Managing Self-Replicating Innovative Goods

Bin Hu
Naveen Jindal School of Management, University of Texas at Dallas, Richardson, TX 75080, USA
bin.hu@utdallas.edu

Zhankun Sun
Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong
zhankun.sun@cityu.edu.hk

Inspired by self-replicating 3D printers and innovative agricultural and husbandry goods, we study optimal production and sales policies for a manufacturer of self-replicating innovative goods with a focus on the unique "keep-or-sell" trade-off, namely whether a newly produced unit should be sold to satisfy demand and stimulate future demand or added to inventory to increase production capacity. We adopt the continuous-time optimal control framework and marry a self-replication model on the production side to the canonical innovation diffusion model on the demand side. By analyzing the model, we identify a condition that differentiates Strong and Weak Replicability regimes wherein production and sales respectively take priority over the other, and fully characterize their distinct optimal policies. These insights prove robust and helpful in several extensions, including backlogged demand, liquidity constraints, stochastic innovation diffusion, launch inventory decision, and exogenous demand. We also find that social marketing strategies are particularly well-suited for self-replicating innovative goods under Strong Replication.

Key words: Innovation diffusion, Bass model, optimal control, production planning, inventory management

1. Introduction

3D printing, also known as additive manufacturing, is an emerging manufacturing technology which operates on the principle of depositing layers of material following 3D blueprints to form objects (Additive Manufacturing 2018). The most common printable materials are various plastics, although technologies have been developed to print metals (The Verge 2017), concrete (CNN 2017), food (NASA 2013), and even living cells (ScienceDaily 2018). Additive manufacturing has made its way into producing a wide range of products from customized shoes (MIT Sloan Management Review 2013) to rocket engines (The Guardian 2017), but one of the most fascinating products made with additive manufacturing is arguably 3D-printed 3D printers.

RepRap (short for *replicating rapid prototyper*) is an open-source project to develop 3D printers that can replicate themselves (RepRap.org 2018). A RepRap printer is designed to comprise in-house engineered plastic parts which are 3D-printable, and other commodities such as metal rods, bolts and nuts, motors, and circuit boards. As such, one can use a RepRap printer to print engineered parts and procure remaining commodities to build an identical or evolved “child” RepRap printer. The RepRap project has had a significant impact on the consumer-grade 3D printer industry. In 2015, Josef Prusa, a core developer of the RepRap project, began selling a RepRap printer kit of his own design called the Original Prusa i3 under his company Prusa Research[1]. All plastic parts in the kit were 3D-printed in the company’s “print farm” which is simply a room filled with the same printers. The printer was an immense success. Its design and quality coupled with an affordable price created a dedicated following, and there was a constant backlog of demand (3DPI 2016). By July 2019, Prusa Research’s print farm had grew from 15 printers to 500 printers, and it had shipped 130,000 3D-printed printers to more than 130 countries (Prusa Research 2020); see Figure 1 for a view of the print farm as of 2018. For his achievements, Josef Prusa was name in Forbes’ 2018 *30 Under 30 - Europe - Technology* list (Forbes 2018). Astonishingly, the company had grown its demand “without a sales team, through word of mouth and with the support of the international maker community (Digital Social Innovation 2017)”. Such dependency on word-of-mouth (previous sales stimulating new demands) rather than conventional marketing efforts is not uncommon in this highly innovative industry. Peter Misek, an analyst at Jefferies, interviewed over 15 3D printer exhibitors at a 3D Printshow in New York, and noted the 3D printer market had become a “word-of-mouth/branding game” (Business Insider 2014).

Prusa Research had taken advantage of the 3D printing technology to drive a business model where the manufacturing mode is self-replication and the demand for their innovative product is heavily driven by word-of-mouth. Self-replicating machines have existed for some time. The Japanese industrial robot manufacturer FANUC uses its own robots to make more robots (Bloomberg 2017). The same is happening in the Chinese factory of ABB, a Swiss-Swedish robotics and automation corporation (Reuters 2018). Yamazaki Mazak, a leading machining tool manufacturer, has machined needed components on their own Flexible Manufacturing Systems for decades (Jaikumar 1989). Such self-replicating innovative products exhibit unique and fascinating operational characteristics and trade-offs. For a conventional manufacturer, the production capacity is usually exogenous or ex-ante determined, the inventory primarily serves as a buffer against variability, and production and sales have no fundamental conflict. (Production and inventory

[1] Derivation and commercialization of open-source products are allowed as long as the sold products remain open-source.

Figure 1 Prusa Research's print farm (Prusa Research 2020)

management is a huge literature; for some recent developments, see Caro and Martínez-de Albéniz 2010, Berling and Martínez-de Albéniz 2011, Mayorga and Ahn 2011, Feng et al. 2013, Feng and Lu 2013, Feng et al. 2020.) For a self-replicating manufacturer like Prusa Research, however, such operational common sense is turned upside down. Prusa Research uses its own finished products for replication, therefore its production capacity is limited by its inventory and can be dynamically adjusted. After each 3D printer is produced, Prusa Research faces the trade-off of adding the new printer to its inventory to increase production capacity versus selling the printer to earn a revenue and stimulate future demand. With production, inventory, and sales deeply intertwined and the unique "keep-or-sell" trade-off, it is not straightforward how to manage the production and sales of self-replicating innovative goods.

Although self-replicating machines carry a futuristic aura, humanity has actually manufactured goods by self-replication since ancient times—in agriculture and husbandry—through planting seeds, rhizomes and cuttings, grafting, and breeding. Admittedly, not all agricultural goods are produced through self-replication (e.g., seedless watermelons are grown from triploid seeds, and mules are crossbred from horses and donkeys; both are sterile), and most agricultural and husbandry goods are staple commodities that face relatively stable demands (such as wheat and beef). Nonetheless, many innovative agricultural and husbandry goods are produced by self-replication and depend on word-of-mouth to drive demands, not unlike self-replicating 3D printers. In 2012, Native Seeds/SEARCH began selling seeds for a translucent rainbow-colored corn called Glass Gem (Figure 2); the corn soon went viral on social media and Native Seeds/SEARCH could not meet demand because they "did not grow out enough to sell" (Business Insider 2012). From 2016 to 2018, ten new dog breeds had debuted at the annual Westminster Dog Show (Bloomberg 2018), and their breeders must breed their dogs to meet growing demands. Such innovative agricultural and husbandry goods exhibit similar operational characteristics that their “inventories” (crops or herds) are used to produce off-springs and thus limit the production capacities, and the producers face a similar “reproduce-or-sell” trade-off.

Figure 2 Glass Gem corn (Native Seeds/SEARCH 2018)

A natural concern with self-replicating goods is that customers may produce their own goods through replication and even compete with the original producer, but in most cases the concern is not serious enough to threaten the business model. In the case of Prusa Research which publishes all printable part design files on their website (Prusa Research 2018), the sold kits contain not only printed parts but also other commodities which are not necessarily easily procurable by customers. Their pricing is also at such a level that replicating and selling 3D printers may not be lucrative for customers, considering the company’s economies of scale and established manufacturing expertise and distribution channels. On the contrary, the open-source printable design allows Prusa Research to push hardware updates to existing customers (who can download and print revised parts to replace old parts) in a manner similar to software developers pushing updates. For many agricultural and husbandry goods, the same economic argument would apply. Where it does not, legal measures may be taken to protect the business model; for example, Monsanto, a former US agricultural biotechnology giant, sued hundreds of farmers for replanting its patented seeds (The Guardian 2013). Therefore, despite the potential concern, the self-replication business model remains viable.

Inspired by these examples, we set out to study optimal production and sales policies for a manufacturer of self-replicating innovative goods. We adopt a continuous-time optimal control model which marries a self-replication model on the production side (where the production rate is limited by the inventory) to the canonical innovation diffusion model for new product adoption on the demand side (where existing sales stimulate new demand generation). Despite the problem's complexity, we are able to identify conditions for two regimes, in which we fully characterize the manufacturer's optimal production and sales policies. In the *Strong Replicability* regime, production takes priority over sales as long as the produced goods will eventually be sold, and sales may be held back even with enough inventory to satisfy demand. In the *Weak Replicability* regime, sales have priority over production and are never held back, although replication using unsold inventory may still be optimal. These insights are found to be robust and help characterize optimal policies in several extensions of the model.

In what follows, we review related literature in Section 2 and analyze the base model in Section 3 to derive our main results and insights. We then show in Section 5 that our main insights are robust and help characterize optimal policies in several extensions, including backlogged demand, liquidity constraints, stochastic innovation diffusion, launch inventory decision, and exogenous demand. We summarize the paper in Section 6. Additional analysis and proofs are relegated to the Appendix.

2. Literature review

This paper studies manufacturing innovative goods by self-replication to meet demand driven by word-of-mouth. Therefore, it is most related to the streams of literature on managing renewable (self-replicating) resources and on manufacturing innovative goods with demand diffusion. The literature on managing renewable resources dates back to Gordon (1954) and Schaefer (1957) who study fishery and Kilkki and Väisänen (1969) who study forestry; see Chapter 10 of Sethi (2019) for a summary of this literature. Their models share with ours the growth of resources through self-replication with two major differences. First, they model a simple unit revenue for the consumption of the resources, implicitly assuming infinite demand, whereas we model limited and endogenous demand that grows through innovation diffusion. Second, they assume uncontrolled natural resource growth (as is appropriate for their applications), whereas we model production as an explicit decision and show that it should be shut off in certain cases.

On the other hand, stemming from the classic paper by Bass (1969) grows a rich literature on innovation diffusion; see Mahajan et al. (1991), Mahajan et al. (2000) and Meade and Islam (2006) for reviews of this literature. The majority of this literature focuses on describing demand growth while ignoring a firm's ability to meet such demand, with a few notable exceptions. Jain et al. (1991) are among the first to consider innovation diffusion under supply constraints, although they take a strictly marketing perspective and focus on modeling how production constraints impede demand growth, rather than studying production decisions. In the operations management literature, Kumar and Swaminathan (2003) and Ho et al. (2002) (with a follow-up note by Ho et al. 2011) study the joint optimal production and sales decisions for innovative goods under supply constraints. Shen et al. (2011, 2014) further include pricing decisions in the model. These papers generally assume fixed production capacities. Our paper can be seen as an extension of the literature that study joint optimal production and sales decisions under demand diffusion and production capacity constraints, where the production capacity is dynamic and endogenous to the decisions. Bilginer and Erhun (2015) in a more stylized two-period discrete diffusion model also incorporate endogenous capacity decisions. By comparison, we adopt the canonical continuous-time optimal control framework, and our endogenous production capacity is driven by past production and sales decisions—a feature unique to the self-replication business model. Another contribution to the innovation diffusion literature is that we extend the model to accounts for diminishing word-of-mouth.

Topic-wise, our paper belongs to an emerging literature on the operations management of 3D printing. Various aspects of 3D printing are being investigated, including flexibility (Dong et al. 2016), customization (Sethuraman et al. 2018), retail (Chen et al. 2017, Arbabian and Wagner 2020), and spare part manufacturing (Song and Zhang 2020). We grow this literature by studying the self-replication business model inspired and enabled by 3D printing.

3. Base model

Following the canonical innovation diffusion literature (Kumar and Swaminathan 2003, Ho et al. 2002, 2011, Shen et al. 2011, 2014), we consider a continuous-time continuous-quantity model [2]. Suppose a firm develops an innovative good that can only be produced through self-replication, and at time 0 when the good is introduced to market, there is $I_L > 0$ units of initial launch inventory. The good can be sold from time 0 to time T which represents its maximum life cycle. The good's replication rate is r, meaning a unit of the good in a unit time can produce up to r units of itself. The cost to produce (replicate) a unit of the good is c. The retail price of the good is fixed at π. We denote the inventory level at time t by $I(t)$, and the inventory holding cost per unit good-time

[2] Although 3D printers are not infinitely divisible, with sufficiently large inventory and sales, a continuous-quantity model is a reasonably accurate approximation of the discrete operations. For reference, by July 2019, Prusa Research had shipped 130,000 3D printers produced in its print farm that had grown to have 500 printers (Prusa Research 2020). The same argument applies to animal herds. Crop grains are nearly infinitely divisible. On the other hand, when the replication cycle is relatively short compared with the product's life cycle and replication is not necessarily in batches, a continuous-time model is a reasonably accurate approximation of the discrete adjustments.

by h. At time T, we assume that all remaining inventory is discarded without loss of generality (it is obvious that assuming any salvage value below c does not change the optimal policies).

We denote the total market size for the good by m, namely at time 0, m consumers are potentially interested in the good. The classical innovation diffusion model is an ordinary differential equation (ODE) $D'(t) = [\alpha + \beta D(t)/m][m - D(t)]$ where $D(t)$ denotes the cumulative demand at t, namely the total number of consumers thus far having wanted to buy the good. The parameter α is called the coefficient of innovation which captures the rate at which potential customers discover the good by themselves. The model states that new demand ($D'(t)$) is generated as a percentage ($\alpha + \beta D(t)/m$) of the remaining market size ($m - D(t)$). The parameter β is called the coefficient of imitation which captures the rate at which "imitators" are influenced by existing customers into wanting the good, namely word-of-mouth. The model yields an S-shaped demand curve which starts slow, accelerates as customers increasingly generate word-of-mouth, but eventually slows down and tapers as the market comes saturated. Ho et al. (2002) and Kumar and Swaminathan (2003) propose a revised model $D'(t) = [\alpha + \beta S(t)/m][m - D(t)]$, where $S(t)$ denotes the cumulative sales at time t, by recognizing that when production is constrained, demand may not equal sales, and it should be cumulative sales rather than demand that generate word-of-mouth. While we adopt this demand model formulation, we make a further observation that the word-of-mouth effect may be the strongest immediately after purchase and diminishes over time. Therefore, in our model we let $S(t)$ denote the *adjusted* (rather than actual) cumulative sales to account for diminishing word-of-mouth; the details of the adjustment will be elaborated later.

