Optimal Growth in Two-Sided Markets

Zhen Lian

S. C. Johnson Graduate School of Management, Cornell University, zl458@cornell.edu

Garrett van Ryzin

Cornell Tech and Lyft, Inc., gjv25@cornell.edu

Abstract

We develop a theoretical model of optimal growth in two-sided markets. The model posits that market output (number of transactions) is a function of the stock of supply and demand. This market output is modeled using a homogeneous function, which can have increasing or decreasing returns to scale. The supply and demand stock levels follow a growth model in which the rate of growth at each point in time is a function of both the surplus each side of the market receives and the attrition of supply and demand (supply and demand lifetimes). The surplus can be apportioned between the two sides of the market by changing the price paid to sellers and the price charged to buyers, which we assume the platform controls. Through these price levers, the platform can pay subsidies to one or both sides of the market. We investigate the behavior of optimal market growth, including the point at which the market becomes self-sustaining and the long-run optimal size of the market. We characterize the optimal balance between supply and demand as the market size grows and determine optimal subsidy policies that maximize discounted total profit. For the case of both increasing and decreasing returns without price constraints, we show the optimal policy is to grow using an impulse of subsidy spending (a "subsidy shock") to move the market immediately to its optimal long-run size. This result is consistent with the "race to growth" observed in many two-sided markets like ride-sharing.

Keywords: two-sided market, optimal growth, market balance, sharing economy

1 Introduction

How to optimally grow a two-sided market is an important strategic problem in marketplace companies. In particular, all major ride-sharing companies (Uber, Lyft, Didi) have large internal organizations focused on how best to manage growth. This includes the acquisition of new passengers and drivers, encouraging more usage from existing passengers and drivers, passenger and driver pricing and how to balance growth spending between the demand and supply sides of the market. Similar growth problems are faced in two-sided markets in hospitality (Airbnb), food delivery (Door Dash), and casual labor (Task Rabbit, Handy). A fundamental problem in growth planning is managing the tradeoff between the cost of subsidizing growth (through prices or incentives) and the benefit of the scale economies that are achieved through growth. We analyze a stylized version of this problem.

In particular, our model assumes outputs (transactions) are "produced" by two primary inputs: the stock of supply and demand, which can be thought of as the number of adopters on each side of the market. This production is modeled using a homogeneous production function, which could have either increasing or decreasing returns to scale. We assume a monopolist platform controls the price of supply and demand. Supply and demand stock levels (number of adopters) in turn evolve according to a growth model in which the rate of growth is a function of both the surplus each side of the market receives as well as losses due to attrition (finite demand and supply "lifetimes"). The platform can apportion this surplus between the two sides of the market by changing the price paid to sellers and the price charged to buyers. The platform can also subsidize the market by setting the price of demand lower than the price of supply. Using this model, we analyze how to optimally grow the market through prices and subsidies to maximize the platform's discounted total profits. Because ours is a monopoly model, there is no competitive benefit to growth (such as preemption); rather, the optimal growth policy is determined solely by whether the benefits of getting to scale quickly justify the costs of subsidizing growth.

Our analysis relies on a novel state space reduction in which we transform our two-state market model (with stock of supply and stock of demand as states) to a model with a single scalar "market size" state. This market size has units of dollars and can be interpreted as the level of investment in the market. Indeed, in our model the platform could liquidate the market by charging prices that reduce the stock of supply and demand to zero and generate a one-time profit equal to the market size.

Having a single market size greatly simplifies the analysis of market growth, since it provides a clear ordering of growth paths; in particular, a growth path which has a market size always greater than another growth path corresponds to "faster" growth. How the market size is apportioned between the supply and demand sides then reduces to a scalar "market balance" control variable. This too is conceptually appealing, since it allows us to cleanly distinguish between the total size of the market and how that market size is allocated between supply and demand. Optimal supply . and demand balance itself is also a fundamental concern of marketplace growth organizations and our reformulation isolates and highlights this important choice variable.

