.89-2\gamma)/(13.53) < 51\%]$
1.0	$16.74+C^\star_{p,s}$	$7.83+2\gamma+C^\star_{p,d}$	$8.92-2\gamma$	$[0\%, (8.92-2\gamma)/(16.74) < 53\%]$
1.25	$19.97+C^\star_{p,s}$	$9.02+2\gamma+C^\star_{p,d}$	$10.95-2\gamma$	$[0\%, (10.95-2\gamma)/(19.97) < 56\%]$
1.5	$23.20+C^\star_{p,s}$	$10.22+2\gamma+C^\star_{p,d}$	$12.98-2\gamma$	$[0\%, (12.98-2\gamma)/(23.20) < 57\%]$
1.75	$26.44+C^\star_{p,s}$	$11.42+2\gamma+C^\star_{p,d}$	$15.02-2\gamma$	$[0\%, (15.02-2\gamma)/(26.44) < 57\%]$
1.95	$29.03+C^\star_{p,s}$	$12.38+2\gamma+C^\star_{p,d}$	$16.65-2\gamma$	$[0\%, (16.65-2\gamma)/(29.03) < 57\%]$

Table 6 Comparison of dual sourcing with DYN-POUT against full off-shoring under IMA(0,1,1) demand.
Parameters: $L=6, p=3.8, h=1, b=9, u=4, m=1.5, \sigma=1, \mu=10$.

Lemma 2 *For DYN-POUT with IMA(0,1,1) demand, the optimal smoothing* $\alpha^\star = f^{-1}_{IMA}[\lambda]$ *where*

$$f_{IMA}[\alpha] = \frac{\sigma_{q|IMA}}{\sigma_{q|IMA} - \sigma_{i|IMA}\left((\alpha^2-1)^2\beta(1-L)\alpha^{L-3} - \frac{(\alpha-1)^2}{\alpha}\right)}. \ \square \tag{37}$$

As with AR(1), the Lemma specifies $\alpha^\star$ in terms of the inverse function and provides a direct graphical solution. Alternatively, $f_{\text{IMA}}[\alpha^\star] = \lambda$ is easily solved numerically for $\alpha^\star \in (-1, 1]$ given $\lambda \in [0, 1]$ and $\beta \in [0, 2)$, as shown in Online Appendix B, which also presents analytic bounds.

After numerically optimizing α, the break-even purchase price (20), p^-, for which dual sourcing with a SpeedFactory and DYN-POUT policy dominates single sourcing from the offshore supplier, can be plotted, see Fig. 8. As demand becomes more non-stationary the break-even curves fall and the strategy space where a SpeedFactory dominates (i.e., the area between the concave breakeven curve and $p = u = 4$) grows significantly. The impact of non-stationary demand is even more significant than that of serially-correlated demand: when $\beta = 1.95$, the SpeedFactory strategy space is approximately *six times* larger than its iid (i.e., $\beta = 0$) counterpart. (Recall that this region is conservative and a lower bound as the optimal dual sourcing policy outperforms DYN-POUT.)

Not only do SpeedFactory solutions become more viable as non-stationarity (i.e., β) increases, their value relative to full off-shoring also increases. Table 6 presents the total costs as a function of β, λ, and the purchasing costs $C^\star_{p,s}$ and $C^\star_{p,d}$ under IMA demand. In this numerical setting, SpeedFactories are more attractive than full off-shoring, and increasingly so as the demand becomes more non-stationary. Note, the IMA(0,1,1) demand has no natural mean and $\{C^\star_{p,d}, C^\star_{p,s}\}$ are indeterminate. So, although $C^\star_{p,d} < C^\star_{p,s}$, it is not possible to determine the range of the possible cost improvement under IMA(0,1,1) in the same way we did for AR(1) demand. However, we are able to give a lower bound of the absolute difference between full off-shoring and the SpeedFactory solution (note, $0 < \gamma < 1$).

7. SpeedFactory decisions in a European apparel company

A European apparel company currently sources casual wear (T-shirts, polos, sweaters, etc.) from Bangladesh. The production and transportation lead-time is 5 months, which together with the

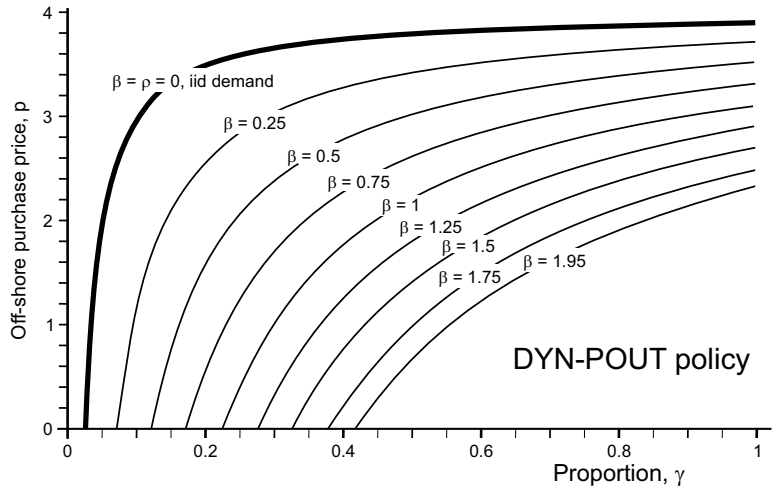


Figure 8 Compared to iid demand ($\beta = 0$, bold line), more non-stationary demand (β increases) significantly increases the viability region where dual sourcing using SpeedFactories under DYN-POUT policies dominates single sourcing. Parameters: $L = 6, h = 1, b = 9, u = 4, m = 1.5, \sigma = 1, \mu = 10$.

review cycle gives a risk period of $L = 6$. We ran their monthly demand data (from January 2013 to June 2018) through the *arima.auto* function in the R software package so as to identify the demand process structure. For illustrative purposes we report on one T-shirt that followed an IMA(0,1,1) demand pattern, with an average demand of 845, a standard deviation of the error term of 514, and $\beta = 1$. T-shirts produced in Bangladesh and imported to the EU had a declared average value of $p =$€1.67 during the 2013-2017 period (`www.ec.europa.eu/eurostat`).

We assumed a life-time of a printable T-shirt of one year, corresponding to a inventory holding cost of $h = p/12$. Target availability levels were set to be 98% as per company policy, implying a backlog cost of $b = 49h$. This resulted in an expected monthly purchasing cost of €1411.15, and inventory cost of €1653.43, providing a total off-shore cost per month of €3064.58. With the same demand, lead-time, and cost information, we quantified whether a SpeedFactory would be attractive. We assumed the European labor flexibility coefficient[16] $m = 1.5$.

Proposition 4 revealed that a SpeedFactory is only viable when the breakeven price p^- is concave. Rewriting Eq. (21), we find that p^- is concave in γ if the local unit production cost u satisfies:

$$u < u^c = \frac{h+b}{m} \cdot \frac{\sigma_{i,s} - \sigma_i}{\sigma_q} \cdot \frac{\phi[z_i]}{\phi[z_q]} = €1.92, \tag{38}$$

otherwise p^- is convex in γ. Re-arranging the p^- (eq.(20)) for u, reveals that the SpeedFactory will become economic for $\gamma = 0.2$ when the local unit production cost $u < u^- = €1.87$ with $\alpha^\star = 0.616$. The cheapest purchase cost (averaged over the 2013-2017 period) for a European-based SpeedFactory is €2.60 in Croatia (`www.ec.europa.eu/eurostat`). Note, this is not a unit production cost,

[16] As labor costs are typically less than 50% of the manufacturing cost of a T-shirt the true m will be smaller than $m = 1.5$. As a consequence our SpeedFactory predictions are likely to be conservative.

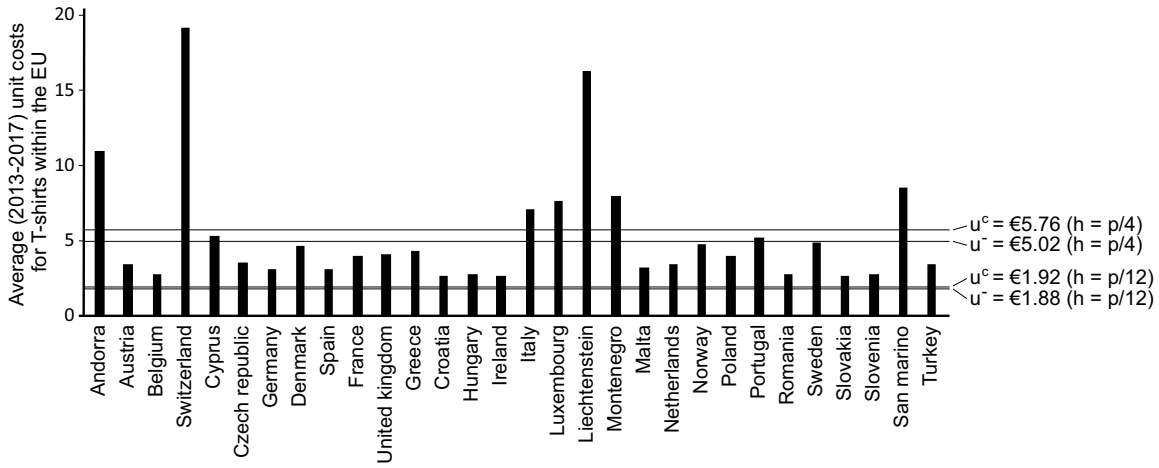


Figure 9 Average unit production costs for T-shirt production within the EU countries during (2013-2017). Based on information available at www.ec.europa.eu/eurostat/.

to which we have no access. However, our SpeedFactory analysis reveals the benchmark unit price for local production to be viable is not the current off-shore price €1.67, but rather €1.92 because of the benefit of tight inventory control.