The firm's objective is to maximize the total discounted profit by controlling instantaneous production rate $p(t)$ and sales rate $s(t)$ subject to constraints. We assume $\pi > c$ and $r > \rho$, where $\rho \geq 0$ is the continuous discounting factor, to rule out uninteresting cases where replication is obviously never optimal. In other words, the firm decides how many units of the good in the inventory are used for replication, how many units are delivered to satisfy demand, and how many units to simply keep in the inventory. In the base model we assume that all unmet demand is lost. The firm faces an optimal control problem, with control variables $p(t)$ and $s(t)$ and state variables $D(t), S(t)$, and $I(t)$:

$$\max_{p(t),s(t)} \int_0^T [\pi s(t) - cp(t) - hI(t)]e^{-\rho t}\,\mathrm{d}t \tag{1}$$

$$\begin{aligned}
\text{s.t.}\quad & D'(t) = d(t) \doteq [\alpha + \beta S(t)/m][m - D(t)],\ S'(t) = s(t) - wS(t),\ I'(t) = p(t) - s(t), \\
& 0 \leq p(t) \leq rI(t),\ 0 \leq s(t) \leq d(t),\ I(T) \geq 0, \\
& D(0) = S(0) = 0,\ I(0) = I_L > 0.
\end{aligned}$$

We would first like to explain the constraint $S'(t) = s(t) - wS(t)$. Solving this ODE under the boundary condition $S(t) = 0$ yields $S(t) = \int_0^t e^{-w(t-u)} s(u)\,\mathrm{d}u$. Therefore, $S(t)$ is the *discounted* integration of past sales rates with continuous discounting factor $w \geq 0$, which we referred to as the *adjusted* cumulative sales. When used in the innovation diffusion equation $D'(t) = [\alpha + \beta S(t)/m][m - D(t)]$, the adjusted cumulative sales capture the effect that distant past sales generate less word-of-mouth than more recent past sales. The parameter w captures the diminishing rate; in particular, when $w = 0$, $S(t) = \int_0^t s(u)\,\mathrm{d}u$ becomes the cumulative sales. Therefore, our model is an extension of that of Ho et al. (2002) and Kumar and Swaminathan (2003) to capture diminishing word-of-mouth. Note that $S(t)$ is not the actual cumulative sales (which would be a non-discounted integration of past sales rates); in particular the latter never decreases over time while the former may decrease (albeit never negative). In Section 4.1 we investigate a similar effect of sales intensity and find it to behave qualitatively similarly.

We further elaborate on three more constraints. First, $p(t) \leq rI(t)$ captures the self-replication characteristic that one's inventory limits its production capacity. Second, $s(t) \leq d(t)$ reflects the lost-sales assumption (see the discussion at the beginning of this section): because past unmet demands have been lost, one can only satisfy current demand. Also, because $s(t)$ is the instantaneous sales *rate* while $I(t)$ is the available inventory *level*, $s(t)$ is not directly constrained by $I(t)$. As long as $I(t) > 0$, any arbitrarily large $s(t)$ is allowed by the inventory *at the instant* t. Finally, one may expect a non-negative inventory constraint $I(t) \geq 0,\ t \leq T$. However, note that for a self-replicating good, whenever the inventory is depleted ($I(t) = 0$), production is no longer possible ($0 \leq p(t) \leq rI(t) = 0$) and the inventory cannot increase henceforth. This unique property means that a non-negative ending inventory ($I(T) \geq 0$) also guarantees a general non-negative inventory ($I(t) \geq 0,\ t \leq T$). Being able to replace the pure-state constraint $I(t) \geq 0$ by a terminal constraint $I(T) \geq 0$ greatly simplifies the optimal control problem.

An assumption implied by the formulation is that after replicating itself, a unit of the good remains in the original state and can be sold as new. This is clearly not true in practice, but one can easily adjust the replication rate parameter to account for wear. For example, suppose a 3D printer can print a printer kit each day, and will wear out after printing 100 kits. Mathematically, it is equivalent to assuming that a 3D printer can print 99/100 of a printer kit each day, but remains brand new after printing 99 kits (essentially replacing itself with one of its "offsprings"). In a continuous model, this logic can be applied in real time, i.e., the worn-out printers are replenished by some of their "offsprings" instantaneously.

We apply standard Pontryagin's maximum principle to Problem (1). The details are relegated to the Appendix. We then solve the problem based on the principle in multiple steps. All non-straightforward proofs are in the Appendix. An immediate result from the analysis is that the optimal control for Problem (1) is *bang-bang*; namely the optimal $p(t)$ is either 0 or $rI(t)$, and the optimal $s(t)$ is either 0 or $d(t)$. As such we will simply use "off" and "on" to refer to these specific levels, respectively. We first show two properties.

Proposition 1. *After production is switched on, it is never optimal to switch it off (until T or inventory depletion).*

Proposition 2. *Under the optimal policy, if $(r-\rho)\pi > cr+h$, then the product life cycle will last until T (i.e., $I(t)>0,\ \forall t<T$) regardless of how large T is; if $(r-\rho)\pi < cr+h$, then the product life cycle will be cut short (i.e., $I(t)=0,\ \tau \le t \le T,\ \exists \tau < T$) for sufficiently large T.*

Proposition 1 characterizes the optimal production policy structure. Proposition 2 further reveals an important condition that defines two fundamentally different parameter domains. When $(r-\rho)\pi > cr+h$, which we refer to as the *Strong Replicability* regime for reasons that will become clear later, it is optimal to keep the inventory positive (by maximizing production and/or curbing sales) and stay in business for arbitrarily large T. On the other hand, when $(r-\rho)\pi < cr+h$, which we refer to as the *Weak Replicability* regime, it may be optimal to sell out the inventory and end the product life cycle prematurely before T.

To understand the intuitive meaning of the strong replicability condition $(r-\rho)\pi > cr+h$, consider the firm with one unit of the good facing ample demand. The firm can sell the good now and earn revenue π. Alternatively, the firm may let the good replicate itself for a small time interval δ. The benefit of doing so is that after δ the firm will be able to sell $1+r\delta$ units of the good and earn $\pi(1+r\delta)$. There are costs though: replication itself costs $cr\delta$; holding the good for δ for replication incurs inventory cost $h\delta$ (omitting higher-order terms); and finally, earning the revenue δ later incurs discounting cost $\rho\pi\delta$ (omitting higher-order terms). By requiring that the benefit of replication outweighs the costs, we have

$$\pi(1+r\delta) - cr\delta - h\delta - \rho\pi\delta > \pi \Rightarrow (r-\rho)\pi > cr+h, \tag{2}$$

thus recovering the Strong Replicability condition in Proposition 2. Therefore, the interpretation of Strong Replicability is that *replication and selling later is more lucrative than immediate sales.* It is worth pointing out that this statement is stronger than "replication is profitable", which means that the value created by replication outweighs the direct cost of replication, or $\pi > c$ but does not account for indirect costs due to holding back sales. On the other hand, Weak Replicability $\pi < (cr + h)/(r - \rho)$ means that, if possible, the firm would sell the good immediately rather than using it for replication and selling it later. Note that Weak Replicability does not preclude profitable replication; i.e., $c < \pi < (cr + h)/(r - \rho)$ is possible.

The Strong and Weak Replicability insights allow us to show the next two theorems which completely characterize the optimal production and sales policies in the two regimes.

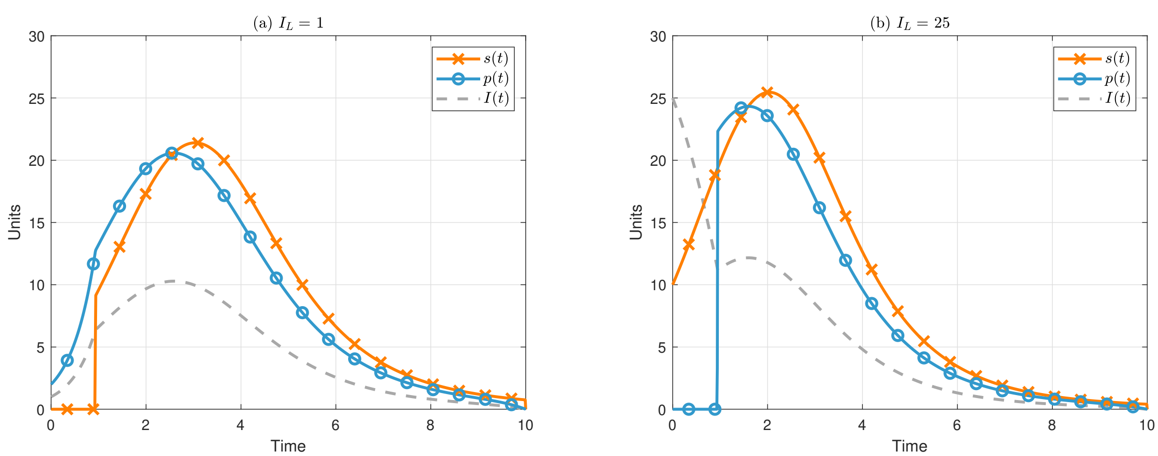
THEOREM 1 **(Strong Replicability)**. *Assume $(r - \rho)\pi > cr + h$. Let $\underline{I}$ be such a level of I_L that always keeping production and sales on leads to inventory depletion at exactly T. Let $\overline{I}$ be such a level of I_L that always keeping production off and sales on leads to inventory depletion at exactly T. Clearly $\underline{I} < \overline{I}$.*

(a) When $I_L < \underline{I}$, it is optimal to always keep production on, and switch and keep sales on at such a time that the inventory is depleted exactly at T.

(b) When $\underline{I} < I_L < \overline{I}$, it is optimal to always keep sales on, and switch and keep production on at such a time that the inventory is depleted exactly at T.

(c) When $I_L > \overline{I}$, it is optimal to always keep sales on and production off, and the inventory is not depleted at T.

Figure 3 Optimal policies of Problem (1) under Strong Replicability



Note. These examples are generated with $m = 100$, $\pi = 2.3$, $c = 2$, $h = 0.2$, $r = 2$, $\alpha = 0.1$, $\beta = 1$, $\rho = 0.1$, $w = 0.25$, and $T = 10$ (discretized into 1,000 periods). Strong Replicability holds: $(r - \rho)\pi = 4.37 > cr + h = 4.2$.

Corollary 1. *Assume* $(r-\rho)\pi > cr + h$. *The optimal production (sales) switch-on time increases (decreases) in the replication rate* r.

As we discussed after Proposition 2, Strong Replicability means that replication and selling later is more lucrative than immediate sales. For this reason, the firm should postpone sales despite lost revenues to maximize production, as long as the built inventory will eventually be sold. Therefore, the optimal policy is generally to always keep production on, and initially hold off sales to be switched on at such a time that the inventory is depleted exactly at T (Case (a) of Theorem 1, illustrated by Figure 3 (a)), and the optimal production (sales) switch-on time increases (decreases) in the replication rate. Unless, if the launch inventory I_L is sufficiently large that always keeping production on is not necessary (will lead to leftover inventory at T), the optimal policy is to always keep sales on, and initially hold off production to be switched on at such a time that the inventory is depleted exactly at T (Case (b) of Theorem 1, illustrated by Figure 3 (b)). In the extreme and unlikely case where the launch inventory I_L is so large that it cannot be sold through T, production is never needed (Case (c) of Theorem 1, not illustrated).

Although closed-form expressions are unavailable due to Problem (1)'s complexity, Theorem 1 completely characterizes the unique optimal production and sales policies, and the optimal time to switch on production/sales can be easily found through a one-dimensional search with linear computational complexity. Corollary 1 further characterizes the optimal switch-on times's sensitivity to the replication rate. (All figures in this paper are however directly evaluated from the original problems without assuming any structure, in order to verify the theoretical results.)

To summarize, under Strong Replicability, the firm should prioritize the future over the present: replication is kept on as long as future demand justifies it, sales are postponed when getting in the way of maximizing production, and the inventory always lasts through the entire product life cycle (as predicted by Proposition 2).

Theorem 2 **(Weak Replicability)**. *Assume* $(r-\rho)\pi < cr + h$. *Define the maximum production interval*

$$\Delta \doteq \begin{cases} -\ln[1-r(\pi-c)/h]/r & \text{if } \rho = 0, \\ -\ln\{1-r[\ln(\rho\pi+h)-\ln(\rho c+h)]/\rho\}/r & \text{if } \rho > 0. \end{cases} \tag{3}$$

Let $\check{I}$ *be such a level of* I_L *that always keeping production and sales on leads to inventory depletion at exactly* $\min(\Delta, T)$. *For* $T \geq \Delta$, *let* $\hat{I}$ *be such a level of* I_L *that always keeping sales on and switching production on at* $T-\Delta$ *leads to inventory depletion at exactly* T. *Let* $\overline{I}$ *be such a level of* I_L *that always keeping production off and sales on leads to inventory depletion at exactly* T. *Clearly* $\check{I} < \hat{I} < \overline{I}$.