While derived from our stylized model, this state space reduction is not unrealistic. Intuitively, it corresponds to a case where it is easier for a platform to shift surplus from the supply to demand side (or vice versa) than it is to grow the total market. That is, the platform can raise price significantly for buyers (demand) and pay suppliers (supply) significantly more by simply passing the revenues from the demand to supply side. This will result in a loss of adoption on the demand side and an increase in adoption on the supply side, but requires no net subsidy. Hence, in the limit we can use such abrupt pricing changes to effectively choose the market balance we want at any point in time. However, such transfers of value between supply and demand do not change the total market value; they simply amount to "cashing out" of one side of the market to "buy up" the other side - leaving the total market value unchanged. True size growth requires increasing the total value of the market overall. Our size and balance decomposition concisely reflects this distinction.

We then investigate the behavior of optimal market growth, including the point at which the market becomes self-sustaining (the "critical size") and the long-run stable size of the market (the "saturation size"). These size thresholds provide important information on what size is necessary to achieve in order for a market to operate subsidy-free when markets have increasing returns to scale and the rational limits to market growth when markets have decreasing returns to scale.

We then characterize the optimal balance between supply and demand as a market grows. In particular, we show that optimal market balance at any point in time is uniquely determined by the market size at that point in time, for both increasing and decreasing returns. That is, regardless of the market size growth path, the optimal market balance choice at each time is determined only by the market size at that time, the structural parameters governing the supply and demand growth process and the platform's discount rate. This balance choice reflects a fundamental trade-off between minimizing loss of size due to attrition and maximizing market output, which increases surplus and adoption. The path of optimal balance as a function of size, however, depends on the relative durability of supply and demand (which has the longer mean lifetime) and whether the market has increasing or decreasing returns to scale. Together, these results provide robust insights into the optimal choice of supply and demand balance as markets grow.

We then analyze how prices should be set to maximize discounted total profits. In general, we show that rapid subsidization is optimal in a range of cases – though not always. In particular, for both increasing and decreasing returns to scale where prices are unconstrained, an impulse of subsidy spending (a "subsidy shock") that takes the market instantaneously from its initial size to its final (steady-state or terminal) size is optimal. When there are constraints on the growth trajectory, we show that in the decreasing returns case, faster growth paths are always better than slower growth paths. For increasing returns, faster paths may not always dominate depending on whether the market potential is above or below a threshold size. However, if the market potential is large enough, a subsidy shock is still optimal.

The fact that faster growth is better closely mirrors the observed growth strategies in many marketplace companies, as noted in the popular business press [1] [2] [3] In ride-sharing in particular, heavy subsidization to incentivize rapid growth has been the norm in the industry; ride-sharing firms spent billions of dollars and multiple years subsidizing riders and drivers to build scale in the cities they serve. Despite the skepticism expressed by some in the popular business press about such growth strategies [4], our results suggest that they may indeed be rational.

The organization of the paper is as follows: Section 2 is a review of related literature; Section 3 is the model setup, consisting of four parts: (1) an output model about how supply and demand

[1] *Uber Aims to Maintain Heavy Spending to Keep Rivals at Bay.* (April 12, 2019). https://www.ft.com/content/8a28ba78-5d09-11e9-9dde-7aedca0a081a

[2] *How A Venture Capitalist Would Look At Uber's Value Today: It's All About The Marketplace.* (May 14, 2019). https://www.forbes.com/sites/mikeghaffary/2019/05/14/how-a-venture-capitalist-would-look-at-ubers-value-today-its-all-about-the-marketplace/

[3] *Blitzscaling.* (April 2016). https://hbr.org/2016/04/blitzscaling

[4] *Uber's Path of Destruction*, (Summer 2019, Vol. 3, No. 2), https://americanaffairsjournal.org/2019/05/ubers-path-of-destruction/

generate market outputs; (2) a growth model about how the growth rates of supply and demand react to wage and price at any time; (3) a market size-and-balance reformulation (4) the optimal growth formulation. Section 4 shows our main results, consisting of three parts: (1) the optimal balance of supply and demand; (2) the stationary solution; (3) the optimal growth policies. Section 5 gives some numerical examples.