If we were to assume a life-time of the printable T-shirt of just four months, rather than a year, the inventory holding cost becomes $h = p/4$. With the same lead-time, service levels, and demand process, the total monthly off-shore cost becomes €6371.45. The break-even purchase price is concave in γ if $u < u^c = €5.76$ and the break-even unit production cost with $\gamma = 0.2$ becomes $u^- = €5.02$ (with $\alpha^\star = 0.503$). Based on average prices in the 2013-2017 period, see Figure 9, the following countries would be suitable for SpeedFactory production: Belgium, Austria, Czech Republic, Germany, Denmark, Spain, France, UK, Greece, Croatia, Hungary, Ireland, Malta, Netherlands, Norway, Poland, Romania, Sweden, Slovakia, Slovenia, and Turkey (`www.ec.europa.eu/eurostat`). This demonstrates our analysis is able to determine whether (and where) a SpeedFactory is attractive given the specific product characteristics and country specific factors such as labor costs, over-time rates, and lead-times.

8. Concluding remarks

As off-shore costs rise, it is not necessary to wait until off-shore costs exceed local costs to bring back production; we can re-shore economically before that occurs with a SpeedFactory. A SpeedFactory, used in combination with an off-shore supplier, should be as small as possible, but able to serve the local order variation. In other words, its scale is of the order of the standard deviation of the local orders. The small, local SpeedFactory can be valuable even when local unit production cost is higher than the off-shore purchase price due to inventory savings accruing from the fast lead time. These savings increase when the SpeedFactory lead time decreases relative to the off-shore

lead-time and when the SpeedFactory labor is more flexible. Rigid local labor works against the benefits of re-shoring.

Re-shoring is strongly influenced by the demand structure; positively autocorrelated and non-stationary demand significantly increases the economic benefit of SpeedFactories. This implies that the upstream supply chain needs to be aligned to market factors as the majority of retail demand is correlated or non-stationary. The *small is beautiful* SpeedFactory also reduces demand variability faced by the off-shore supplier. While the off-shore suppliers demand volume has been reduced, the decreased variability reduces any capacity idling/over-time required at the off-shore supplier, preserving the off-shore business case.

We considered two different types of orders issued to the off-shore supplier, the TBS-type with constant off-shore orders and the dynamic off-shore orders based on forecasted demand. Dynamic orders were the only feasible way to manage non-stationary demand. Our results are likely to be conservative as we have not accounted for the value of late customization or location, nor for customization or multi-product flexibility. There is also a reputation benefit from having a local SpeedFactory; people are often willing to pay for a made-within-their-own-country label. Furthermore, the structure of the SpeedFactory decision shows that limited editions/very short shelf-life products benefit particularly strongly from SpeedFactory production. We emphasize that identifying the best-fitting time series is the first step in our approach. Once the time-series is identified, we have presented a causal framework to solve the capacity and production and sourcing process of the SpeedFactory.

Our work could also provide insights into a range of scenarios where there are two lead-times. For example, in supply chains where there are two transportation modes (for example, air/ship, or road/rail). The interesting nuance in this problem is that often the longest lead-time has the least environmental impact and the emphasis is on shifting product to the slow route, lengthening lead-times, rather than using the more environmentally damaging fast route. Re-manufacturing applications could also exist as the production of new items have one lead-time, and the re-manufacture of product returns has another lead time.

References

Ali, M. M., J. E. Boylan, A. A. Syntetos. 2012. Forecast errors and inventory performance under forecast information sharing. *International Journal of Forecasting* **28** 830–841.

Allon, G., J. A. Van Mieghem. 2010. Global dual sourcing: Tailored base-surge allocation to near- and offshore production. *Management Science* **56**(1) 110–124.

Aviv, Y. 2007. On the benefits of collaborative forecasting partnerships between retailers and manufacturers. *Management Science* **55**(5) 777–794.

Axsäter, S. 2013. When is it feasible to model low discrete demand by a normal distribution? *OR Spectrum* **35** 153–162.

Bain, M. 2017. Nike's plan to shave a month off its shoe delivery times. https://qz.com/1112641/nike-is-racing-adidas-to-speed-up-sneaker-manufacturing/.

Bissell, C. 1992a. Pioneers of control: An interview with Arnold Tustin. *IEE Review* **June** 223–226.

Bissell, C. 1992b. Yakov Tyspkin: A Russian life in control. *IEE Review* **Sept** 313–316.

Boute, R. N., J. A. Van Mieghem. 2015. Global dual sourcing and order smoothing: The impact of capacity and lead times. *Management Science* **61**(9) 2080–2099.

Box, G.E.P., G.M. Jenkins, G.C. Reinsel. 2008. *Time Series Analysis, Forecasting, and Control*. Holden-Day, San Francisco.

Cui, R., G. Allon, A. Bassamboo, J.A. Van Mieghem. 2015. Information sharing in supply chains: An empirical and theoretical valuation. *Management Science* **61**(11) 2803–2824.

Dekimpe, M. G., D. M. Hanssens. 1995. Empirical generalizations about market evolution and stationarity. *Marketing Science* **14**(3) 109–121.

Deziel, D. P., S. Eilon. 1967. A linear production-inventory control rule. *Production Engineer* **46**(2) 93–104.

Disney, S. M., M. R. Lambrecht. 2008. On replenishment rules, forecasting and the bullwhip effect in supply chains. *Foundations and Trends in Technology, Information and Operations Management* **2**(1) 1–80.

Disney, S. M., A. Maltz, X. Wang, R. D. H. Warburton. 2016. Inventory management for stochastic lead times with order crossovers. *European Journal of Operational Research* **248** 473–486.

Ellram, L.M., W.L. Tate, K.J. Petersen. 2013. Offshoring and reshoring: an update on the manufacturing location decision. *Journal of Supply Chain Management* **49**(2) 14–22.

Gaalman, G. 2006. Bullwhip reduction for ARMA demand: The proportional order-up-to policy verses the full-state-feedback policy. *Automatica* **42** 1283–1290.

Graves, S. C. 1999. A single-item inventory model for a non-stationary demand process. *Manufacturing & Service Operations Management* **1**(1) 50–61.

Hanssens, D. M. 1998. Order forecasts, retail sales, and the marketing mix for consumer durables. *Journal of Forecasting* **17** 327–346.

Heath, D.C., P.L. Jackson. 1994. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Transactions* **26**(3) 17 – 30.

Hosoda, T., S.M. Disney. 2018. A unified theory of the dynamics of closed-loop supply chains. *European Journal of Operational Research* **269**(1) 313 – 326.

Hua, Z., Y. Yu, W. Zhang, X. Xu. 2015. Structural properties of the optimal policy for dual-sourcing systems with general lead times. *IIE Transactions* **47**(8) 841–850.

Huh, W.T., M. Nagarajan. 2010. Technical note - Linear inflation rules for the random yield problem: Analysis and computations. *Operations Research* **58**(1) 244–251.

Janakiraman, G., S. Seshadri, A. Sheopuri. 2015. Analysis of tailored base-surge policies in dual sourcing inventory systems. *Management Science* **61**(7) 1547–1561.

Kinkel, S., S. Maloca. 2009. Drivers and antecedents of manufacturing offshoring and backshoring: A german perspective. *Journal of Purchasing & Supply Management* **15** 154–165.

Margolis, J. 2017. Made in England: why some manufacturing is coming home from China. https://www.ft.com/content/141418ba-88b1-11e7-afd2-74b8ecd34d3b.

Nise, N. S. 2004. *Control Systems Engineering*. John Wiley & Sons, New Jersey.

Rosenshine, M., D. Obee. 1976. Analysis of a standing order inventory system with emergency orders. *Operations Research* **24**(6) 1143–1155.