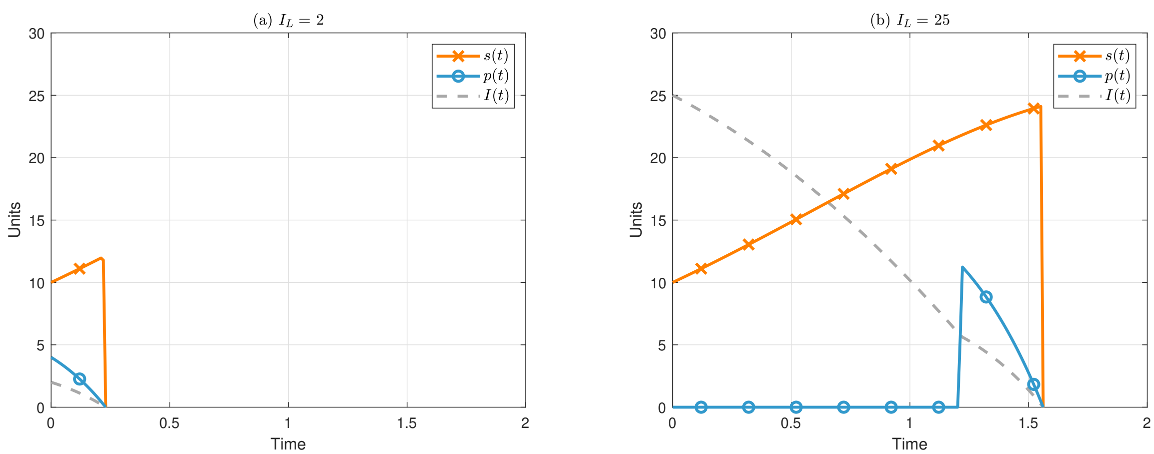
(a) When $I_L < \check{I}$*, it is optimal to always keep production and sales on, and the inventory is depleted before* $\min(\Delta, T)$*.*

(b) When $T \geq \Delta$ *and* $\check{I} < I_L < \hat{I}$*, it is optimal to always keep sales on, and switch and keep production on at such a time that production lasts for exactly* Δ *until inventory depletion.*

(c) When $T \geq \Delta$ *and* $\hat{I} < I_L < \overline{I}$*, or when* $T < \Delta$ *and* $\check{I} < I_L < \overline{I}$*, it is optimal to always keep sales on, and switch production on at such a time later than* $\max(T-\Delta, 0)$ *that production lasts shorter than* $\min(\Delta, T)$ *and the inventory is depleted exactly at* T*.*

(d) When $I_L > \overline{I}$*, it is optimal to always keep production off and sales on, and the inventory is not depleted at* T*.*

Figure 4 Optimal policies of Problem (1) **under Weak Replicability**



Note. These examples are generated with $m = 100$, $\pi = 2.1$, $c = 2$, $h = 0.2$, $r = 2$, $\alpha = 0.1$, $\beta = 1$, $\rho = 0.1$, $w = 0.25$, *and* $T = 10$ *(discretized into 1,000 periods, not fully plotted). Weak Replicability holds:* $(r - \rho)\pi = 3.99 < cr + h = 4.2$*. The theoretical prediction of the maximum production interval is* $\Delta = -\ln\{1 - r[\ln(\rho\pi + h) - \ln(\rho c + h)]/\rho\}/r \approx 0.3405$*. The numerically observed maximum production interval in Case (b) is* 0.3401*.*

Corollary 2. *Assume* $(r - \rho)\pi < cr + h$*. The maximum production interval* Δ *defined in* (3) *increases in the replication rate* r *and the retail price* π*, and decreases in the production cost* c *and inventory holding cost* h*.*

Opposite to Theorem 1, Weak Replicability means that immediate sales are more lucrative than replication and selling later. For this reason, the firm should always keep sales on even if the inventory will be depleted before T with all potential future demand lost, thus the optimal sales policy in Theorem 2. The more interesting question is whether replication may ever be optimal, and if so, when. While inventory that can be sold immediately should not be used for replication, replication using leftover inventory after satisfying all demand may still be profitable given that $\pi > c$. However, since all demand is already being satisfied with existing inventory, additional units produced through replication are only to be sold after existing inventory is depleted, leading to inventory cost being accumulated over the total production interval. This intuition suggests that despite Weak Replicability, replication may still happen toward the end of the product life cycle, but should never last longer than a maximum interval when the accumulated inventory cost exactly offsets the profit margin. Theorem 2 formalizes this intuition and derives the closed-form expression for the maximum production interval Δ. Generally, with moderate levels of launch inventory I_L, production should be switched on at such a time that it will last for exactly Δ before the inventory is depleted at $\tau \leq T$ (Case (b) of Theorem 2, illustrated by Figure 4 (b); the theoretical prediction of Δ is verified numerically with very high precision). Unless, if the launch inventory is sufficiently small, production simply cannot last for Δ even if it is always kept on (Case (a) of Theorem 2, illustrated by Figure 4 (a)). Conversely, in the unlikely extreme cases where the launch inventory is sufficiently large, production does not need to last for Δ (Case (c) of Theorem 2, not illustrated) or does not need to happen at all (Case (d) of Theorem 2, not illustrated) because all demand can already be satisfied. Similar to Theorem 1, Theorem 2 completely characterizes the unique optimal production and sales policies, and the optimal time to switch on production/sales can be easily found through a one-dimensional search with linear computational complexity. Corollary 2 further characterizes the maximum production interval Δ's sensitivity to key model parameters. Since the maximum production interval is determined by the tradeoff between the profit from replication and the cost to hold inventory, it is intuitive that Δ increases in r and π and decreases in c and h.

We also want to distinguish certain cases of Weak Replicability (Theorem 2 (b), (c) and (d), Figure 4 (b)) from certain cases of Strong Replicability (Theorem 1 (b) and (c), Figure 3 (b)). In both groups of cases, sales are always on since the beginning while production is switched on later or kept off. The causes of the similar behavior are however distinct. Under Weak Replicability, sales intrinsically take priority and thus are *always on* while production is generally switched on later, as a result sales *may not last through the product life cycle*. Under Strong Replicability, production intrinsically take priority and thus are generally switched on before sales to make sure that sales *always last through the product life cycle*; the reason in these cases of Strong Replicability why sales are switched on before production is the abundant launch inventory, which means that production does not need to be always on to meet all demand through the product life cycle.

To summarize, under Weak Replicability, the firm should prioritize the present over the future: demand is immediately satisfied as long as inventory is available despite slowing down replication and causing more future demand to be lost, and the inventory generally does not last through the entire product life cycle (as predicted by Proposition 2). Interestingly, replication generally still occurs toward the end of the product life cycle, but will never last longer than a maximum interval.

Our analysis of the base model reveals several crucial insights about the self-replication business model. First, we show that the "keep-or-sell" trade-off boils down to a simple yet informative condition (2) which compares the value of replication against not only the replication cost, but also crucially the inventory holding and discounting costs. In the Strong Replicability regime when (2) holds, a firm should generally hold back sales despite lost demand. This finding is reminiscent of Kumar and Swaminathan (2003)'s finding that a myopic policy of selling as much as possible may not be optimal in capacitated innovation diffusion, and that a build-up policy is optimal with lost sales. The driving forces are however not the same. In their setting, holding back sales serves two purposes: building up inventory, and decelerating demand growth. In our self-replication setting, holding back sales also serves a third purpose: accelerating production. We also show that the optimal time to switch on sales should balance life-cycle demand and supply. These results provide a guideline for producers of self-replicating innovative goods in their capacity ramp-up and ramp-down decisions. On the other hand, in contrast with Kumar and Swaminathan (2003) who show in their setting with lost sales that it is always optimal to build up inventory before sales begin, we show in our self-replication setting that sales should begin right away in the Weak Replicability regime when (2) does not hold. Interestingly, replication may still be optimal near inventory depletion, although it would never last longer than a maximum interval. This non-straightforward insight cautions firms that not all self-replicating innovative goods should prioritize replication over sales, yet even those that do not may still benefit from limited replication near inventory depletion. This insight is unique to self-replication and enriches our understanding of capacitated innovation diffusion.

Lastly, we note that (2) is a local condition independent of the demand process. It suggests that the aforementioned insights should apply to a much wider range of settings beyond our base model, which are explored in Section 5.

4. Numerical studies

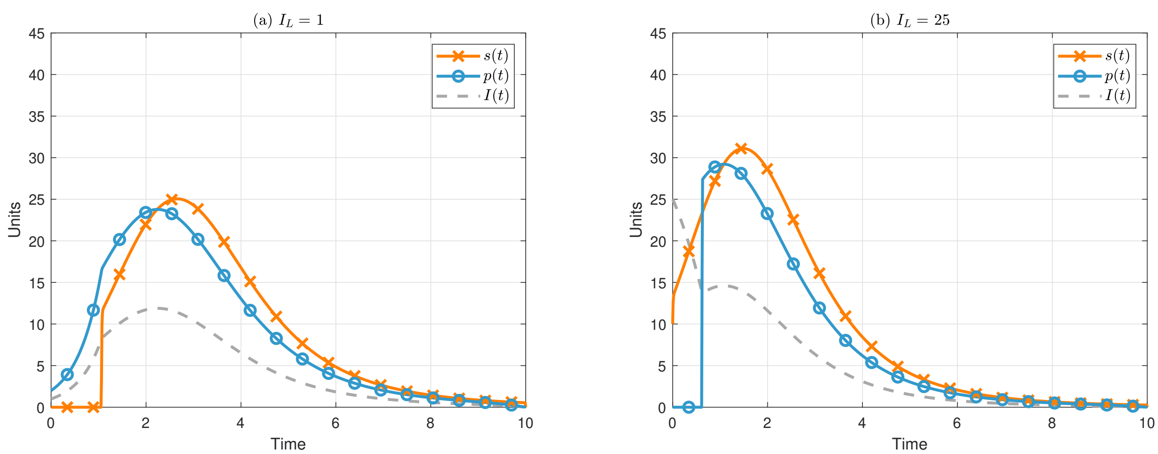
4.1. Effect of sales intensity

In the base model (1) we capture diminishing word-of-mouth by defining the adjusted cumulative sales. In this section we consider a similar effect of sales intensity. The motivation is that aside from the regular word-of-mouth generated by cumulative sales, the sudden increase of a product's sales also tends to create a trend and generate word-of-mouth (especially in the age of social networking). To capture this effect, we modify the classical demand diffusion model by including the sales intensity $s(t)$ alongside the cumulative sales $S(t)$ to generate word-of-mouth:

$$D'(t) = d(t) \doteq [\alpha + (\beta_1 S(t) + \beta_2 s(t))/m][m - D(t)], \tag{4}$$

where β_1 and β_2 are the corresponding coefficients of imitation for the cumulative sales and sales intensity, respectively. With $\beta_2 = 0$ the diffusion model is reduced to that in Ho et al. (2002) and Kumar and Swaminathan (2003). Analyzing this model is challenging, and thus we resort numerical experiments. Figures 5 and 6 respectively illustrate the optimal policies with demand diffusion model (4) under Strong and Weak Replicability. One can see that the optimal policies are qualitatively similar to that of the base model (Figures 3 and 4), which is unsurprising considering that the effect of sales intensity is similar to the effect of diminishing word-of-mouth in that more recent sales generate more word-of-mouth than those from distant past.

Figure 5 Optimal policies with sales intensity under Strong Replicability

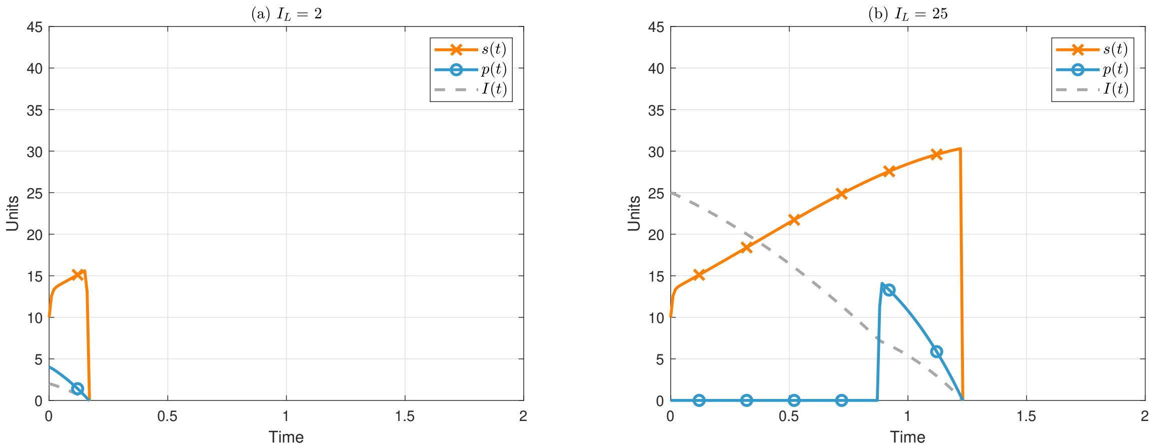


Note. These examples are generated with $m = 100$, $\pi = 2.3$, $c = 2$, $h = 0.2$, $r = 2$, $\alpha = 0.1$, $\beta_1 = 1$, $\beta_2 = 0.25$, $\rho = 0.1$, $w = 0.25$, *and* $T = 10$ *(discretized into 1,000 periods). Strong Replicability holds:* $(r - \rho)\pi = 4.37 > cr + h = 4.2$.