2 Related literature

There is extensive literature in economics on two-sided markets. Among all, the price allocation between the two sides of the market is a central question. Caillaud and Jullien (2003) study the competition among intermediaries who perform matchmaking service between two groups of users. The authors characterize the intermediary's optimal pricing strategy as "divide-and-conquer", i.e. subsidizing the participation of one group of users and recovering the profit loss from the other group. Rochet and Tirole (2003) looks at the a two-sided market that exhibits usage externalities and the platform charges per-transaction fees on both sides of the market (e.g. the credit card industry). Armstrong (2006) looks at a two-sided market that exhibits membership externalities and the platform charges a lump-sum membership fee on both sides of the market (e.g. night clubs and shopping malls). From both papers, the optimal price allocation for a monopoly platform follows a standard Lerner formula with some reinterpretation of marginal costs, implying that the price on one side of the market is inversely related to this side's elasticity of demand. Rochet and Tirole (2006) builds a unifying model that combine the usage externality and the membership externality. Weyl (2010) extends Rochet and Tirole (2006) by incorporating user heterogeneity on each side of the platform. Hagiu (2004) builds on Caillaud and Jullien (2003) and looks at the commitment issue in a two-sided market with one side of the market arriving earlier than the other side of the market.

Due to the static nature of these models, they do not consider the process of acquiring users; once the price is set by the platform, users instantly join as long as they gain positive utility. This may be a good characterization for those two-sided platforms that are already well-known and mature, but for new platforms, it takes time for market penetration and product diffusion. Moreover, although a subsidy policy is widely discussed in these papers, it is mostly driven by the asymmetry of price elasticities and cross externalities in two sides of the market, and thus follows the "divide-and-conquer" strategy mentioned above. In contrast, we look at the trade-off between the cost of subsidizing growth and the benefit of the scale economy achieved by growth. In our paper, the definition of subsidy is a net profit loss (see 3.5. Definitions). We show that it is generally optimal for the platform to apply heavy subsidies at the early stage of the market growth. In other words, the optimal policy may involve the platform losing money on both sides of the market simultaneously for some time. Clearly, such a subsidy cannot be optimal in the static models we have discussed in the platform literature.

Our model shares many similarities in our macro view of two-sided markets with those in the modern economic growth literature. For example, Solow (1956) tries to explain long-run economic growth by representing a complex economy as a simple one-good economy and abstracting away from individual decisions (Acemoglu, 2009). In the Solow model, the output of the economy is generated by labor and capital following a Cobb-Douglas production function with constant returns to scale, and the economy grows following the law of motion in the capital-labor ratio. The neoclassical growth literature builds on the Solow model by endogenizing the consumer's consumption decision. Under the representative household's utility maximization, the optimal path of the capital-labor ratio and consumption are identified. Seeking to explain the empirical evidence of balanced growth (Kaldor, 1957), this stream of literature put an emphasis on policies that lead to a constant capital-output ratio and constant output growth rate. In other words, their approach is to identify conditions under which a balanced growth path is optimal (Acemoglu, 2009).

In contrast, our paper seeks to provide insights for the optimal growth of modern day two-sided markets like ride-sharing, for which rapid growth and heavy subsidies are typical. Therefore, we do not constrain our analysis to only balanced growth paths. Moreover, in the growth literature, the state variable is often a single variable, the capital-labor ratio. This is due to a common assumption in this literature that the population is constant or grows at an exogenous rate and that the fundamental policy objective is per-capita output. Therefore, their optimization problem is inherently a one-dimensional problem. However, these assumptions do not make sense in the context of two-sided markets like ride-sharing. Total profit is the objective for a profit-making platform not profit per unit of supply or demand. Also, compared to a one-sided market, how to balance the growth of supply and demand is a unique challenge for two-sided markets, and holding any one side of the market constant is not relevant. It is essential to consider the growth of supply and demand simultaneously, and this raises the methodological challenge of characterizing the optimal policy for a problem with a two-dimensional state variable and a two-dimensional control variable, of which a closed-form solution for both steady states and transitional dynamics are not commonly available. In this sense, our paper adds to the neoclassical growth literature by providing a stylized and tractable formulation for two-dimensional growth.