Shotter, J., L. Whipp. 2016. Robot revolution helps Adidas bring shoemaking back to Germany. https://www.ft.com/content/7eaffc5a-289c-11e6-8b18-91555f2f4fde.

Song, J.-S., L. Xiao, H. Zhang, P. Zipkin. 2017. Optimal policies for a dual-sourcing inventory problem with endogenous stochastic lead times. *Operations Research* **65**(2) 379–395.

Song, J.-S., P. Zipkin. 2009. Inventories with multiple supply sources and networks of queues with overflow bypasses. *Management Science* **55**(3) 362–372.

Tsypkin, Y. Z. 1964. *Sampling Systems Theory and its Application*, vol. 2. Pergamon Press, Oxford.

Whittemore, A.S., S. C. Saunders. 1977. Optimal inventory under stochastic demand with two supply options. *SIAM Journal on Applied Mathematics* **32**(2) 293–305.

Wikipedia. 2018. $z-$transform. https://en.wikipedia.org/wiki/Z-transform.

Xin, L., D.A. Goldberg. 2018. Asymptotic optimality of tailored base-surge policies in dual-sourcing inventory systems. *Management Science* **64**(1) 437–452.

Yao, M., S. Minner. 2017. Review of multi-supplier inventory models in supply chain management: An update. *Available at SSRN: https://ssrn.com/abstract=2995134* **June 29**.

Ziady, Hanna. 2019. Adidas is closing hi-tech sneaker factories in germany and the us. https://edition.cnn.com/2019/11/12/business/adidas-speedfactory-plants-closing/index.html.

Zipkin, P. H. 2000. *Foundations of Inventory Management*. McGraw-Hill Companies.

Appendix A. Proofs

Proof of Proposition 1 The safety stock and the inventory costs will be derived. We depart from the expected, per period, inventory holding and backlog costs,

$$\mathbb{E}[C_t^i] = b\int_{-\infty}^{0}(-x)f[x]dx + h\int_{0}^{\infty}xf[x]dx, \tag{39}$$

where $f[x]$ is the pdf of the inventory levels. Using a linear TBS-POUT policy with iid normal demand, the inventory levels are normally distributed with mean μ_i and standard deviation σ_i, $f[x] = \phi[x|\mu_i, \sigma_i]$. Closing the integral and using the pdf and cdf of the standard normal distribution, $\{\phi[x], \Phi[x]\}$, leads to,

$$\mathbb{E}[C_t^i] = (b+h)\left(\mu_i \Phi\left[\frac{\mu_i}{\sigma_i}\right] + \sigma_i \phi\left[\frac{\mu_i}{\sigma_i}\right]\right) - b\mu_i. \tag{40}$$

We wish to control the mean inventory levels to minimise the expected inventory costs. The derivative of $\mathbb{E}[C_t^i]$ w.r.t. μ_i is

$$\frac{d\mathbb{E}[C_t^i]}{d\mu_i} = (b+h)\Phi\left[\frac{\mu_i}{\sigma_i}\right] - b. \tag{41}$$

Setting the derivative to zero and solving for μ_i reveals the optimal safety stock level, $i^\star = \mu_i$,

$$i^\star = \sigma_i z_i; \quad z_i = \Phi[b/(b+h)]. \tag{42}$$

Using (42) in (40) and simplifying leads to the following expression for the minimised costs,

$$C_i^\star = \sigma_i (h+b)\phi[z_i]. \tag{43}$$

The optimal capacity level $k^\star$ and the minimised capacity costs are derived as follows. The expected capacity costs are

$$\mathbb{E}[C_t^q] = uk + \mathbb{E}[um[q_t - k]^+] = uk + um\int_k^\infty (x-k)f[x])dx, \tag{44}$$

where $f[x]$ is the pdf of the local orders. For normally distributed orders with mean $\mu\gamma$ and standard deviation σ_q, $f[x] = \phi[x|\mu\gamma, \sigma_q]$; using this and simplifying (44), leads to,

$$\mathbb{E}[C_t^q] = uk + m\left((k-\mu\gamma)\Phi\left[\frac{k-\mu\gamma}{\sigma_q}\right] + \sigma_q\phi\left[\frac{k-\mu\gamma}{\sigma_q}\right] + \mu\gamma - k\right). \tag{45}$$

The derivative of $\mathbb{E}[C_t^q]$ w.r.t. k is

$$\frac{d\mathbb{E}[C_t^q]}{dk} = u\left(1 + m\left(\Phi\left[\frac{k-\mu\gamma}{\sigma_q}\right] - 1\right)\right). \tag{46}$$

Setting the derivative to zero and solving for k reveals the optimal capacity level,

$$k^\star = \mu\gamma + \sigma_q z_q; \quad z_q = \Phi^{-1}[(m-1)/m]. \tag{47}$$

Substituting (47) into (45) and simplifying leads to the minimised local capacity costs,

$$C_q^\star = u\mu\gamma + um\sigma_q\phi[z_q]. \tag{48}$$

Proof of Proposition 2 The derivative of (15) with respect to α is

$$\frac{dC^{\pi,k}}{d\alpha} = \varphi\left(\frac{(1-\lambda)\frac{d\sigma_i^2}{d\alpha}}{2\sigma_i} + \frac{\lambda\frac{d\sigma_q^2}{d\alpha}}{2\sigma_q}\right). \tag{49}$$

From (49), the necessary first order optimality condition for $\alpha^\star$ is

$$\frac{(\lambda-1)\frac{d\sigma_i^2}{d\alpha}}{2\sigma_i} = \frac{\lambda\frac{d\sigma_q^2}{d\alpha}}{2\sigma_q}. \tag{50}$$

Rearranging (50) for λ provides, (17), the stated inverse function for α. □

Proof of Proposition 4 Expression (21) is readily obtained from the second derivative of p^-. □

Appendix B. Methodology: Using *Z*-transforms to calculate variances

To study the significant impact of correlated and non-stationary demand series, we adopt Z-transforms[17]. These transforms are the discrete-time analogue of Laplace and Fourier transforms and are at the foundation of linear control theory. This section presents the essence of this methodology.

Linear discrete-time systems are governed by linear difference equations. System analysis requires the solution of these equations for a given input function. This solution can be characterized in the time-domain or, using the Z-transform, in the frequency domain. The benefit of the latter is that the difference equation is transformed into an easier-to-solve polynomial equation[18]. The Z-transform of a discrete time series x_t is defined as[19]

$$X[z] = \mathscr{Z}[x_t] = \sum_{t=0}^{\infty} x_t z^{-t}. \tag{51}$$

Linearity implies that the system output y_t for any input function x_t is fully described by the impulse response function g_t, which is the solution of the system's difference equation when the input is the unit impulse function (also called the Kronecker delta function) $\delta[t]$; $\delta[t=0]=1$ and $\delta[t\neq 0]=0$. Indeed, given that any input function is the superposition of scaled and delayed

[17] $Z-$transforms were developed independently during the second world war for military needs: In the UK for gun target systems (Tustin), Bissell [1992a], the US (Ragazzini and Zadeh) for radar and also in Russia (Tsypkin), Bissell [1992b]. The z name originated from the US team, Wikipedia [2018].

[18] We could have used a range of methods to study our problem: the Martingale Method of Forecast Evolution [Heath and Jackson 1994], state space methods [Gaalman 2006], the expectation operator [Hosoda and Disney 2018], the backwards difference operator, [Box et al. 2008], to name just a few. However, we choose the z-transform method due to its well-established tool-kit for understanding complex linear systems. Regardless of method used, the end expressions are equivalent.