4.2. Sensitivity analysis

Most parameters in our model are exogenous, such as the production cost, the inventory holding cost, the rate of replication, the market size, and the diminishing rate of word-of-mouth. These parameters are difficult to change. By contrast, the coefficient of innovation α and the coefficient

Figure 6 Optimal policies with sales intensity under Weak Replicability



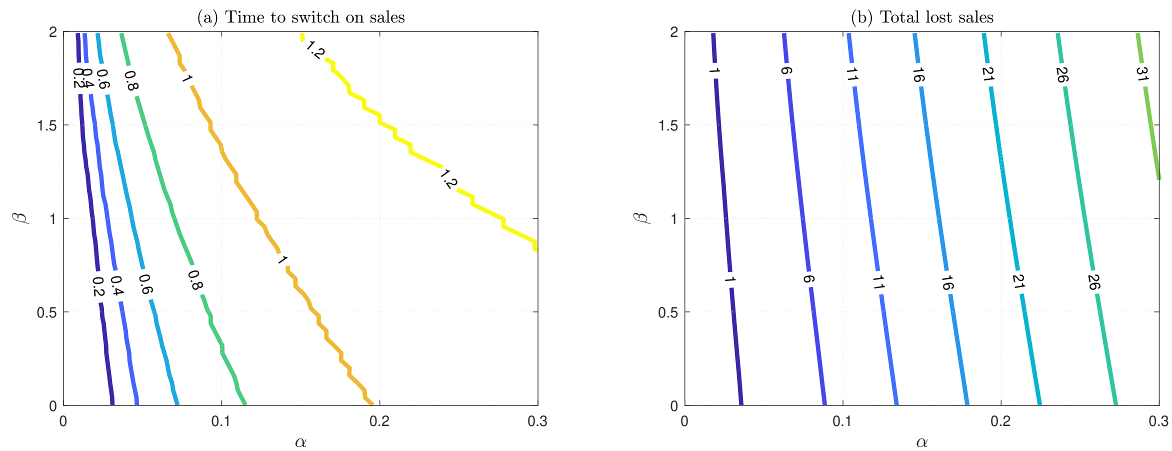
Note. These examples are generated with $m = 100$, $\pi = 2.1$, $c = 2$, $h = 0.2$, $r = 2$, $\alpha = 0.1$, $\beta_1 = 1$, $\beta_2 = 0.25$, $\rho = 0.1$, $w = 0.25$, *and* $T = 10$ *(discretized into 1,000 periods). Weak Replicability holds:* $(r - \rho)\pi = 3.99 < cr + h = 4.2$.

of imitation β are more easily manipulated. In particular, α can be increased through conventional marketing efforts such as running ads and commercials in media channels, whereas β can be increased through social marketing efforts such as rewarding customers who share their experiences in their social circles and referral bonuses. We numerically study the impact of changing α and β to drive insights into marketing innovative self-replicating goods.

Lilien et al. (2017)'s Exhibit 5.8 lists estimated coefficients of innovation and imitation for a range of goods. Two cases most relevant to this paper are camcorders (representing innovative electronics) with $\alpha = 0.044$ and $\beta = 0.304$, and hybrid corns (representing innovative agriculture) with $\alpha = 0.000$ and $\beta = 0.789$. The ranges of the parameters across all products are $0 \leq \alpha \leq 0.265$ and $0 \leq \beta \leq 1.390$. Based on these estimates, we limit our numerical studies within the range of $0 \leq \alpha \leq 0.3$ and $0 \leq \beta \leq 2$ for practical relevance.

We first investigate the case of Strong Replication with limited launch inventory (which is more practically relevant than abundant launch inventory). Recall that in this case production is always on, and sales are switched on when there is enough inventory to last through the entire product life cycle. Figure 7 illustrates the optimal time to switch on sales and total lost sales for varying α and β. The general observation is that the greater demand potential, the *later* sales should be switched on and the *more* sales should be given up. This observation can seem counterintuitive: facing greater demand potential, one may feel instinctive to start meeting demand earlier. However, the nature of self-replicating goods is such that giving up earlier demands is necessary to preserve inventory for greater production to satisfy later demands. The numerical study reaffirms the need to resist the urge to switch on sales too early, which would only lead to more severe shortage later.

Figure 7 Optimal time to switch on sales and total lost sales under Strong Replicability

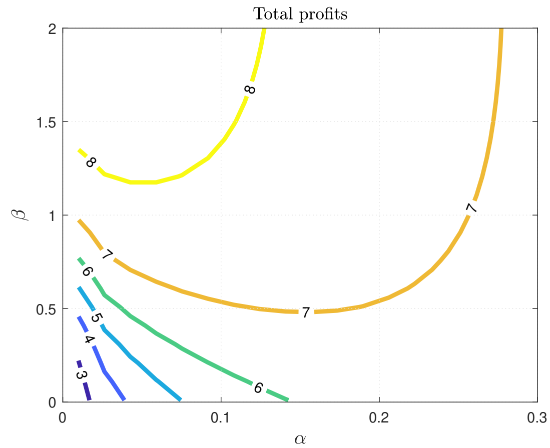


Note. This example is generated with $I_L = 1$, $m = 100$, $\pi = 2.3$, $c = 2$, $h = 0.2$, $r = 2$, $\rho = 0.1$, $w = 0.25$, and $T = 10$ (discretized into 1,000 periods). Strong Replicability holds.

Figure 8, which illustrates the optimal total profits for varying α and β under Strong Replicability, shows a more interesting pattern. Increasing β through social marketing efforts unsurprisingly improves total profits. However, increasing α through conventional marketing efforts, especially when combined with high β values, may backfire and *reduce* total profits. The contrast highlights fundamentally different demand growth patterns driven by conventional and social marketing efforts.

Consider the innovation diffusion equation $D'(t) = [\alpha + \beta S(t)/m][m - D(t)]$. First we consider $\alpha \gg \beta$ so that the demand is mostly driven by conventional marketing efforts. In this case, the demand growth is the fastest at product launch and follows an exponential decay. Then we consider $\alpha \ll \beta$ so that the demand is mostly driven by social marketing efforts. In this case, the demand growth is minimum before sales are switched on (because $S(t) \equiv 0$), and after sales are switched on the demand growth initially increases exponentially and then flattens out before finally decaying exponentially, forming the classic S-shaped innovation diffusion curve. On the other hand, recall that the self-replication production mode is characterized by an initial supply shortage period followed by an exponential supply growth. It is easy to see that the early gradual exponential demand growth of social marketing is a good match with the early gradual exponential supply growth of self-replication. It means that even with aggressive social marketing efforts, the lost sales

Figure 8 Optimal total profits under Strong Replicability



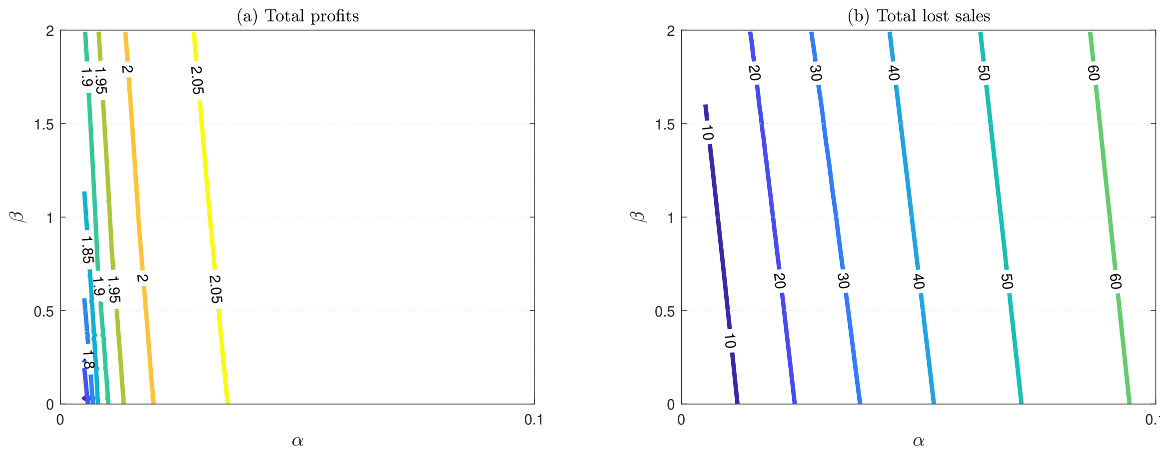
Note. This example is generated with $I_L = 1,\ m = 100,\ \pi = 2.3,\ c = 2,\ h = 0.2,\ r = 2,\ \rho = 0.1,\ w = 0.25,$ *and* $T = 10$ *(discretized into 1,000 periods). Strong Replicability holds.*

are relatively minor. On the other hand, aggressive conventional marketing efforts mean that the highest initial demand growth is met with the most severe initial supply shortage, and significant sales lost will ensue. This insight explains why in Figure 8 large α values can backfire and reduce total profits whereas large β values do not. The explanation is also consistent with Figure 7 (b) which shows that the total lost sales increase dramatically for larger α values, but only modestly for larger β values.

Next we plot the optimal total profits and lost sales for varying α and β under Weak Replicability. Figure 9(b) resembles Figure 7(b). Figure 9(a) however differs from Figure 8 in that under Weak Replication increasing α no longer reduces total profits. The reason is that in this case the firm prioritizes sales over replication. As a result, the inventory tends to be depleted before neither the demand nor the production enters the exponential growth stage, meaning that the effect which causes conventional marketing efforts to backfire under Strong Replicability is absent under Weak Replicability.

To summarize, the sensitivity analysis reveals an interesting insight: under Strong Replicability, social marketing which boosts the word-of-mouth effect is a particularly well-suited marketing strategy for self-replicating innovative goods, while conventional marketing strategies should be used conservatively and cautiously. The insight matches our key motivating examples—the Prusa 3D printers and the Glass Gem corns—which are self-replicating innovative goods promoted almost exclusively through social marketing strategies.

Figure 9 Total profits and total lost sales under Weak Replicability



Note. This example is generated with $I_L = 1,\ m = 100,\ \pi = 2.1,\ c = 2,\ h = 0.2,\ r = 2,\ \rho = 0.1,\ w = 0.25,$ *and* $T = 10$ *(discretized into 1,000 periods). Weak Replicability holds.*

5. Extensions

In this section, we show how our main insights from the base model remain robust and help characterize optimal policies in a range of extensions. These extensions are technically more complicated to analyze. For brevity, we will forgo full analyses and focus on how the insights from the base model can inform optimal policies in these extensions.

5.1. Backlogged demand

Our base model assumes lost sales. In line with Kumar and Swaminathan (2003) and Shen et al. (2011, 2014), we consider backlogged demand in this section. In particular, we adopt the partial-backlog model of Shen et al. (2014) and assume that a $\gamma \in [0,1]$ fraction of unmet demand is backlogged and $1-\gamma$ is lost. Backlogged demand incurs backlog cost b per unit good-time. Revenue from backlogged demand is earned upon delivery of the good. We assume that orders are filled first-in-first-out, namely backlogged demand is satisfied before new demand, following prevalent business practices including at Prusa Research. To model such a system, we need an additional state variable $B(t)$ to capture the cumulative backlog at time t. Let $\mathbb{1}_{B(t)} \doteq 1$ indicate $B(t) > 0$, and $\mathbb{1}_{B(t)} \doteq 0$ indicate otherwise.

Note that for our self-replicating product, once the inventory is depleted, no further production or sales is possible. In such a case the firm should announce the end of life for the product and no longer be penalized for the backlog. In other words, allowing a backlog necessitates formally cutting short the product life cycle at inventory depletion. (It was not necessary in the base model because without a backlog the firm incurs no cost after inventory depletion.) We present the modified model below, where τ represents the end of the product life cycle. Note that the definition of τ depends on the property that once $I(t)$ reaches zero it can never become positive again. The fact that the product life cycle ends whenever the inventory is depleted also eliminates the need for both the pure state constraint $I(t) \geq 0$ and the terminal condition $I(T) \geq 0$.

$$\max_{p(\cdot),s(\cdot)} \int_0^\tau [\pi s(t) - cp(t) - hI(t) - bB(t)]e^{-\rho t}\,\mathrm{d}t \tag{5}$$

$$\begin{aligned} \text{s.t.}\quad & D'(t) = d(t) \doteq [\alpha + \beta S(t)/m][m - D(t)],\ S'(t) = s(t) - wS(t),\ I'(t) = p(t) - s(t), \\ & B'(t) = \gamma[d(t) - s(t)] - \mathbb{1}_{B(t)}(1-\gamma)s(t), \\ & 0 \leq p(t) \leq rI(t),\ s(t) \geq 0,\ s(t)[1 - \mathbb{1}_{B(t)}] \leq d(t), \\ & D(0) = S(0) = B(0) = 0,\ I(0) = I_L > 0,\ \tau \doteq \min(T, \sup\{t | I(t) > 0\}). \end{aligned}$$

Note the differential equation governing backlogged demand: $B'(t) = \gamma[d(t) - s(t)] - \mathbb{1}_{B(t)}(1-\gamma)s(t)$. When there is no backlog ($B(t) = 0$), the equation becomes $B'(t) = \gamma[d(t) - s(t)]$ because any unmet demand is partially backlogged. (In this case the constraint $s(t)[1 - \mathbb{1}_{B(t)}] \leq d(t)$ becomes $s(t) \leq d(t)$ which implies $B(t) \geq 0$.) When there is backlog ($B(t) > 0$), the equation becomes $B'(t) = \gamma d(t) - s(t)$. Recall that backlogged demand is satisfied before new demand. Therefore, all sales go toward reducing backlog while a γ fraction of new demand adds to backlog.

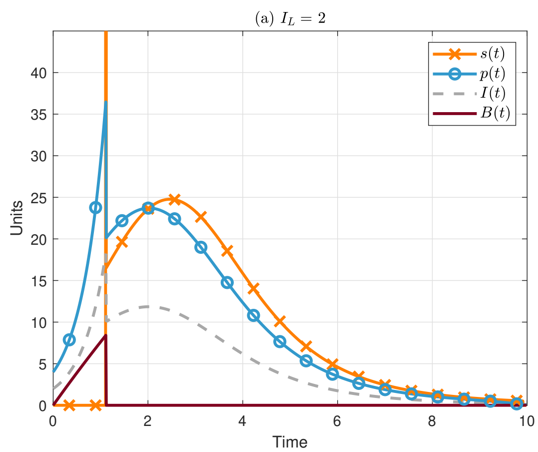
Strong Replicability from Section 3 states that the value of replication outweighs the production, inventory holding and discounting costs. In the backlog model, there is an additional backlog cost to be weighed against, and Strong Replicability is defined by

$$\pi(1 + r\delta) - cr\delta - h\delta - \rho\pi\delta - \gamma b\delta > \pi \Rightarrow (r - \rho)\pi > cr + h + \gamma b. \tag{6}$$

Therefore, with Condition (2) replaced by (6), Theorems 1 and 2 fully characterize the optimal policies with partial backlog. This observation contrasts Kumar and Swaminathan (2003) who show in their setting that a build-up policy is not always optimal with backlog.