In the operations management field, there is a burgeoning literature that studies pricing issues in the context of ride-sharing platforms or online two-sided markets. Cachon et al. (2017) discuss the role of dynamic pricing in a platform with service providers that have self-scheduling ability and how it influences users' welfare. Banerjee et al. (2015) conclude dynamic pricing does not necessarily yield higher performance but is more robust than static pricing under imperfect information. Castillo et al. (2017) argue that surge pricing can help avoiding the matching failure due to long-distance inefficient pickup while keeping low prices at most times and thus solves the "wild goose chase" problem. Gurvich et al. (2019) study the platform's optimal compensation structure in the presence of self-scheduling agents, which addresses the trade-off between the flexibility from self-scheduling and the cost for sustaining the agent pool. Taylor (2018) studies how the customer's delay sensitivity and the driver's independence influence the platform's optimal pricing and wage decisions. Bai et al. (2018) study the optimal wage and price ratio using a queuing model. Hu and Zhou (2017) study the fixed commission contract and find that the flat commission can be optimal or near-optimal if well designed. Interestingly, they show that the two-dimensional price and wage optimization can be reduced to a one-dimensional matching quantity optimization problem. Bimpikis et al. (2019) model the ride-sharing market as a network of locations with riders heterogeneous in destination preferences and drivers who self-relocate to maximize expected payoff.

Our paper is also related to the literature on market thickness. Early theoretical work in labor markets commonly assumed increasing or constant returns to scale (e.g. Diamond (1982)). However, there is more recent work on returns to scale in the context of innovative marketplaces. Kabra et al. (2016) use data from a ride-hailing market and show evidence of increasing returns to scale. Cullen and Farronato (2014) use data from TaskRabbit, an online peer-to-peer freelance marketplace and finds no evidence of increasing returns to scale. Li and Netessine (2019) use data from an online holiday rental platform and find that increased market thickness actually leads to lower matching efficiency, which implies decreasing returns to scale. Nikzad (2017) studies how market thickness and competition influence the market equilibrium in the ride-hailing market.

Our paper is based on optimal control theory and is related to a number of applications of optimal control theory in the field of management science. For example, Vidale and Wolfe (1957) study a firm's optimal advertising expenditure problem when the effect of advertising carries over but diminishes over time. The rate of sales is affected by the advertising effort in two ways: new adoption from the unsold portion of the market due to advertising, and loss from the sold portion of the market over the time. Dhebar and Oren (1986) study a single supplier's pricing decision for a new product that exhibits demand side network externalities, subject to dynamics of market growth. In our model, we consider the optimal growth problem as a profit maximization problem with growth rates of supply and demand driven by users' adoption and attrition from both sides of the market over time.

Our work is distinct from the above literature because; 1) we focus on optimal long-term profit of the platform; 2) we analyze the optimal balance between supply and demand, and 3) we characterize optimal growth strategies under both increasing and decreasing returns to scale. Our aim is to provide a concise but comprehensive theory of optimal growth in two-sided markets that provides useful insights and a framework on which other studies can build.

3 Model Setup

We begin by defining how supply and demand combine to produce output in a two-sided market. We then define how the supply and demand sides of the market grow as a function of the surplus each side receives from the market as well as attrition losses due to finite supply and demand lifetimes. Next, we show how this two-sided model can be reduced to a one-dimensional model in which a scalar market size is the only state variable and market balance (the ratio of supply size to demand size) becomes a control. The reformulation simplifies our analysis of growth and provides useful intuition. Lastly, we define the optimal growth problem for our model.

3.1 Supply, demand and production

Let $s(t)$ denote the stock of supply and $d(t)$ denote the stock of demand at time t. It is most natural to think of these stocks as being the number of suppliers and buyers who have joined the platform (adopters), but the measure of supply and demand could be more general. For example, in ride-sharing networks, supply is typically measured in driver-hours and demand is measured as the number of unique app sessions which could convert to a request for a ride.

To keep the terminology generic, we refer to a unit of market output as a *transaction* (e.g. a transaction is a ride in a ride-sharing network). We assume the total number of transactions generated by a given stock of supply s and demand d follows a homogeneous production function, denoted as $g(s, d)$.

$$g(ks, kd) = k^{\alpha} g(s, d) \tag{1}$$

where $g(s, d)$ is homogeneous of degree α.