[19] Notation: As is usual practice in the control literature we use lower case letters for state variables in the time domain and UPPER case letters for the corresponding state variables in the complex frequency (z) domain.

impulse functions, its output is similarly the superposition of scaled and delayed impulse response functions. Equivalently, the output is the convolution of the input and impulse response:

$$y_t = \sum_{i=0}^{t} x_i g_{t-i}. \tag{52}$$

The Z-transform of the impulse response function is called the transfer function $G(z)$ and convolutions are transformed into simple multiplications so that the $Z-$transform of (52) yields

$$Y(z) = G(z)X(z). \tag{53}$$

The impulse response function directly allows the exact computation of the variance of the output:

Lemma 3 (Tsypkin [1964]) *If the input x_t to a linear system with impulse response function g_t is an iid random process with variance σ_x^2, then the long-run variance of the output y_t is*

$$\sigma_y^2 = \sigma_x^2 \sum_{t=0}^{\infty} (g_t)^2. \tag{54}$$

Proof: A sketch of the proof removes potential mystery. Denote $\mathbb{E}[x_t] = \bar{x}$ and $\lim_{t\to\infty} \mathbb{E}[y_t] = \bar{y}$. Taking expectations and limits in (52) yields $\bar{y} = \bar{x} \sum_{t=0}^{\infty} g_t$. Indeed, linearity means that a centered input $x_t - \bar{x}$ yields asymptotically centered output $y_t - \bar{y}$. Similarly:

$$\begin{aligned}
\sigma_y^2 &= \lim_{t\to\infty} \mathbb{E}[(y_t - \bar{y})^2] && \text{(by definition of a variance)} \\
&= \lim_{t\to\infty} \mathbb{E}\left[\left(\sum_{i=0}^{t} (x_i - \bar{x}) g_{t-i}\right)^2\right] && \text{(using (52))} \\
&= \lim_{t\to\infty} \mathbb{E}\left[\left(\sum_{i=0}^{t} (x_i - \bar{x}) g_{t-i}\right)\left(\sum_{j=0}^{t} (x_j - \bar{x}) g_{t-j}\right)\right] && \text{(expand the square)} \\
&= \lim_{t\to\infty} \sum_{i=0}^{t}\sum_{j=0}^{t} \mathbb{E}[(x_i - \bar{x})(x_j - \bar{x})] g_{t-i} g_{t-j} \\
& && \text{(expected value of a sum is the sum of its expected addends)} \\
&= \sigma_x^2 \sum_{i=0}^{\infty} g_i^2. && (\mathbb{E}[(x_i - \bar{x})(x_j - \bar{x})] = 0 \text{ if } i \neq j \text{ for iid input } x_t)
\end{aligned}$$

□

Note, Lemma 5 holds regardless of the distribution of x_t; all that is required is that x_t is iid.

In the on-line technical companion, Online Appendix A shows how we derived the impulse response functions and related variances for our inventory systems. Online Appendix B shows the detailed analytical optimization of the AR(1) and IMA(0,1,1). Online Appendix C details the VBA code for a Microsoft Excel Add-in that can be used to find $\alpha^\star$ numerically for AR(1) and IMA(0,1,1) demand. The Add-in is based on the Regula-Falsi method.

Online Appendix A. Deriving variances using Z-transforms

This Appendix gives an overview of how Z-transforms, a basic analysis tool in control theory, were used to calculate the inventory and order variance. For the unfamiliar reader, we first derive the familiar variances for the full off-shoring setting, before tackling dual sourcing systems.

8.1. Deriving the full off-shoring order and inventory variances

We start with the $Z-$transform of the inventory balance equation,

$$i_t = i_{t-1} - d_t + q^g_{t-L}, \tag{55}$$

for the single sourcing setting:

$$I_s[z] = z^{-1}I_s[z] - D[z] + z^{-L}Q_g[z] \Rightarrow I_s[z] = \left(z^{-L}Q_g[z] - D[z]\right)\frac{z}{z-1}. \tag{56}$$

This defines the inventory transfer function, $I_s[z]$, as a function of the order and demand transforms $Q_g[z]$ and $D[z]$. This inventory relationship can be graphically represented in a control block diagram[20] by blocks labeled b, c, d in Fig. 10 (the letters refer to the short-hand notation in the top-left corner of the blocks in Fig. 10). In our single off-shore supplier setting, the OUT replenishment policy (75) is the optimal linear policy. The order transfer function is found by first transforming the inventory position (the sum of current inventory and outstanding receipts) to yield

$$IP[z] = I_s[z] + \sum_{i=1}^{L-1} z^{-i}Q_g[z] \Rightarrow IP[z] = I_s[z] + \frac{1-z^{1-L}}{z-1}Q_g[z], \tag{57}$$

represented by block e in Fig. 10. Next, we convert the demand and forecasting difference equations ($d_t = \mu + \epsilon_t$ and $\hat{d}^\star_{t+n,t} = \mu$ for iid; and (22)-(30) for AR(1)), into transfer functions. The transform of the demand ($D[z]$) and the forecast of demand over $L+1$ periods ($A[z]$) is shown in Table 7 for all three demand processes (iid, AR(1), and IMA(0,1,1)). These transforms correspond to block d and a in Fig. 10, completing the block diagram.

Remark. In the time domain, under a random demand, the long-run expected inventory $\mathbb{E}[i_t] = i^\star$. As the variance is the expected squared deviation from the mean, the specific value of $i^\star$ has no impact on var$[i_t]$, which we aim to calculate. Likewise, the long-run expected demand, $\mathbb{E}[d_t]$ and the long-run expected replenishment quantity $\mathbb{E}[q^g_t]$ have no influence on the var$[d_t]$ and var$[q^g_t]$ respectively. □

The $Z-$transform of the order-up-to replenishment policy (75) is

$$Q_g[z] = A[z]D[z] - IP[z]. \tag{58}$$

[20] Block diagrams are a standard control theory technique, see Nise [2004].

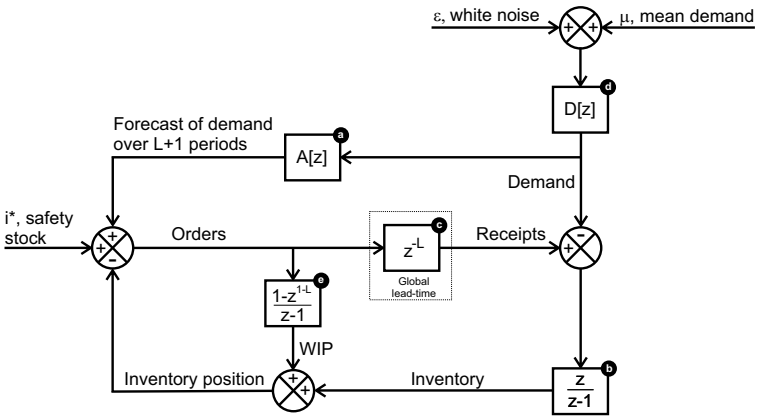


Figure 10 Block diagram of the full off-shoring setting.

Transfer function	iid demand	AR(1) demand	IMA(0,1,1) demand
A[z]	0	$\frac{\rho(\rho^L-1)}{\rho-1}$	$\frac{z\beta L}{z-1+\beta}$
D[z]	1	$\frac{z}{z-\rho}$	$1+\frac{\beta}{z-1}$

Table 7 Full off-shoring transfer functions.

Substituting (56) and (57) into (58) results in

$$Q_g[z] = A[z]D[z] - IP[z] \tag{59}$$

$$= A[z]D[z] - I_s[z] - \frac{1-z^{1-L}}{z-1}Q_g[z] \tag{60}$$

$$= A[z]D[z] - (z^{-L}Q_g[z] - D[z])\frac{z}{z-1} - \frac{1-z^{1-L}}{z-1}Q_g[z]. \tag{61}$$

Collecting together all the $Q_g[z]$ and simplifying gives

$$Q_g[z]\left(1 + z^{-L}\frac{z}{z-1} + \frac{1-z^{1-L}}{z-1}\right) = A[z]D[z] + D[z]\frac{z}{z-1} \tag{62}$$

$$Q_g[z] = \frac{A[z]D[z] + D[z]\frac{z}{z-1}}{1 + z^{-L}\frac{z}{z-1} + \frac{1-z^{1-L}}{z-1}} \tag{63}$$

$$Q_g[z] = D[z]\left(1 + A[z] - \frac{A[z]}{z}\right). \tag{64}$$

After substituting in the relevant expressions for $A[z]$ and $D[z]$ as given in Table 7, simple algebra leads to the transfer functions for the global orders $Q_g[z]$. Knowing $Q_g[z]$, the inventory transfer function $I_s[z]$ can be found from (56).

We are now ready to determine the variances of the orders and inventory levels from the relevant transfer function using Tsypkin's Relation (Lemma 3). Putting $A[z] = 0$ and $D[z] = 1$ into (64) provides the transfer function of the global orders under iid demand:

$$Q_g[z] = 1. \tag{65}$$

Taking the inverse z-transform of (65) gives the order impulse response function

$$q_t^g = \mathscr{Z}^{-1}[1] = \delta[t], \tag{66}$$

which equals the impulse function (Kronecker delta). Using Tyspkin's relation,

$$\sigma_{q,s}^2 = \sigma^2 \sum_{t=0}^{\infty} \delta[t]^2 = \sigma^2. \tag{67}$$

Note, the OUT policy behaves as a *pass-on-orders* system and as expected $\sigma_{q,s}^2 = \sigma^2$.