Figure 10 illustrates Theorem 1's Case (a) with partial backlog, and compares interestingly with Figure 3 (a). The backlog model's optimal policies share a similar overall structure: sales are held back in the beginning to maximize production; once enough inventory is built so that all future demand can be satisfied, sales are switched on and the inventory is depleted at exactly T. However, in Figure 10, as sales are held back in the beginning, a backlog is accumulated. As a result, a larger inventory needs to be built compared with Figure 3 before sales can be switched on, at which time a chunk of the inventory is used to instantly satisfy the entire backlog, leading to a drop of $I(t)$ in the amount of $B(t)$, a drop of $B(t)$ to zero, and an infinite spike of $s(t)$—a policy known technically as *impulse control* (Sethi 2019, p. 19). (In discretized numerical experiments the spike of $s(t)$ is large but finite; it is truncated in Figure 10.) No demand is ever backlogged past this point. We omit illustrating Theorem 1's Cases (b) and (c) and Theorem 2, because with abundant launch inventory, and under Weak Replicability where immediate sales have priority over production, no demand will ever be backlogged.

Figure 10 Optimal policies of Problem (5) **under Strong Replicability**



Note. This example is generated with $m = 100$, $\pi = 2.5$, $c = 2$, $h = 0.2$, $b = 0.5$, $r = 2$, $\alpha = 0.1$, $\beta = 1$, $\rho = 0.1$, $\gamma = 0.8$, $w = 0.25$, *and* $T = 10$ *(discretized into 1,000 periods). Strong Replicability holds:* $(r - \rho)\pi = 4.75 > cr + h + \gamma b = 4.6$.

The analysis and numerical example show that under Strong Replicability, when demand may be backlogged, the firm should resist the urge to reduce the ever-growing backlog, and instead focus on maximizing replication and only satisfy all backlog at once when enough inventory (production capacity) is built for the product's remaining life cycle.

5.2. Liquidity constraint

In our base model, to focus on the most fundamental insights of self-replicating goods, we only considered the self-replication production constraint. Indeed, a corporate giant such as Monsanto producing an innovative agricultural good through self-replication faces few other production constraint. Yet for a startup with limited financial resources such as Prusa Research, the firm faces the liquidity constraint, namely that it need to carefully manage their cash flows to stay solvent at all times.

In this section, we modify the base model (1) to incorporate the liquidity constraint as follows. Suppose at time 0 the firm has an initial cash level (or a line of credit) L_0. Production and inventory costs deplete the cash deposit whereas sales replenish it, leading to cash level $L(t)$ at time t. We require that at any time the cash level cannot be negative.[3] The problem formulation is

$$
\begin{aligned}
&\max_{p(t),s(t)} \int_0^T [\pi s(t) - cp(t) - hI(t)]e^{-\rho t}\,\mathrm{d}t && (7)\\
&\text{s.t.}\quad D'(t) = d(t) \doteq [\alpha + \beta S(t)/m][m - D(t)],\\
&\qquad S'(t) = s(t) - wS(t),\ I'(t) = p(t) - s(t),\ L'(t) = \pi s(t) - cp(t) - hI(t),\\
&\qquad 0 \le p(t) \le rI(t),\ 0 \le s(t) \le d(t),\ L(t) \ge 0,\ I(T) \ge 0,\\
&\qquad D(0) = S(0) = 0,\ I(0) = I_L > 0,\ L(0) = L_0 \ge 0.
\end{aligned}
$$

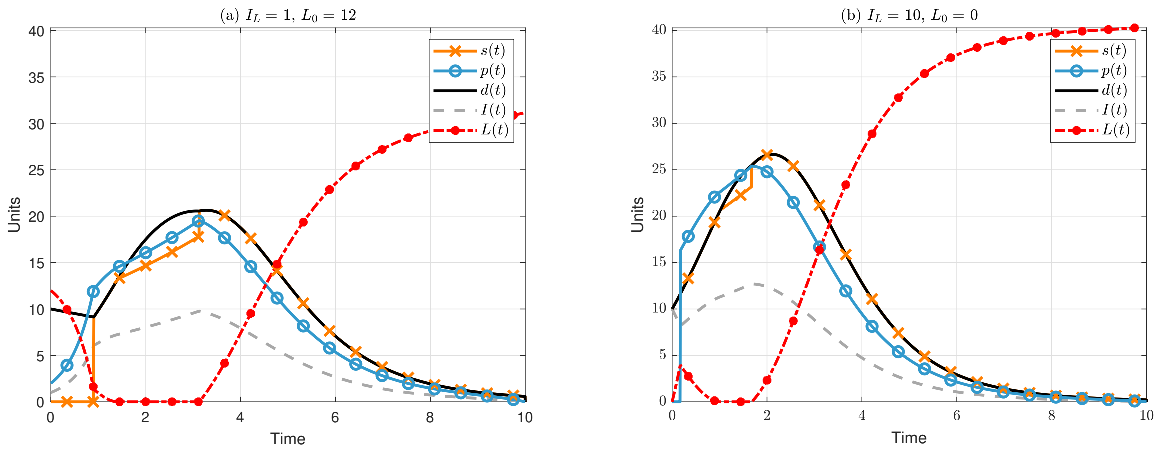
The liquidity constraint is irrelevant in most cases under Weak Replicability because sales take priority over production. Unless there is a huge launch inventory such that the inventory cost drains cash faster than sales revenue replenishes it, the liquidity constraint will never become binding. Henceforth, we focus on Strong Replicability where replication takes priority over sales and the cash level may dip significantly during the early production ramp-up stage. Figure 11 illustrates the optimal policies under Strong Replicability and the liquidity constraint. The two cases in Figure 11 respectively correspond to those of Figure 3, between which a comparison can be drawn.

One can see that the basic patterns of Figure 3 are preserved in Figure 11: in Case (a) with a low launch inventory, sales are initially held back to maximize production; in Case (b) with a higher launch inventory, sales are always on and production does not need to be switched on initially; and in both cases the inventory is depleted at T. This shows that the liquidity constraint does not alter the fundamental Strong Replicability insight and the basic structural result of Theorem 1.

There is however an astonishing observation from Figure 11: when the liquidity constraint is binding (the firm runs dry of cash), the optimal policy is to *scale back sales* to a level that provides just enough revenue to cover the cost of inventory and production, as such the cash position remains at zero for a period. This observation is counter-intuitive. When a firm runs out of cash, common sense may dictate that sales be maximized so as to resolve the cash crisis. Nonetheless, this observation can actually be explained with the Strong Replicability insight.

[3] More generally, the liquidity constraint should be modeled as a terminating condition, namely the product life cycle is cut short whenever the cash position reaches zero; however note that if at this time there is remaining inventory, the firm can always cease production and sell off the inventory to improve the objective while staying solvent, meaning that it is never optimal to cut short the product life cycle because of the liquidity constraint. Therefore our constraint of requiring the cash level to be non-negative is without loss of generality.

Figure 11 Optimal policies of Problem (7) under Strong Replicability



Note. These examples are generated with $m = 100$, $\pi = 2.3$, $c = 2$, $h = 0.2$, $r = 2$, $\alpha = 0.1$, $\beta = 1$, $\rho = 0.1$, $w = 0.25$, *and* $T = 10$ *(discretized into 1,000 periods). Strong Replicability holds:* $(r - \rho)\pi = 4.37 > cr + h = 4.2$.

Recall that Strong Replicability requires production and inventory-building to be maximized until enough inventory is built. When a firm runs out of cash, production is constrained and needs to be financed by sales. However, sales also drain inventory which tightens the replication constraint on production. The optimal policy to maximize production, therefore, should be to relieve the cash shortage without limiting replication—in other words, to sell just enough inventory to finance production and keep the cash level at zero, until enough inventory is built for the remainder of the product life cycle when sales can be thereafter maximized, as is observed in Figure 11. This insight may prove instructive for innovative startups such as Prusa Research.

5.3. Stochastic innovation diffusion

The majority of the capacitated innovation diffusion literature, as well as our base model, assume no uncertainty for analytical tractability. However, real-life markets can be highly uncertain, especially those for innovative goods which motivate this paper. Among the three market parameters in our models, arguably, the market size m and the coefficient of innovation α are relatively easy to estimate through established marketing techniques, whereas the coefficient of imitation β, which is deeply rooted in human behavior, may be the most elusive parameter to estimate. In this section, we investigate a model where β is a random variable.

We adopt the approach of Kanniainen et al. (2011) and Shen et al. (2014) and multiply β by a random noise $X(t) \sim \text{Uniform}[0, 2]$ which is independently and identically distributed over time.

In other words, the realized coefficient of imitation may be as low as 0 or as high as 2β, with the average being β. The modified problem formulation is

$$\begin{aligned}
&\max_{p(t),s(t)} \int_0^T [\pi s(t) - cp(t) - hI(t)]e^{-\rho t}\,\mathrm{d}t \\
&\text{s.t.}\quad D'(t) = d(t) \doteq [\alpha + \beta X(t)S(t)/m][m - D(t)],\ S'(t) = s(t) - wS(t),\ I'(t) = p(t) - s(t), \\
&\qquad 0 \le p(t) \le rI(t),\ 0 \le s(t) \le d(t),\ I(T) \ge 0,\ X(t) \sim \text{Uniform}[0,2], \\
&\qquad D(0) = S(0) = 0,\ I(0) = I_L > 0.
\end{aligned}$$

We note that this formulation is in fact technically ill-defined. The demand rate $d(t)$ cannot be adapted to a finite-variation process and is thus not a semimartingale. This means that $d(t)$ is non-integrable in common (e.g., Itô or Stratonovich) stochastic calculus variations. However, once we discretize the decision horizon to numerically evaluate the model, the problem disappears. Because of the model's non-integrality, its behavior and the heuristics' performances do not converge as the interval approaches zero. As such, we fix the discretization period to be 1,000 and make horizontal comparisons of different heuristics.

We will focus on the most relevant case—Strong Replicability with a small launch inventory—namely Case (a) of Theorem 1 and Figure 3, where the key decision is when to switch on sales. We will restrict our numerical considerations among bang-bang policies, namely to satisfy either all or no demand. Under stochastic innovation diffusion, finding the optimal policy requires solving a complex stochastic dynamic program (SDP) which is computationally challenging. Instead, we consider an oracle—who knows the entire realized coefficient of imitation trajectory and chooses the optimal policy based on a deterministic model—as a benchmark, which is much easier to compute. Clearly, such an oracle will always outperform the optimal SDP solution, and a heuristic's performance loss against the oracle (known as the regret) is an upper bound of its performance loss against the optimal SDP solution. Note that the oracle does not exist in real life and is only conceived for benchmarking purposes.

We propose two heuristics. The first, which we refer to as the deterministic heuristic or H1, is that the firm simply ignores all uncertainty (replacing all $X(t)$ with 1) at time 0, finds the "optimal" time to switch on sales based on the deterministic model, and executes the policy in the uncertain environment. The computational complexity of H1 is the same as Theorem 1, namely linear. A potential issue with H1 is that it does not dynamically readjust the strategy. Therefore, we propose another heuristic which we refer to as the rolling deterministic heuristic or H2. The heuristic is such that in *every* period the firm takes the realized cumulative demand and sales as given, ignores all future uncertainty (replacing all future $X(t)$ with 1), finds the "optimal" time to switch on sales based on the deterministic model for the remaining periods, and executes the policy for the current period. In other words, H2 involves running H1 in every period for all remaining periods. As such H2 has a quadratic computational complexity. We extensively evaluate the average performance losses (regrets) of H1 and H2 against the oracle over 1,000 simulation trials for a wide range of parameter combinations, and have observed highly consistent patterns. The percentage regrets range from insignificant (e.g., 3%) to very significant (e.g., 50%) depending on parameters. Percentage regrets tend to be high when the profit margin is low (e.g., with a smaller π), which is understandable because the oracle's profit, the basis for the percentage regret, becomes smaller. By comparison, the absolute regret varies much less when the profit margin varies. Some representative examples are provided in Table 1.

Table 1 Average percentage regrets of H1 and H2 against the oracle for varying r and π

	$r=2$						$r=3$					
π	2.3	4.3	6.3	8.3	10.3	12.3	2.3	4.3	6.3	8.3	10.3	12.3
H1 (%)	13.87	4.88	3.87	3.49	3.29	3.17	17.26	9.39	6.91	6.02	5.57	5.29
H2 (%)	19.98	6.97	5.52	4.96	4.66	4.49	47.10	11.92	8.65	7.48	6.85	6.46
H2/H1	1.44	1.43	1.43	1.42	1.42	1.42	2.73	1.27	1.25	1.24	1.23	1.22

Note. These examples are generated with $m=100$, $c=2$, $h=0.2$, $\alpha=0.1$, $\beta=5$, $\rho=0.1$, $T=1$ *(discretized into 1,000 periods),* $w=0.25$, *and* $I_L=1$ *over 1,000 simulation trials. The horizon is discretized into 100 periods. Strong Replicability holds for all cases.*

The most surprising observation however is that H2 consistently performs worse than H1 with at least 20% larger regrets. Why does the static heuristic H1 outperform the dynamic heuristic H2? Recall that H1 makes a single decision in the first period ignoring all future uncertainty. H2 seemingly remedies H1's static nature by repeatedly performing H1 in each period based on the realized cumulative demand and sales. However, despite being apparently dynamic, H2 is not forward-looking, in the sense that the decision at each period fails to account for future decisions. For example, H2's first-period decision is identical to H1, thus also as bad as H1. In future periods, H2 continues to introduce noises based on the realized uncertainty and past decisions. As a result, the dynamic H2 is actually noisier than the static H1 and is consistently outperformed by the latter. The numerical experiments show that practitioners can simply use the static deterministic heuristic H1 to address uncertainty, which performs particularly well for high-margin goods, whereas dynamically adjusting the heuristic without being forward-looking actually backfires and worsens the performance. The usefulness of the deterministic heuristic is consistent with the findings of Shen et al. (2014).

5.4. Launch inventory decision

The base model assumes an exogenous launch inventory I_L. This assumption is equivalent to assuming that demand diffusion begins at product launch (time 0)—an appropriate assumption when the firm is not the only player in the market, such as in the case of Prusa Research. To see the equivalence, consider the following scenario. Assume the firm has inventory I_L at time 0, and recall that the only way to increase inventory is by self-replication. Should the firm want to have more launch inventory, it needs to postpone the product launch to allow replication. However, if the firm is not the only player in the market, its potential customers will still be gradually lost to competing firms, and the product will still be obsolete at time T, despite the postponed launch. Such a scenario is no different from the firm holding back sales and maximizing production after a product launch at time 0 as in Theorem 1's Case (a). In this sense, the firm has no real ability to choose the product launch time or, equivalently, the launch inventory.

However, if the innovative product faces no competition and demand diffusion does not begin until the firm launches the product, such as the case of the Glass Gem corn, the firm will have the option to postpone the product launch and produce enough launch inventory, and the product's life cycle can still last for T. The cost of doing so is that revenue will be postponed and discounted. Due to the self-replicating nature of the good, we still need to assume an initial "seed" inventory $I_0 > 0$ from the research and development process. Let t_L denote the product launch time. The problem formulation is

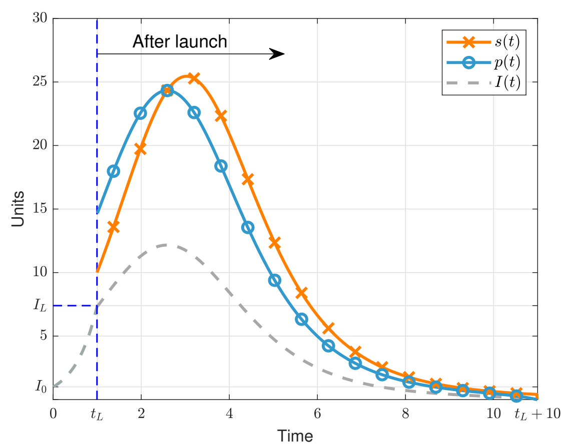
$$\begin{aligned}
&\max_{t_L \geq 0} \left[-I_0(cr+h)\int_0^{t_L} e^{(r-\rho)t}\,\mathrm{d}t + e^{-\rho t_L} \max_{p(t),s(t)} \int_0^T [\pi s(t) - cp(t) - hI(t)]e^{-\rho t}\,\mathrm{d}t \right] \\
&\text{s.t.} \quad D'(t) = d(t) \doteq [\alpha + \beta S(t)/m][m - D(t)],\ S'(t) = s(t) - wS(t),\ I'(t) = p(t) - s(t), \\
&\qquad 0 \leq p(t) \leq rI(t),\ 0 \leq s(t) \leq d(t),\ I(T) \geq 0, \\
&\qquad D(0) = S(0) = 0,\ I(0) = I_0 e^{rt_L} > 0.
\end{aligned}$$

In this formulation, the term $-I_0(cr+h)\int_0^{t_L} e^{(r-\rho)t}\,\mathrm{d}t$ is the production and inventory cost of the launch inventory preparation period t_L. The subsequent life-cycle value is discounted by $e^{-\rho t_L}$ due to the postponement. In exchange, the launch inventory is increased to $I_0 e^{rt_L}$.

It is clear that under Weak Replicability where replication is not a priority, or under Strong Replicability but with a sufficiently large seed inventory I_0, the firm should immediately launch the product. Let us focus on the case of Strong Replicability and a small seed inventory I_0; more specifically, Theorem 1's Case (a) should we force $t_L = 0$. In this case, should the firm postpone the product launch to build more launch inventory, and if yes, by how long? The answer is provided by the following proposition and illustrated in Figure 12.

PROPOSITION 3 **(Launch inventory)**. *Assume $(r-\rho)\pi > cr+h$, and $I_0 < \underline{I}$ which is defined as such a level of I_L that always keeping production and sales on leads to inventory depletion at exactly T after the product launch. It is optimal to postpone the product launch by $t_L = (\ln \underline{I} - \ln I_0)/r$ until the launch inventory is $I_L = \underline{I}$, and keep both production and sales on after launch until inventory depletion at exactly $t_L + T$.*

Figure 12 Optimal launch inventory under Strong Replicability



Note. This example is generated with $I_0 = 1,\ m = 100,\ \pi = 2.3,\ c = 2,\ h = 0.2,\ r = 2,\ \alpha = 0.1,\ \beta = 1,\ \rho = 0.1,\ w = 0.25,$ and $T = 10$ (discretized into 1,000 periods). Strong Replicability holds: $(r-\rho)\pi = 4.37 > cr + h = 4.2$. The optimal launch time and inventory are $t_L = 1.0$ and $I_L = 7.39$, respectively.

The proposition states that under Strong Replicability and with a small seed inventory, it is optimal to postpone the product launch so that sales are on right after the launch and no demand is lost. While it is intuitive to avoid losing demand, it is non-trivial that doing so is justified despite delayed and thus discounted revenue. The explanation is again tied to Strong Replicability, which states that the benefit of growing inventory outweighs the production, inventory holding and discounting cost. This result is potentially helpful for firms determining product launch times for their exclusive innovations, such as Native Seeds/SEARCH which regretted that they "did not grow out enough [Glass Gem corns]" before launching the product (Business Insider 2012).

5.5. Exogenous demand

Our research problem is inspired by self-replicating innovative goods such as the Original Prusa i3 3D printers and the Glass Gem corns. The challenge of growing a small launch inventory to meet future demand is particularly relevant in the early stages of innovative goods; for mature self-replicating goods with stable demands such as regular wheat, there is typically already sufficient inventory to sustain replication to meet the demands. This is why we adopt a demand diffusion model appropriate for early stages of innovative goods. However, our key finding of the Strong Replicability condition (2) contains no demand model parameters, suggesting that the self-replicating insights technically also apply to exogenous (non-diffusion) demands. In this section we analyze self-replicating goods facing exogenous demands, the simplest of which is stable demand. Consider the following formulation with a stable demand stream d:

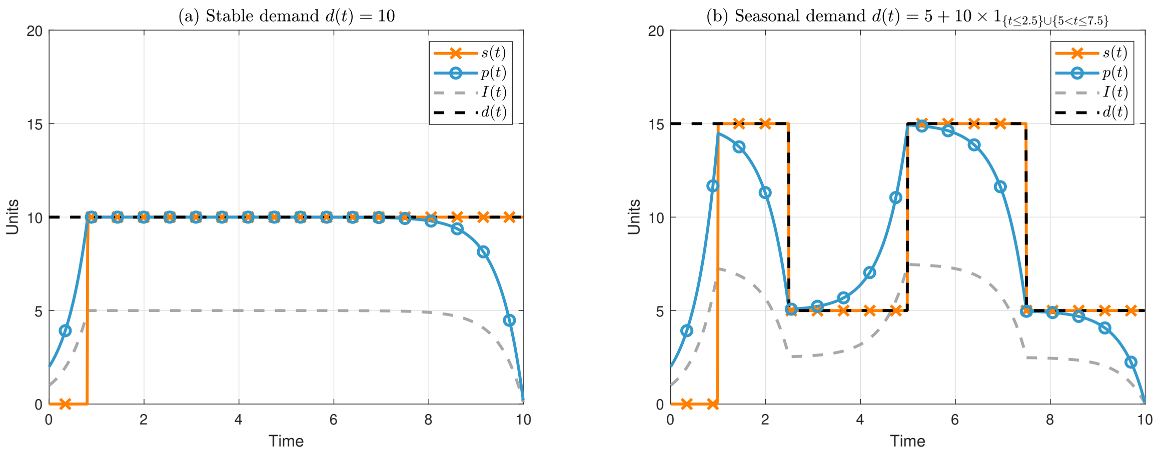
$$\max_{p(t),s(t)} \int_0^T [\pi s(t) - cp(t) - hI(t)]e^{-\rho t}\,\mathrm{d}t$$
$$\text{s.t. } I'(t) = p(t) - s(t),\ 0 \le p(t) \le rI(t),\ 0 \le s(t) \le d,\ I(T) \ge 0,$$
$$I(0) = I_L > 0.$$

We focus on the most interesting case of Strong Replicability with insufficient launch inventory to meet the stable demand through self-replication (the insights for other cases similarly carry over). The next proposition presents the optimal policies.

Proposition 4 **(Stable demand)**. *Assume $(r-\rho)\pi > cr + h$ and $rI_L < d$. It is optimal to always keep production on, and switch and keep sales on at such a time $t_s < (\ln d - \ln I_L - \ln r)/r$ that the inventory is depleted exactly at T; $\lim_{T\to\infty} t_s = (\ln d - \ln I_L - \ln r)/r$, $\lim_{T\to\infty} I(t_s) = d/r$.*

Figure 13 (a) illustrates Proposition 4. Facing stable demands, the firm should build (nearly) enough inventory to meet all demand all demand through replication while sustaining the inventory level, until the end-of-life-cycle sell-off. In Figure 13 (b), we numerically evaluate the optimal policies for a highly seasonal exogenous demand pattern. One can see that the basic Strong Replicability insights still apply: production is always on, sales are initially held back and then switched and kept on, and the inventory runs out exactly at the end of the product life cycle. Admittedly, it is less likely in practice that an innovative self-replicating good with a small launch inventory would already face a stable or exogenous demand pattern. Nevertheless, the fact that our main insights carry over and help characterize optimal policies with exogenous demands speaks to their robustness and usefulness.

Figure 13 Optimal policies with exogenous demands under Strong Replicability



Note. This example is generated with $I_0 = 1,\ m = 100,\ \pi = 2.3,\ c = 2,\ h = 0.2,\ r = 2,\ \alpha = 0.1,\ \beta = 1,\ \rho = 0.1,\ w = 0.25,$ *and* $T = 10$ *(discretized into 1,000 periods). Strong Replicability holds:* $(r - \rho)\pi = 4.37 > cr + h = 4.2$.

6. Conclusion

Inspired by self-replicating 3D printers and innovative agricultural and husbandry goods, we study optimal production and sales policies for a manufacturer of self-replicating innovative goods. Such a production system is operationally fascinating: the inventory serves as the production facility and limits the production capacity, and the firm faces the unique "keep-or-sell" trade-off for each newly-produced unit—should it be sold to satisfy demand and stimulate future demand, or should it be added to inventory to increase production capacity?

We adopt the continuous-time optimal control framework and marry a self-replication model on the production side to the canonical innovation diffusion model on the demand side. By analyzing the model, we identify two regimes: the Strong Replicability regime where production takes priority over sales as long as the produced goods will eventually be sold, and the Weak Replicability regime where sales have priority over production and are never held back. Following these insights, we fully characterize their distinct optimal production and sales policies. Generally speaking, under Strong Replicability, sales are initially held back to maximize production and the inventory is depleted exactly at the end of the product life cycle, whereas under Weak Replicability, sales are never held back, the inventory may be depleted within the product life cycle, and production only takes place near inventory depletion.

These insights prove robust and helpful in several extensions, including backlogged demand, liquidity constraints, launch inventory decision, and exogenous demand. We also numerically evaluate the performance of deterministic heuristics under stochastic demand diffusion, and under a seasonal exogenous demand pattern. The insights and policies derived in this paper are potentially instructive for manufacturers of innovative self-replicating goods. A firm should first evaluate if it resides in the Strong or Weak Replicability regime, and then apply the respective priorities (production or sales) to derive appropriate policies. The models can provide further quantitative support. We also show that social marketing strategies are particularly well-suited for self-replicating innovative goods under Strong Replicability.

Our model assumes a fixed retail price. The assumption is partly motivated by the fact that the Prusa 3D printers are sold at fixed retail prices during their primary life cycles. In general, pricing can be a powerful lever to balance demand and supply in innovation diffusion as shown by Shen et al. (2011, 2014). On the other hand, dynamic pricing could lead to strategic consumer behavior (e.g., speculative waiting) and actually worsen supply-demand imbalance, which may explain Prusa Research's pricing policy. A promising future research direction is to allow dynamic pricing and account for resulting strategic customer behavior. It would require a substantial development of the model and methodology, but will likely also reveal substantial insights.

References

3DPI. 2016. Interview with Josef Prusa, CEO and Founder of Prusa Research. `https://3dprintingindustry.com/news/interview-josef-prusa-ceo-founder-prusa-research-81650/`. Retrieved October 19, 2020.

Additive Manufacturing. 2018. What is additive manufacturing? `http://additivemanufacturing.com/basics/`. Retrieved October 19, 2020.

Arbabian, Mohammad E, Michael R Wagner. 2020. The impact of 3D printing on manufacturer–retailer supply chains. *European Journal of Operational Research* **285**(2) 538–552.

Bass, Frank M. 1969. A new product growth for model consumer durables. *Management science* **15**(5) 215–227.

Berling, Peter, Victor Martínez-de Albéniz. 2011. Optimal inventory policies when purchase price and demand are stochastic. *Operations research* **59**(1) 109–124.