Using (65) inside of (56) provides the inventory transfer function,

$$I_s[z] = \frac{z(z^{-L}-1)}{z-1} \tag{68}$$

which has a time domain response given by it inverse z-transform,

$$i_t^s = \mathscr{Z}^{-1}\left[\frac{z(z^{-L}-1)}{z-1}\right] = -\mathscr{Z}^{-1}\left[\sum_{t=0}^{L-1} z^{-t}\right] = h[t-L] - h[t], \tag{69}$$

where $h[\cdot]$ is the unit step function; that is, $h[t<0]=0$, otherwise $h[t\geq 0]=1$. Applying Tyspkin's relation from Lemma 3 provides

$$\sigma_{i,s}^2|_{\text{IID}} = \sigma^2 \sum_{t=0}^{\infty} (h[t-L] - h[t])^2 = \sigma^2 L. \tag{70}$$

The variance of inventory in the presence of AR(1) and IMA(0,1,1) demand is obtained in exactly the same manner. As these expressions can be quite lengthy, we used Mathematica (Wolfram Research) to help with the algebra involved. For clarity, we also provide a sketch of the derivation of the remaining variance expressions in what remains of Online Appendix A.

8.1.1. Deriving the full off-shoring inventory variance under AR(1) demand Departing from (64), we substitute in the transfer function of demand, $D[z] = z/(z-\rho)$, and the transfer function of the mechanism that converts the demand into the forecast of demand over the lead-time, $A[z] = \rho(\rho^L-1)/(\rho-1)$; simplifying yields the order transfer function,

$$Q_{g,s}[z] = \frac{z}{z-\rho}\left(1 + \frac{\rho(\rho^L-1)}{\rho-1} - \frac{\rho(\rho^L-1)}{z(\rho-1)}\right) = \frac{\rho - z + (z-1)\rho^{1+L}}{(z-\rho)(\rho-1)}. \tag{71}$$

Using (71) inside (56) and simplifying provides the inventory transfer function,

$$I_s[z] = \frac{\rho^{L+1}}{\rho-1}\left(\frac{z^{1-L}}{z-\rho}\right) + \frac{1}{1-\rho}\left(\frac{z^{1-L}}{z-1} + \frac{(\rho-1)z^2}{(z-1)(z-\rho)}\right). \tag{72}$$

Taking the inverse $Z-$transform of reveals the inventory impulse response,

$$i_t^s = \frac{\rho^{L+1}}{\rho-1}\big(\rho^{t-L}h[t-L]\big) + \frac{1}{1-\rho}\big(h[t-L] + \rho^{t+1} - 1\big) = \begin{cases} \frac{\rho^{t+1}-1}{1-\rho} & \text{if } t < L, \\ 0 & \text{if } t \geq L. \end{cases} \tag{73}$$

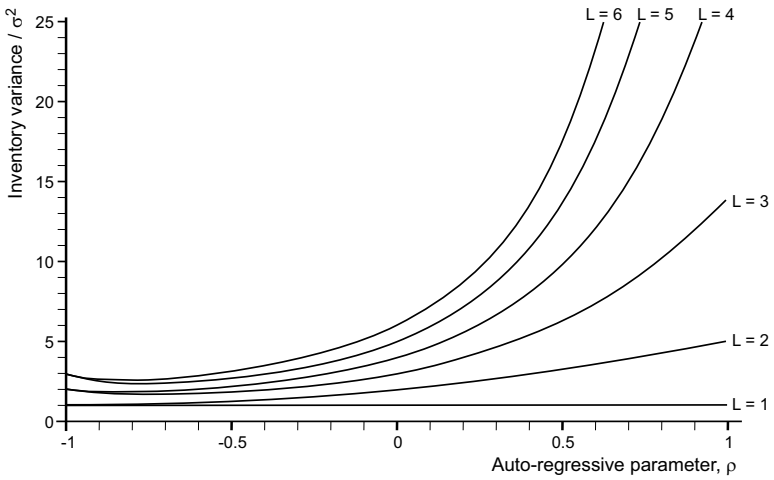


Figure 11 The effect of auto-correlated demand on the inventory in a fully off-shored supply chain. (Recall that $\sigma^2_{i,s}/\sigma^2 = L$ for iid demand).

Tyspkin's sum of squares then provides the inventory variance,

$$\sigma^2_{i,s}|_{\mathrm{AR}} = \frac{\sigma^2}{(\rho-1)^2}\left(\frac{\rho(\rho(\rho^L-1)-2)(\rho^L-1)}{\rho^2-1}+L\right) \tag{74}$$

which matches the inventory variance given by Disney and Lambrecht [2008]. With unit lead time $L=1,\ \sigma^2_{i,s}=\sigma^2$. As with iid demand, the inventory variance is increasing in lead time L, but superlinearly if $\rho>0$ and sublinearly if $\rho<0$. It also is convex in ρ, see Figure 11.

8.1.2. Deriving the single sourcing inventory variance under IMA(0,1,1) demand

For IMA(0,1,1) demand and its forecast introduced in §6.2, single-sourcing under a linearized OUT replenishment policy becomes

$$q^g_t = \underbrace{\left(i^\star + \sum_{i=1}^{L}\hat{d}^\star_{t+i,t}\right)}_{\text{Target inventory position}} - \underbrace{\left(i_t + \sum_{i=1}^{L-1} q^g_{t-i}\right)}_{\text{Inventory position}} \tag{75}$$

$$= i^\star + L\hat{d}^\star_{t+n,t} - i_t - \sum_{i=1}^{L-1} q^g_{t-i}. \tag{76}$$

Remark 1. At time $t=0$, the long-run variance of the demand and forecast is infinite. As the forecast is added directly into the replenishment order (see (76)), the long-run variance of the orders, q^g_t, is also infinite. We assume the supplier can provide items at a unit purchase price, p, and the supplier either has sufficient in-house capacity, or can subcontract excess demand to another supplier with the same lead time, quality, and cost characteristics. □

Remark 2. The demand variance under IMA(0,1,1) demand is given by $\sigma^2_d = \sigma^2(1+\sum_{n=1}^{\infty}\beta^2)$; the variance of the inventory maintained by the OUT policy is $\sigma^2_{q,g} = \sigma^2((1+L\beta)^2+\sum_{n=1}^{\infty}\beta^2)$. The difference $\sigma^2_{q,g}-\sigma^2_d = \sigma^2 L\beta(2+L\beta)$ reveals bullwhip is always generated as $\beta>0$ and $L\geq 1$. □

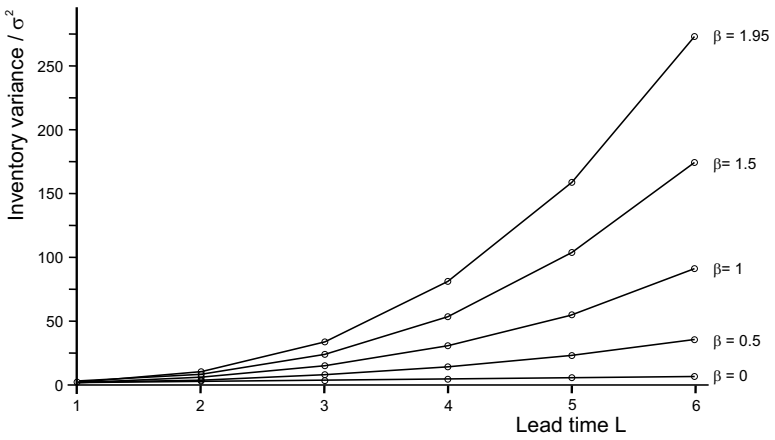


Figure 12 The variance of the inventory levels when outsourcing all production to an off-shore supplier under IMA(0,1,1) demand.

Graves [1999] shows that the variance of the inventory levels under single sourcing is given by (33). With unit lead time $L=1$, $\sigma^2_{i,s}=\sigma^2$. As noted by Graves [1999] and shown in Figure 12, the inventory variance, which was linear in the lead-time L for iid demand, is now convex increasing in L and in β. This means when β is sufficiently large, the inventory costs will also be convex in L. When $\beta=0$, the iid expression is recovered. When $\beta=1$, the IMA(0,1,1) demand process degenerates into the AR(1) process with $\rho=1$ and $\sigma_{i,s}=\sigma^2(L^2+L(2L^2-3L+1))/6$.

To verify the inventory variance expression of Graves [1999] with our z-transform approach, we depart from (64), this time substituting in $D[z]=1+\beta/(z-1)$, for the demand and, $A[z]=z\beta L/(z-1+\beta)$, for the lead-time demand forecast generating mechanism. Simplifying yields the transfer function of the orders,

$$Q_{g,s}[z]=1+\beta\left(L+\frac{1}{z-1}\right). \tag{77}$$

Placing (77) inside (56) and simplifying yields the IMA(0,1,1) inventory transfer function,

$$I_s[z]=\frac{z}{z-1}\left(z^{-L}\left(1+\beta L+\frac{\beta}{z-1}\right)-\frac{\beta}{z-1}-1\right). \tag{78}$$

Taking the inverse $Z-$transform gives the inventory transfer function in the time domain;

$$i^s_t=(1+\beta L+\beta(t-L))h[n-L]-t\beta-1=\begin{cases}-1-t\beta & \text{if } t<L,\\ 0 & \text{if } t\geq L.\end{cases} \tag{79}$$

Finally, summing the squared impulse response provides the inventory variance,

$$\sigma^2_{i,s}|_{\text{IMA}}=\sigma^2\sum_{t=0}^{L-1}\left(-1-t\beta\right)^2=\sigma^2L\big(1+\beta(L-1)+\beta^2(L-1)(2L-1)/6\big), \tag{80}$$

confirming Graves [1999] and used in (33).

Transfer function	iid demand	AR(1) demand	IMA(0,1,1) demand
$A[z]$	0	0	$\frac{\beta}{1-(1-\beta)z^{-1}}$
$B[z]$	0	ρ	$\frac{\beta}{1-(1-\beta)z^{-1}}$
$C[z]$	0	0	$\frac{\beta z^{1-L}}{1-(1-\beta)z^{-1}}$
$D[z]$	1	$\frac{z}{z-\rho}$	$\frac{\beta+z-1}{z-1}$

Table 8 Dual sourcing transfer functions.

8.2. Obtaining the dual sourcing order and inventory variances

We now consider the variance of the inventory levels and the replenishment orders in the dual sourcing setting. A new block diagram is required, see Fig. 13. This is based on the dual sourcing inventory balance equation (1), the demand processes and their forecasts, (22)-(30), the replenishment rules, (9) and (32), and the off-shore orders, (8) and (31). Specific demand processes require substitutions within the block diagram as detailed in Table 8. The OUT policy can be accessed by setting $\alpha = 0$, POUT requires $-1 < \alpha \geq 1$ for stability.

Equation (63) hints at another way to obtain the transfer function for the local orders directly from the block diagram. First sum all the *feed-forward routes*, from the white noise input to the output, the local orders; $(bd - cd + dfh - adefh)$. Next, divide by $1 -$ *sum of the feedback loops*, from the local orders back to the local orders leading to

$$\begin{aligned} Q[z] =& (bd - cd + dfh - adefh)/(1 + fgh) \\ =& \frac{D[z]z^{-L}\left(A[z](\alpha-1)z + z^L(B[z](z-1) + C[z](z-1) - \alpha z + z)\right)}{z-\alpha}. \end{aligned} \tag{81}$$

After substituting and simplifying, we arrive at transfer functions for the local replenishment orders, $Q[z]$; Table 8 provides the required information. The levels $\{i^\star, \mu\gamma, (1-\gamma)\mu\}$ have no consequence and can be ignored when determining the variance. The transfer function for the inventory levels in dual sourcing mode, $I[z]$, can be found from

$$I_d[z] = \frac{z}{z-1}\big(Q[z]z^{-1} + D[z](A[z]z^{-L} - 1)\big). \tag{82}$$

8.2.1. Dual sourcing variances under iid demand Placing $A[z] = B[z] = C[z] = 0$ and $D[z] = 1$, see Table 8, into (81) and simplifying yields the transfer function of the orders:

$$Q_d[z] = \frac{z(1-\alpha)}{z-\alpha}. \tag{83}$$

Remark. The poles and zeros (the roots of the denominator and numerator w.r.t. z respectively) of a system's transfer function, see (84) or (84), are required to lie within the unit circle in the complex plane for stability. Thus, stability[21] is achieved when $-1 < \alpha \leq 1$. □

[21] Poles and zeros are allowed to lie on the unit circle at $z = 1$; hence the inequality $\alpha \leq 1$.

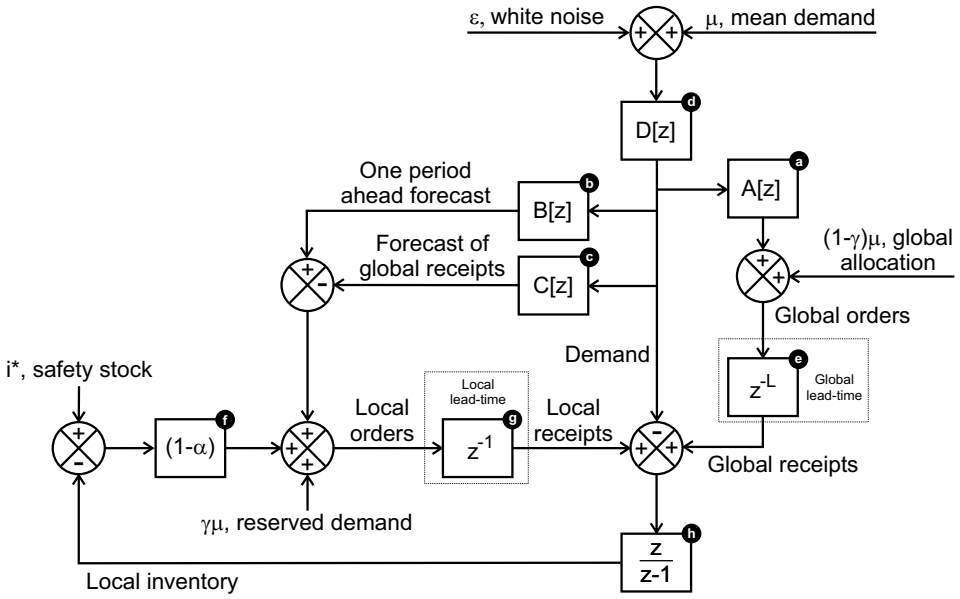


Figure 13 Generic block diagram of the dual sourcing for both the TBS and DYN settings.

Taking the inverse $Z-$transform of the transfer function of the orders, (83) results in

$$q_t^d = \mathscr{Z}^{-1}\left[\frac{z(1-\alpha)}{z-\alpha}\right] = (1-\alpha)\mathscr{Z}^{-1}\left[\frac{1}{1-\alpha z^{-1}}\right] = (1-\alpha)\mathscr{Z}^{-1}\left[\sum_{t=0}^{\infty}\alpha^t z^{-t}\right] = (1-\alpha)\alpha^t. \tag{84}$$

Applying Tsypkin's relation to (84), we obtain the variance expression,

$$\sigma_q^2|_{\text{IID}} = \sigma^2\sum_{t=0}^{\infty}((1-\alpha)\alpha^t)^2 = \sigma^2\left(\frac{1-\alpha}{1+\alpha}\right). \tag{85}$$

Putting (83) into (82) and simplifying gives the inventory transfer function,

$$I_d[z] = \frac{z}{\alpha - z}. \tag{86}$$

The variance of the inventory levels requires us to take the inverse z-transform of the inventory level transfer function. For the iid case,

$$i_t^d = \mathscr{Z}^{-1}\left[\frac{z}{\alpha-z}\right] = (1-\alpha)\mathscr{Z}^{-1}\left[\sum_{t=0}^{\infty}\alpha^t z^{-t}\right] = -\alpha^t, \tag{87}$$

to which we apply Tsypkin's relation to obtain the following expression for the inventory variance,

$$\sigma_i^2|_{\text{IID}} = \sigma^2\sum_{t=0}^{\infty}(-\alpha^t)^2 = \sigma^2\left(\frac{1}{1-\alpha^2}\right). \tag{88}$$

8.2.2. Dual sourcing variances under AR(1) demand The local order transfer function in the dual sourcing setting is found by substituting $A[z]$, $B[z]$, $C[z]$, and $D[z]$ from Table 8 into (81) and simplifying,

$$Q[z] = z\left(\frac{z-1}{\alpha-z} + \frac{z}{z-\rho}\right). \tag{89}$$

Taking the inverse $Z-$transform provides the impulse response of the local orders in response in the dual sourcing scenario under AR(1) demand,

$$q_t = \mathscr{Z}^{-1}[Q[z]] = \alpha^t - \alpha^{t+1} + \rho^{t+1}, \tag{90}$$

which can be used inside Tsypkin's relation, (54), to provide the variance of the local orders under AR(1) demand shown in (26).

The transfer function of inventory levels is found by placing (89) into (82) and simplifying,

$$I[z] = \frac{z}{\alpha - z}. \tag{91}$$

Taking the inverse $Z-$transform of (91) provides the impulse response the inventory levels under AR(1) demand in the time domain,

$$i_t = -\alpha^t. \tag{92}$$

By Tsypkin's relation the sum square of (92) provides the variance of the FGI in the dual sourcing setting given by (27).

8.2.3. Dual sourcing variances under IMA(0,1,1) demand The transfer function of the local orders in the dual sourcing setting is obtained by substituting the relevant $A[z]$, $B[z]$, $C[z]$, and $D[z]$ from Table 8 into (81) and simplifying:

$$Q_d[z] = \frac{z\left(\alpha(1-\beta) + \beta z^{1-L}(\alpha - z) + z(\beta - \alpha + 1) - 1\right)}{(z-1)(z-\alpha)}. \tag{93}$$

The time-domain impulse response of the local orders is given by the inverse $Z-$transform of (93),

$$q_t = \mathscr{Z}^{-1}[Q_d[z]] = \frac{(\beta - \alpha\beta)h[t+1-L] + (\alpha - \beta - 1)\left(1 - \alpha^{t+1}\right) - (\alpha(\beta-1)+1)\left(\alpha^t - 1\right)h[t-1]}{\alpha - 1}. \tag{94}$$

Summing the square of (94) over $t = 0$ to ∞ gives the variance expression in (34).