Bilginer, Özlem, Feryal Erhun. 2015. Production and sales planning in capacitated new product introductions. *Production and Operations Management* **24**(1) 42–53.

Bloomberg. 2017. This company's robots are making everything—and reshaping the world. https://www.bloomberg.com/news/features/2017-10-18/this-company-s-robots-are-making-everything-and-reshaping-the-world. Retrieved October 19, 2020.

Bloomberg. 2018. The newest breeds strut their fluff at the Westminster Dog Show. https://www.bloomberg.com/news/photo-essays/2018-02-14/the-newest-breeds-strut-their-fluff-at-the-westminster-dog-show. Retrieved October 19, 2020.

Business Insider. 2012. This summer get ready for the prettiest type of corn ever. https://www.businessinsider.com/glass-gem-corn-2012-5. Retrieved October 19, 2020.

Business Insider. 2014. Jefferies: 3D printing is about to become a big market story. https://www.businessinsider.com/3d-printer-stocks-2014-2. Retrieved October 19, 2020.

Caro, Felipe, Victor Martínez-de Albéniz. 2010. The impact of quick response in inventory-based competition. *Manufacturing & Service Operations Management* **12**(3) 409–429.

Chen, Li, Yao Cui, Hau L Lee. 2017. Retailing with 3D printing. *Available at SSRN 3031566* .

CNN. 2017. This robot can 3D print a building in 14 hours. https://money.cnn.com/2017/05/02/technology/3d-printed-building-mit/index.html. Retrieved October 19, 2020.

Digital Social Innovation. 2017. Prusa Research: The maker company open sourcing the 3D printing industry. https://digitalsocial.eu/case-study/24/prusa-research. Retrieved October 19, 2020.

Dong, Lingxiu, Duo Shi, Fuqiang Zhang. 2016. 3D printing vs. traditional flexible technology: Implications for manufacturing strategy. *Available at SSRN 2847731* .

Feng, Qi, Lauren Xiaoyuan Lu. 2013. The role of contract negotiation and industry structure in production outsourcing. *Production and Operations Management* **22**(5) 1299–1319.

Feng, Qi, Sirong Luo, J George Shanthikumar. 2020. Integrating dynamic pricing with inventory decisions under lost sales. *Management Science* **66**(5) 2232–2247.

Feng, Qi, Sirong Luo, Dan Zhang. 2013. Dynamic inventory–pricing control under backorder: Demand estimation and policy optimization. *Manufacturing & Service Operations Management* **16**(1) 149–160.

Forbes. 2018. 30 Under 30 - Europe - Technology 2018: Josef Prusa. https://www.forbes.com/profile/josef-prusa/#15b2881825ff. Retrieved October 19, 2020.

Gordon, H Scott. 1954. The economic theory of a common-property resource: the fishery. *Classic Papers in Natural Resource Economics*. Palgrave Macmillan, London, 178–203.

Ho, Teck-Hua, Sergei Savin, Christian Terwiesch. 2002. Managing demand and sales dynamics in new product diffusion under supply constraint. *Management Science* **48**(2) 187–206.

Ho, Teck-Hua, Sergei Savin, Christian Terwiesch. 2011. Note: A reply to "new product diffusion decisions under supply constraints". *Management Science* **57**(10) 1811–1812.

Jaikumar, Ramchandran. 1989. Yamazaki Mazak (A). Harvard Business School Case 9-686-083.

Jain, Dipak, Vijay Mahajan, Eitan Muller. 1991. Innovation diffusion in the presence of supply restrictions. *Marketing Science* **10**(1) 83–90.

Kanniainen, Juho, Saku J Mäkinen, Robert Piché, Alok Chakrabarti. 2011. Forecasting the diffusion of innovation: A stochastic bass model with log-normal and mean-reverting error process. *IEEE Transactions on Engineering Management* **58**(2) 228–249.

Kilkki, Pekka, Unto Väisänen. 1969. Determination of the optimum cutting policy for the forest stand by means of dynamic programming. *Acta Forestalia Fennic* **102** 1–23.

Kumar, Sunil, Jayashankar M Swaminathan. 2003. Diffusion of innovations under supply constraints. *Operations Research* **51**(6) 866–879.

Lilien, Gary L, Arvind Rangaswamy, Arnaud De Bruyn. 2017. *Principles of Marketing Engineering and Analytics*. 3rd ed. DecisionPro.

Mahajan, Vijay, Eitan Muller, Frank M Bass. 1991. New product diffusion models in marketing: A review and directions for research. *Diffusion of technologies and social behavior*. Springer, 125–177.

Mahajan, Vijay, Eitan Muller, Yoram Wind. 2000. *New-product diffusion models*, vol. 11. Springer Science & Business Media.

Mayorga, Maria E, Hyun-Soo Ahn. 2011. Joint management of capacity and inventory in make-to-stock production systems with multi-class demand. *European Journal of Operational Research* **212**(2) 312–324.

Meade, Nigel, Towhidul Islam. 2006. Modelling and forecasting the diffusion of innovation–a 25-year review. *International Journal of forecasting* **22**(3) 519–545.

MIT Sloan Management Review. 2013. With 3-D printing, the shoe really fits. https://sloanreview.mit.edu/article/with-3-d-printing-the-shoe-really-fits/. Retrieved October 19, 2020.

NASA. 2013. 3D printing: Food in space. https://www.nasa.gov/directorates/spacetech/home/feature_3d_food.html. Retrieved October 19, 2020.

Native Seeds/SEARCH. 2018. Glass gem. https://shop.nativeseeds.org/products/zp103/. Retrieved October 19, 2020.

Prusa Research. 2018. Prusa i3 printable parts and upgrades. https://www.prusa3d.com/prusa-i3-printable-parts/. Retrieved October 19, 2020.

Prusa Research. 2020. Prusa research: About us. https://www.prusa3d.com/about-us/. Retrieved October 19, 2020.

RepRap.org. 2018. https://reprap.org/wiki/RepRap. Retrieved October 19, 2020.

Reuters. 2018. Robots to make robots at ABB's new $150 million factory in china. https://www.reuters.com/article/us-abb-robots/robots-to-make-robots-at-abbs-new-150-million-factory-in-china-idUSKCN1N109X. Retrieved October 19, 2020.

Schaefer, Milner B. 1957. Some considerations of population dynamics and economics in relation to the management of the commercial marine fisheries. *Journal of the Fisheries Board of Canada* **14**(5) 669–681.

ScienceDaily. 2018. 3D printing of living cells. https://www.sciencedaily.com/releases/2018/02/180201092233.htm. Retrieved October 19, 2020.

Sethi, S. 2019. *Optimal Control Theory–Applications to Management Science and Economics*. 3rd ed. Springer, Cham.

Sethuraman, Nagarajan, Ali K Parlakturk, Jayashankar M Swaminathan. 2018. Personal fabrication as an operational strategy: Value of delegating production to customer. *Available at SSRN 3170011* .

Shen, Wenjing, Izak Duenyas, Roman Kapuscinski. 2011. New product diffusion decisions under supply constraints. *Management Science* **57**(10) 1802–1810.

Shen, Wenjing, Izak Duenyas, Roman Kapuscinski. 2014. Optimal pricing, production, and inventory for new product diffusion under supply constraints. *Manufacturing & Service Operations Management* **16**(1) 28–45.

Song, Jing-Sheng, Yue Zhang. 2020. Stock or print? Impact of 3-D printing on spare parts logistics. *Management Science* **66**(9) 3860–3878.

The Guardian. 2013. Monsanto sued small famers to protect seed patents, report says. https://www.theguardian.com/environment/2013/feb/12/monsanto-sues-farmers-seed-patents. Retrieved October 19, 2020.

The Guardian. 2017. New Zealand launches into space race with 3D-printed rocket. https://www.theguardian.com/world/2017/may/25/new-zealand-launches-space-race-3d-printed-rocket. Retrieved October 19, 2020.

The Verge. 2017. 3D-printed titanium parts could save Boeing up to $3 million per plane. https://www.theverge.com/2017/4/11/15256008/3d-printed-titanium-parts-boeing-dreamliner-787. Retrieved October 19, 2020.

Appendix

Pontryagin's maximum principle for Problem (1). We apply Pontryagin's maximum principle for the current-value formulation (Sethi 2019, p. 80) to Problem (1). For state variables D, S and I, respectively denote their co-state variables by λ_D, λ_S and λ_I. The current-value Hamiltonian is

$$\begin{aligned}
&H(D,S,I,p,s,\lambda_D,\lambda_S,\lambda_I,t)\\
&\doteq \pi s(t) - cp(t) - hI(t) + \lambda_D(t)d(t) + \lambda_S(t)[s(t) - wS(t)] + \lambda_I(t)[p(t) - s(t)]\\
&= \lambda_D(t)[\alpha + \beta S(t)/m][m - D(t)] - hI(t) - \lambda_S(t)wS(t) + [\lambda_I(t) - c]p(t) + [\pi + \lambda_S(t) - \lambda_I(t)]s(t).
\end{aligned} \tag{8}$$

Note that the Hamiltonian (8) is linear in $p(t)$ and $s(t)$. Since Pontryagin's maximum principle requires that the Hamiltonian must be maximized at any time by the optimal controls, we immediately come to the conclusion that the optimal control for Problem (1) is bang-bang.

Because Problem (1) contains constraints on control variables $p(t)$ and $s(t)$, we also need to define a Lagrangian for the Hamiltonian. Let μ_{pL} (μ_{pU}) and μ_{sL} (μ_{sU}) be the Lagrange multipliers for the lower (upper) constraints of $p(t)$ and $s(t)$, respectively. Let ν be the Lagrange multiplier for the transversality constraint $I(T) \geq 0$. The current-value Lagrangian is

$$\begin{aligned}
&L(D,S,I,p,s,\lambda_D,\lambda_S,\lambda_I,\mu_{pL},\mu_{pU},\mu_{sL},\mu_{sU},t)\\
&\doteq H(D,S,I,p,s,\lambda_D,\lambda_S,\lambda_I,t) + \mu_{pL}(t)p(t) + \mu_{pU}(t)[rI(t) - p(t)] + \mu_{sL}(t)s(t) + \mu_{sU}(t)[d(t) - s(t)]\\
&= [\lambda_D(t) + \mu_{sU}(t)][\alpha + \beta S(t)/m][m - D(t)] + [r\mu_{pU}(t) - h]I(t) - \lambda_S(t)wS(t)
\end{aligned}$$

$$+ [\lambda_I(t) + \mu_{pL}(t) - \mu_{pU}(t) - c]p(t) + [\pi + \lambda_S(t) - \lambda_I(t) + \mu_{sL}(t) - \mu_{sU}(t)]s(t).$$

Pontryagin's maximum principle requires the following conditions for any t for optimality:

Maximum conditions:

$$p(t) = rI(t) \Leftrightarrow \lambda_I(t) > c, \quad p(t) = 0 \Leftrightarrow \lambda_I(t) < c,$$

$$s(t) = d(t) \Leftrightarrow \pi + \lambda_S(t) > \lambda_I(t), \quad s(t) = 0 \Leftrightarrow \pi + \lambda_S(t) < \lambda_I(t).$$

First-order conditions:

$$\lambda_I(t) + \mu_{pL}(t) - \mu_{pU}(t) - c = \pi + \lambda_S(t) - \lambda_I(t) + \mu_{sL}(t) - \mu_{sU}(t) = 0.$$

Complementary slackness:

$$\mu_{pL}(t), \mu_{pU}(t), \mu_{sL}(t), \mu_{sU}(t) \geq 0,$$

$$\mu_{pL}(t)p(t) = \mu_{pU}(t)[rI(t) - p(t)] = \mu_{sL}(t)s(t) = \mu_{sU}(t)[d(t) - s(t)] = 0,$$

Adjoint conditions:

$$\lambda_I'(t) = \rho\lambda_I(t) + h - r\mu_{pU}(t),$$

$$\lambda_D'(t) = \rho\lambda_D(t) + [\lambda_D(t) + \mu_{sU}(t)][\alpha + \beta S(t)/m],$$

$$\lambda_S'(t) = (\rho + w)\lambda_S(t) - [\lambda_D(t) + \mu_{sU}(t)][\beta - \beta D(t)/m].$$

Transversality conditions:

$$\lambda_D(T) = \lambda_S(T) = 0, \quad \lambda_I(T) = \nu \geq 0, \quad \nu I(T) = 0.$$

□

Proof of Proposition 1. Since Problem (1) does not contain pure state constraints, all co-state variables are continuous in t under optimality. Consider any time t where $I(t) > 0$ and $p(t) = 0$, such that $p(t) < rI(t)$. Due to the complementary slackness, $\mu_{pU}(t) = 0$. The adjoint condition for λ_I becomes

$$\lambda_I'(t) = \rho\lambda_I(t) + h \Rightarrow \lambda_I(t) = Ce^{\rho t} - h/\rho$$

where C is a constant. We argue that in a non-trivial setting $C > 0$, otherwise λ_I would always be negative and production can never take place (recall the maximum condition that $p(t) > 0 \Leftrightarrow \lambda_I(t) > c$). Therefore $\lambda_I'(t) > 0$ whenever $p(t) = 0$. This implies that $p(t)$ can never be switched off as long as there is inventory, thus the proposition. □

Proof of Proposition 2. If at optimality $p(t) \equiv 0$, then the optimal sales policy is clearly $s(t) \equiv d(t)$ (selling off I_L), and thus the inventory will run out before T for sufficiently large T. Now consider $p(t) > 0$ for some t at optimality.