The inventory variance is obtained by first using (93) in (82) to find the inventory transfer function,

$$I[z] = \frac{z}{\alpha - z}. \tag{95}$$

Taking the inverse $Z-$transform reveals the inventory impulse response,

$$i_t = \mathscr{Z}^{-1}[I[z]] = -\alpha^t. \tag{96}$$

Squaring (96) inside the Tsypkin sum provides the inventory variance expression in (34).

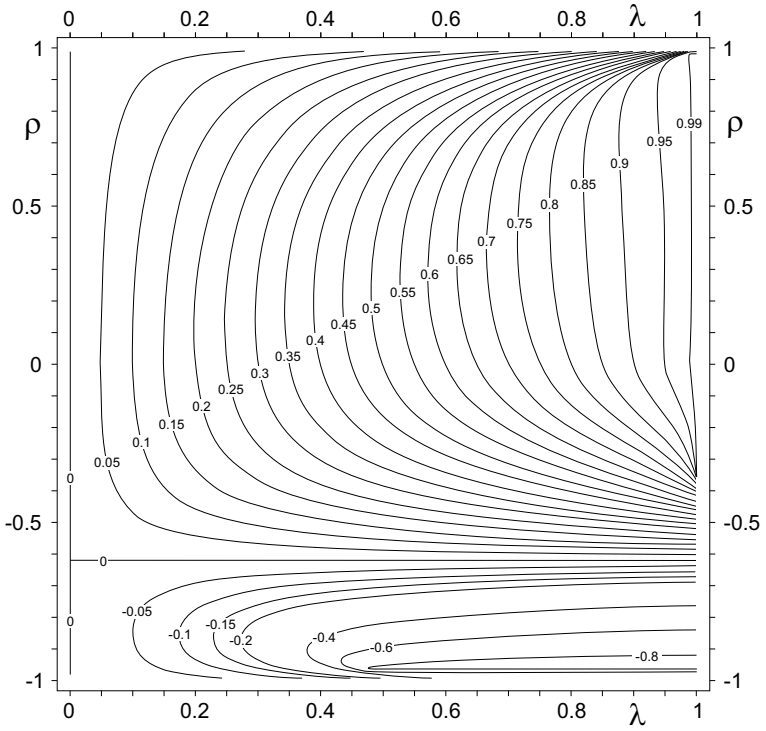


Figure 14 The contour lines of the optimal smoothing $\alpha^\star$ as a function of capacity intensity λ (horizontal axis) and serial correlation ρ (vertical axis) for the TBS-POUT policy under AR(1) demand. Higher capacity intensity requires more production smoothing. Around the natural neighborhood of $\rho > 0$, higher serial correlation leads to less smoothing.

Online Appendix B. Analytic optimization under AR(1) and IMA(0,1,1) demand

This Appendix presents the analytic optimization of α under AR(1) and IMA(0,1,1).

8.3. Correlated AR(1) demand

$f_{\rm AR}[\alpha^\star] = \lambda$ is easily solved numerically for $\alpha^\star \in [-1, 1]$ given $\lambda \in [0, 1]$ and $\rho \in [-1, 1]$, as shown in Fig. 14. Unfortunately, the analytic solution of $f_{\rm AR}[\alpha^\star] = \lambda$ requires the roots of a 7-th order polynomial and we know that no general analytic solutions exist for polynomials above 4-th order. For the practicing manager/analyst Online Appendix C provides the VBA code for a Microsoft Excel Add-in that can numerically determine $\alpha^\star$ for a given cost function under AR(1) demand. The algorithm is based on the Regula-Falsi method. We also provide an approximate analytic solution that is a bound:

Lemma 4 *For TBS-POUT with AR(1) demand, the optimal smoothing $\alpha^\star = g_{AR}[\rho]\lambda + o[\lambda]$ where*

$$g_{AR}[\rho] = 1/f'_{AR}[0] = \frac{1+\rho-\rho^2}{\sqrt{\frac{2\rho^3-2\rho-1}{\rho^2-1}}}. \tag{97}$$

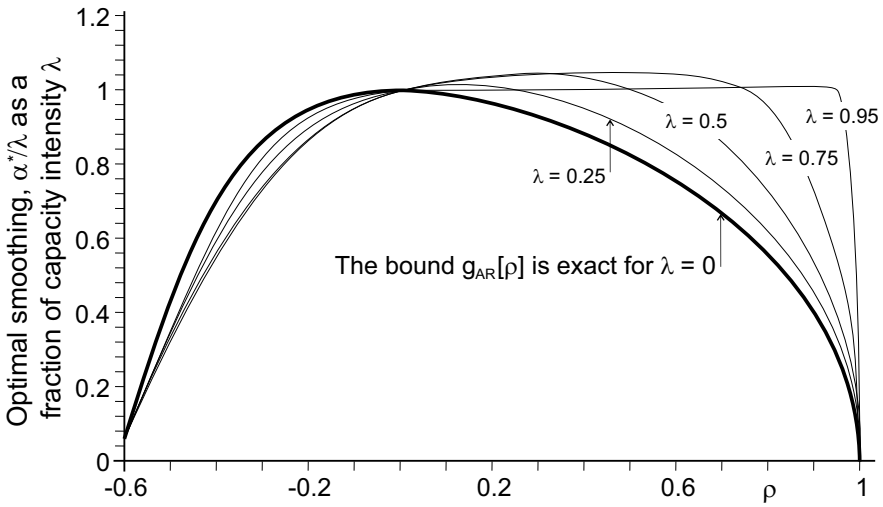


Figure 15 **The graph shows $\alpha^\star/\lambda$ for five capacity intensities $\lambda \in \{0, 0.25, 0.5, 0.75, 0.95\}$ and the linear approximation $g_{\mathbf{AR}}[\rho]$ which is exact for $\lambda = 0$ and a bound otherwise. For practical (positive) correlation values, AR(1) demand series with larger serial correlation require less local order smoothing.**

The linear approximation $g_{AR}[\rho]\lambda$ is a lower (upper) bound for $\alpha^\star$ when serial correlation ρ is positive (negative) and is asymptotically correct for (i) $\lambda \to 0$ and (ii) $\rho \to 0$. □

Lemma 4 shows that, in the natural neighborhood of $\rho > 0$, higher serial correlation leads to less smoothing, as shown in Fig. 15. This is sensible as more negative serial correlation typically results in higher demand fluctuations for which inventory benefits from more smoothing. Lemma 4 also shows that $\alpha^\star$ is positive if $\rho > \overline{\rho} = -0.6180$ (the negative zero of $g_{\mathrm{AR}}[\rho]$ which interestingly is the reciprocal of the golden ratio). Positive smoothing is the natural regime to consider as negative smoothing implies that deviations from the target inventory are over-corrected in each replenishment decision. As most real demand patterns are positively auto-correlated, a positive α is likely in practice.

8.4. IMA(0,1,1) demand

$f_{\mathrm{IMA}}[\alpha^\star] = \lambda$ is easily solved numerically for $\alpha^\star \in [-1, 1]$ given $\lambda \in [0, 1]$ and $\beta \in [0, 2]$, as shown in Fig. 16 for different lead times L. Online Appendix D provides the VBA code to determine $\alpha^\star$ numerically with the Regula-Falsi method in a Microsoft Excel Add-in. If $L = 2$, optimal smoothing roughly equals the capacity intensity λ, almost irrespective of β. Yet as lead-times increase, the optimal smoothing decreases as the non-stationarity β increases, except for high capacity intensity for which smoothing is extremely high (and insensitive to β). Unfortunately, there is no general analytic solution of $f_{\mathrm{IMA}}[\alpha^\star] = \lambda$ but we can provide an approximate analytic solution that is a bound:

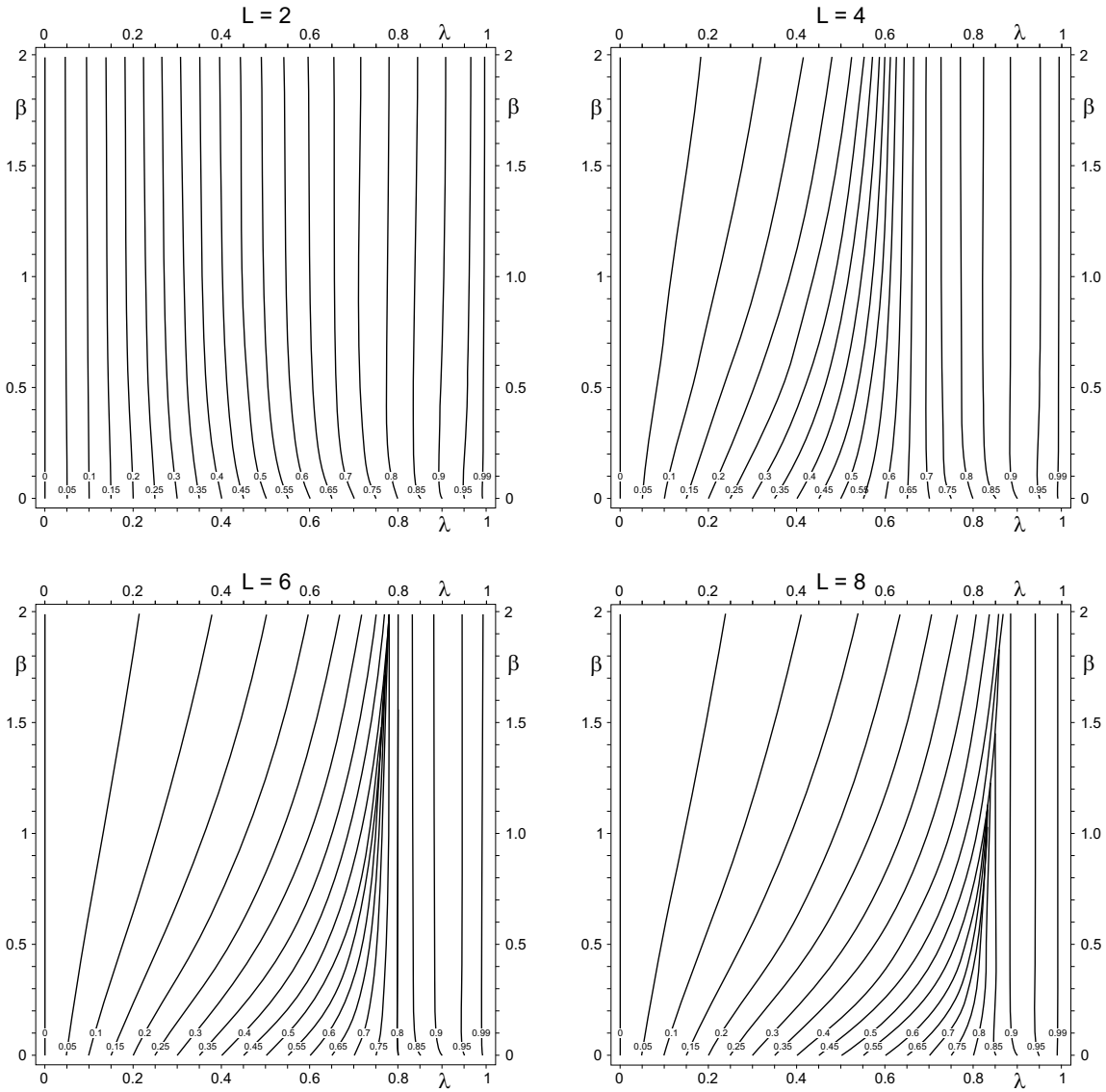


Figure 16 The contour lines of the optimal smoothing $\alpha^\star$ under IMA(0,1,1) demand: as lead-times increase, the optimal smoothing decreases as β increases, except for high capacity intensity for which smoothing is extremely high (and insensitive to β).

Lemma 5 *For DYN-POUT with IMA(0,1,1) demand, the optimal smoothing* $\alpha^\star = g_{IMA}[\beta, L]\lambda + o[\lambda]$ *where*

$$g_{IMA}[\beta, L] = 1/f'_{IMA}[0] = \begin{cases} 1 & \text{if } L = 2 \\ 1/\sqrt{(L-1)\beta^2 + 2\beta + 1} & \text{if } L > 2 \end{cases}. \tag{98}$$

The linear approximation $g_{IMA}[\beta, L]\lambda$ *is a lower bound for* $\alpha^\star$ *and is asymptotically correct for (i)* $\lambda \to 0$ *and (ii)* $\beta \to 0$. □

Lemma 5 shows that for small capacity intensities λ, optimal smoothing decreases in β, as shown in Fig. 17. Indeed, $\alpha^\star$ is roughly proportional to $1/\beta$ and to $1/\sqrt{L}$, implying that optimal smoothing decreases in both the non-stationarity β and the off-shore lead time L.

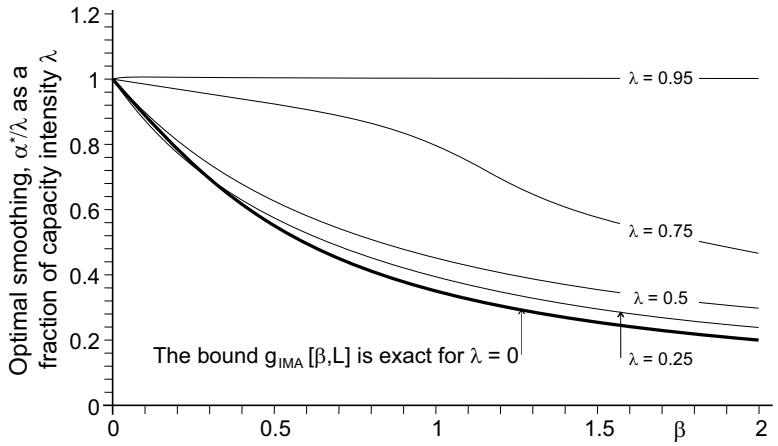


Figure 17 **The graph shows $\alpha^\star/\lambda$ for $L=6$ and for five capacity intensities $\lambda \in \{0, 0.25, 0.5, 0.75, 0.95\}$ and the linear approximation $g_{\mathsf{IMA}}[\beta, L]$ which is exact for $\lambda=0$ and a lower bound otherwise. IMA(0,1,1) demand series with higher non-stationarity require less local order smoothing.**

Online Appendix C. Finding $\alpha^\star$ numerically under AR(1) and IMA(0,1,1) demand with the Regula-Falsi method

A user defined function can be added to Microsoft Excel for numerically finding $\alpha^\star$ using the following VBA code. The function is based on the Regula-Falsi method and is guaranteed to find a solution. For AR(1) demand the $\alpha^\star$ for the TBS-POUT policy can be accessed with "=DSAlphaStar(0,ρ,λ)"; for IMA(0,1,1) demand, the $\alpha^\star$ for the DYN-POUT policy can be accessed by "=DSAlphaStar(L,β,λ)". Due to an asymptote in the AR(1) case, the method is not reliable when $\rho < 0.61803$, so we a have included an escape that outputs a warning when this issue has the potential to happen.

```
Option Explicit
Function Fp(p As Double, L As Integer, br As Double, lambda As Double)
Dim p1, p2 As Double

If L > 0 Then
    p1 = ((-1 + 2 / (1 + p) + br * (2 - 2 * p ^ (L - 1) + (L - 1) * br)) ^ 0.5)
    p2 = ((((p - 1) ^ 2) * (1 / (1 - p ^ 2)) ^ 0.5 * (p ^ 2 + (L - 1) * (p ^ L) * ((1 + p) ^ 2) * br)) / (p ^ 3))
Else
    p1 = (-(-1 + p + (-2 + p + p ^ 2) * br - 2 * p * br ^ 2 + 2 * br ^ 3) / ((1 + p) * (-1 + p * br) * (-1 + br * br))) ^ 0.5
    p2 = (((1 / (1 - p ^ 2)) ^ 0.5) * (-1 + p) ^ 2 * (1 + br * (1 + p ^ 2 - br - 2 * p * br))) / (p * (-1 + p * br) ^ 2)
End If
Fp = (p1 / (p2 + p1)) - lambda
End Function

Function DSAlphaStar(L As Integer, br As Double, lambda As Double)
Dim A, B, C, Fa, Fb, Fc, p1, p2 As Double
Dim grandloop As Integer
Dim p As Double

If br < -0.61803398875 Then
    DSAlphaStar = "When rho<-0.61803 this method is not reliable"
Else:
    A = -0.999999
    B = 0.999999
    C = 0
    For grandloop = 1 To 1000
       p = A
       Fa = Fp(p, L, br, lambda)
        p = B
       Fb = Fp(p, L, br, lambda)
        C = (A * Fb - B * Fa) / (Fb - Fa)
        p = C
        Fc = Fp(p, L, br, lambda)
        If Fc * Fa > 0 Then
```

```
            A = C
        Else: B = C
        End If
        If Abs(Fc) < 0.00000000001 Then
           grandloop = 1000
    End If
Next grandloop
DSAlphaStar = C
End If
End Function
```