When production is on, due to the complementary slackness, $\mu_{pL}(t) = 0$. Due to the first-order condition, $\mu_{pU}(t) = \lambda_I(t) - c$. The adjoint condition for λ_I becomes

$$\lambda_I'(t) = \rho\lambda_I(t) + h - r[\lambda_I(t) - c] \Rightarrow \lambda_I(t) = C'e^{-(r-\rho)t} + \frac{cr+h}{r-\rho}$$

where C' is a constant.

When $T \to \infty$, $\lambda_I(T) \to (cr+h)/(r-\rho)$. Due to the transversality condition, $\lambda_S(t) \to 0$. If $(r-\rho)\pi > cr+h$, we know $\pi + \lambda_S(t) > \lambda_I(t)$ for sufficiently large t. Due to the optimality condition, $s(t) = d(t)$ for sufficiently large t. This implies that $I(t)$ is never depleted before T and production and sales are maximized until T. If $(r-\rho)\pi < cr+h$, we have $\pi + \lambda_S(t) < \lambda_I(t)$ for sufficiently large t. Due to the optimality condition, $s(t) = 0$ for sufficiently large t. It is however clearly suboptimal to stop sales before the inventory is depleted, implying that the inventory cannot be positive, namely $I(t) = 0$, $\tau \le t \le T$, $\exists \tau < T$, for sufficiently large T. □

Proof of Theorem 1. First, note that it is clearly not optimal to keep both production and sales off. Consider the optimal trajectory. Suppose at time t, $I(t) > 0$ and production and sales are both off for an interval δ. The discounted-profit-to-go at $t+\delta$ must be positive, otherwise $I(t) > 0$ would not have been possible (it would have been optimal to cut the product life cycle short). We argue that bringing all future production and sales controls forward by δ has three effects: 1. it saves inventory cost $I(t)h\delta$; 2. it saves discount cost by bringing forward a positive discounted-profit-to-go; and 3. it effectively increases the remaining market size from $m - D(t+\delta)$ to $m - D(t)$, which means that all future production and sales controls remain admissible. Effects 1 and 2 have positive impacts on the total discounted profit and Effect 3 has a non-negative impact on the total discounted profit, thus the overall value is improved. Therefore any control policies with production and sales both off can be improved.

When $I_L > \overline{I}$, it is clearly optimal to never produce and maximize sales, thus the policy.

When $\underline{I} < I_L < \overline{I}$, consider the policy to always keep sales and production on, which would lead to leftover inventory at T. Now consider shutting off production over certain time periods while still having leftover inventory at T. Due to Proposition 1, it is never optimal to switch production back on, which implies that shutting off production can only take place in an initial time interval (namely postponing switching on production). Doing so does not affect sales because inventory is never depleted, but clearly saves production cost and inventory cost. Therefore one can improve the policy by shutting off production in such an initial time interval such that the inventory is depleted exactly at T while always keeping sales on.

We then argue that this is the optimal policy. Note that the aforementioned analysis reveals that anytime there are only three possible optimal controls: only production on; only sales on; both production and sales on, and that production can never be switched off until the inventory becomes depleted. Therefore, the only changes allowed on the policy without violating optimality are 1. shut off production for a longer initial period; 2. shut off sales for some interval during production. Change 2 is clearly not optimal because doing so causes reduced revenue during this interval and future time (through reduced demand). Change 1 would cause inventory depletion before T. To see why this is not optimal, consider the short time interval prior to when production is switched on. In this time interval, sales is on and production is off. Because of Strong Replicability, we know that it would increase the total discounted profit to postpone sales and use the inventory for replication and sell the good in the immediate future. The problem is that the immediate future demand is being satisfied already. However, by the same argument, the immediate future sales can be pushed further for replication to increase the total discounted profit. By making the same argument recursively, the inventory depletion time will be pushed back which is feasible given it happens before T. Therefore, as long as the inventory is depleted before T, we can use this argument to switch production on earlier and push back inventory depletion while increasing the total discounted profit, until the inventory is depleted exactly at T.

When $I_L < \underline{I}$, consider the policy to always keep sales and production on, which would lead to inventory depletion before T. Consider a small time interval after time 0 during which sales is on. Because of Strong Replicability, we know that it would increase the total discounted profit to postpone sales and use the inventory for replication and sell the good in the immediate future. The problem is that the immediate future demand is being satisfied already. However, our earlier reasoning shows that all future sales can be postponed leading to inventory depletion being pushed back, while increasing the total discounted profit. Therefore, as long as the inventory is depleted before T, the time to switch on sales can be postponed until the inventory is depleted exact at T. We argue that this is the optimal policy following the same reasoning as above. □

Proof of Corollary 1. Denote the optimal production switch-on time by $t_p^*(r)$ and the optimal sales switch on time by $t_s^*(r)$ when the replication rate is r. Theorem 1(a) implies that $t_p^*(r) = 0$ when the launch inventory I_L is lower than $\underline{I}$. When $t < t_s^*(r)$, because of the bang-bang nature of the optimal policy, we have $p(t) = rI(t)$ and $s(t) = 0$, $I'(t) = rI(t)$, and $I(t) = I_L e^{rt}$. With a larger replication rate $r' > r$, the inventory level at $t_s^*(r)$ is $I_L e^{r' t_s^*(r)} > I_L e^{r t_s^*(r)}$. Since the inventory is always depleted at T, sales must be switched on earlier, i.e., $t_s^*(r') < t_s^*(r)$.

Theorem 1(b) implies that $t_s^*(r) = 0$ when $\underline{I} < I_L < \overline{I}$. When $t < t_p^*(r)$, because of the bang-bang nature of the optimal policy, we have $p(t) = 0$ and $s(t) = d(t)$, and $I'(t) = -d(t)$. Since the demand diffusion process does not depend on production, neither does the inventory given $p(t) = 0$. Therefore, $I(t)$ does not depend on r when $t < t_p^*(r)$. With a larger replication rate $r' > r$, the total production from $t_p^*(r)$ to T is larger. Since the inventory is always depleted at T, production must be switched on later, i.e., $t_p^*(r') > t_p^*(r)$. $\square$

Proof of Theorem 2. Weak Replicability implies that it is never optimal to postpone sales for replication, and the proof of Theorem 1 shows that it is never optimal to keep both production and sales off, thus it is optimal to always keep sales on.

When $I_L > \overline{I}$, it is clearly optimal to never produce, thus the policy.

When $I_L < \overline{I}$, the inventory would be depleted at some τ before T without production. Consider a small time interval δ before τ. Consider switching production on at $\tau - \delta$. Doing so increases the revenue by $r\pi\delta$ while incurring production cost $rc\delta$, plus other higher-order terms of δ. Since $\pi > c$, production should be switched on when it is sufficiently close to τ (at the same time pushing back τ). The question is how far back should production be switched on. When production lasts for a non-trivial time interval, there are three additional effects to be accounted for. First, the produced units will incur additional inventory costs. Second, because production takes place before sales, revenue and inventory costs need to be discounted. Third, the exponential growth due to production during this period needs to be considered.

We first consider effects 1 and 2. A unit of the good should be produced no longer than a maximum production interval $\bar{\Delta}$ prior to being sold, where $\bar{\Delta}$ is such that effects 1 and 2 exactly offset the profit margin $\pi - c$, namely

$$c + h \int_0^{\bar{\Delta}} e^{-\rho t}\,\mathrm{d}t = e^{-\bar{\Delta}\rho}\pi \Rightarrow \bar{\Delta} = \begin{cases} (\pi - c)/h & \text{if } \rho = 0, \\ [\ln(\rho\pi + h) - \ln(\rho c + h)]/\rho & \text{if } \rho > 0. \end{cases}$$

We call $\bar{\Delta}$ the effective maximum production interval because it is solved assuming a fixed unit of the good being held in inventory during the interval without considering effect 3, the exponential growth of self-replication. In reality, a unit of the good to be sold at τ was only e^{-rt} units at $\tau - t$. Therefore, the actual maximum production interval Δ for which production should take place is be longer than the effective maximum production interval $\bar{\Delta}$, with the per-unit average maximum production interval being equal to the latter, namely

$$\int_0^{\Delta} e^{-rt}\,\mathrm{d}t = \bar{\Delta} \Rightarrow \Delta = \begin{cases} -\ln[1 - r(\pi - c)/h]/r & \text{if } \rho = 0, \\ -\ln\{1 - r[\ln(\rho\pi + h) - \ln(\rho c + h)]/\rho\}/r & \text{if } \rho > 0. \end{cases}$$

The meaning of Δ is the maximum production period prior to inventory depletion. Producing longer than Δ and the revenue cannot justify the production and inventory costs. In other words, the optimal policy for Weak Replicability with $I_L < \overline{I}$ can be generally stated as to always keep sales on, and switch production on at such a time that production lasts for exactly Δ, and the inventory will be depleted before T. This outcome will occur in Case (b). On the other hand, several reasons may cause production to be unable to last for Δ. If I_L is sufficiently small in Case (a), even if production is always kept on, it may not last for Δ. If I_L is sufficiently large in Case (c), production lasting for Δ will lead to leftover inventory at T which is clearly suboptimal, and thus production lasts shorter than Δ. These arguments are formalized as Theorem 2. □

Proof of Corollary 2. By Theorem 2, we have $\Delta = -\ln(1-r\theta)/r,\ r \in (0, 1/\theta)$, where $\theta = (\pi - c)/h$ if $\rho = 0$, and $\theta = [\ln(\rho\pi + h) - \ln(\rho c + h)]/\rho$ if $\rho > 0$. It is easy to verify that $\lim_{r\to 0} \Delta = \theta > 0$. Taking the derivative of Δ with respect to r yields

$$\frac{\mathrm{d}\Delta}{\mathrm{d}r} = \frac{1}{r^2}\left[\ln(1-\theta r) + \frac{\theta r}{1-\theta r}\right], \quad \lim_{r\to 0}\frac{\mathrm{d}\Delta}{\mathrm{d}r} = \frac{\theta^2}{2} > 0.$$

Let $f(r) \doteq \ln(1-\theta r) + \frac{\theta r}{1-\theta r}$ and we have

$$\lim_{r\to 0} f(r) = 0,\ \text{and}\ f'(r) = r\left(\frac{\theta}{1-\theta r}\right)^2 > 0,\ \forall r \in (0, 1/\theta),$$

hence $f(r) > 0$ and $\frac{\mathrm{d}\Delta}{\mathrm{d}r} > 0,\ \forall r \in (0, 1/\theta)$.

Taking the derivatives of θ with respect to π, c and h yields

$$\frac{\mathrm{d}\theta}{\mathrm{d}\pi} = \frac{1}{\rho\pi + h} > 0, \quad \frac{\mathrm{d}\theta}{\mathrm{d}c} = -\frac{1}{\rho c + h} < 0, \quad \frac{\mathrm{d}\theta}{\mathrm{d}h} = \frac{-(\pi - c)}{(\rho\pi + h)(\rho c + h)} < 0, \quad \forall \rho \geq 0.$$

Since $\frac{\mathrm{d}\Delta}{\mathrm{d}\theta} = (1 - r\theta)^{-1} > 0$, we have

$$\frac{\mathrm{d}\Delta}{\mathrm{d}\pi} = \frac{\mathrm{d}\Delta}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}\pi} > 0, \quad \frac{\mathrm{d}\Delta}{\mathrm{d}c} = \frac{\mathrm{d}\Delta}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}c} < 0, \quad \frac{\mathrm{d}\Delta}{\mathrm{d}h} = \frac{\mathrm{d}\Delta}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}h} < 0,$$

hence the corollary. □

Proof of Proposition 3. Note that the proof of Theorem 1 does not depend on the demand process. Therefore, we know that it is optimal to keep production on (both pre- and post-launch), and switch and keep sales on at such a time that the inventory is depleted at the end of the product's life cycle. Clearly, the optimal t_L should not be past $(\ln \underline{I} - \ln I_0)/r$, otherwise there will be leftover inventory at $t_L + T$. What remains is to show that the optimal t_L should not be before $(\ln \underline{I} - \ln I_0)/r$ either.

Suppose the optimal $t_L' < (\ln \underline{I} - \ln I_0)/r$, and sales are switched on sometime $t' > 0$ after the product launch. The firm's profit in this case is the same as the optimal profit if the total market size m is increased to $m' = me^{\alpha t'_L}$ (calculated from the demand diffusion equation with $S(t) \equiv 0$) and the product's life cycle T is increased by t'_L, but demand diffusion began at time 0. On the other hand, consider $t''_L = (\ln \underline{I} - \ln I_0)/r$. The firm's profit in this case is the same as the optimal profit if the total market size m is increased to $m'' = me^{\alpha t''_L} > m'$ and the product's life cycle T is increased by $t''_L > t'_L$, but demand diffusion began at time 0. We argue that the latter optimal profit is strictly larger than the former. First, Theorem 1 states that under Strong replicability, the optimal policies must ensure that the product is sold through the entire life cycle T. Therefore, an optimal policy for T can be implemented to generate the same profit for a longer T, which cannot be optimal because production does not last for the entire T, implying that a longer T strictly increases the optimal profit. Similarly, an optimal policy for m can be implemented to generate the same profit for a larger m, implying that a larger m cannot decrease the optimal profit. As a result, we know that the optimal $t_L = (\ln \underline{I} - \ln I_0)/r$. $\square$

Proof of Proposition 4. Note that the proof of Theorem 1 does not depend on the demand process. Therefore, we know that it is optimal to keep production on, and switch and keep sales on at such a time that the inventory is depleted at the end of the product's life cycle. With $T = \infty$, the inventory level to meet stable demand d through replication is d/r, and $I_L e^{t_s r} = d/r$ yields the optimal time to switch on sales $(\ln d - \ln I_L - \ln r)/r$. For finite T, because the inventory will be gradually sold off to meet some demand, there is no need to build the inventory up to d/r, and thus $t_s < (\ln d - \ln I_L - \ln r)/r$. $\square$