The factor α represents the *total returns to scale* of the market. If $\alpha > 1$, then transactions exhibit increasing returns to scale, meaning if we were to increase supply and demand by the same factor k, total transactions would increase by more than a factor k. If $\alpha < 1$, then transactions exhibit decreasing returns to scale, meaning if we were to increase the supply and demand by the same factor k, total transactions would increase by less than a factor k. However, total returns to scale can vary by type of market and stage of growth, in which case our model should be interpreted as describing the optimal growth problem within a given growth stage.

Increasing returns to scale corresponds to a market where as the supply and demand increase, transactions between buyers and sellers get more efficient. For example, in a ride-sharing market, if growth occurs due to more riders and drivers operating in the same geographical area, this leads to an increase in the density of available (open) drivers, which in turn leads to shorter pickup times. The decreased time to pick up each rider means drivers spend a higher proportion of their time on trip (higher on-trip utilization), enabling them to complete more trips per hour on the platform. Hence, in this case, increasing both supply and demand by a factor k will lead to a increase in rides by more than a factor k.

This scale economy effect was popularized in the famous “Uber napkin diagram” (Chen (2015)) to explain a virtuous cycle in ride-sharing – greater volume leading to higher density, leading to higher driver utilization, leading to lower costs, leading to lower prices, leading to even greater volume – that was used by Uber insiders to explain its early success. This virtuous growth cycle is typical of the early phases of ride-share growth in a city.

In contrast, decreasing returns to scale can occur when growth involves expanding the market to increasingly difficult-to-serve segments or harder-to-match buyers and sellers, such that transactions among sellers and buyers become less efficient. Consider ride-sharing again. Once growth in a city has saturated the city "core", additional growth tends to occur by expanding the geographical coverage of service beyond the city core to outskirts and surrounding suburban areas with lower population density. Such growth can result in a decrease in the average density of available drivers, which leads to increases in average pickup time and consequently reduced driver on-trip utilization. If driver density declines due to geographical growth, then an increase in supply and demand by a factor k will result in less than a factor k increase in rides. This type of growth by geographical expansion is typical of the more mature growth phase of ride-sharing service in a city.

Lastly, we note that one can combine our results from these two cases into a single model of growth in which we assume the market makes a transition from increasing to decreasing returns after it grows beyond a transition "market size". This is described in more detail below in Section 4.3.3.

3.2 Prices and supply/demand growth

Let p_s denote the price paid to the seller for a transaction and p_d denote the price charged to the buyer for a transaction. We assume the platform controls both the prices p_s and p_d and can set $p_d < p_s$ if it wants to subsidize the market. Both buyers and sellers have homogeneous utilities; buyers receive a gross utility v for each transaction and sellers incur a cost c for each transaction. Hence the net seller surplus per unit of supply is

$$w = \frac{(p_s - c)g(s,d)}{s} \tag{2}$$

and the net buyer surplus per unit of demand is

$$u = \frac{(v - p_d)g(s,d)}{d} \tag{3}$$

We then model the growth rate of adopters on the supply and demand side as a linear function of these net utilities

$$\dot{s}(t) = (-\beta_0^s + \beta_1^s w)s(t) \tag{4}$$

$$\dot{d}(t) = (-\beta_0^d + \beta_1^d u)d(t) \tag{5}$$

where here we assume all the coefficients above are positive. Note net utilities w and v can be negative in this formulation, which produces a negative growth effect.

The interpretation of the coefficients β_0^s and β_0^d are as attrition rates of sellers and buyers, respectively. Specifically, $1/\beta_0^s$ and $1/\beta_0^d$ can be interpreted as the average life time of adopters on the supply and demand side. Consider the example where supplier surplus is zero, $w = 0$, so that suppliers are indifferent to staying in the market or exiting. In this case we have $\dot{s}(t)/s(t) = -\beta_0^s$, so the stock of supply adopters declines by a constant β_0^s percent per unit time, i.e. suppliers have a mean time to exit of $1/\beta_0^s$ and the stock of supply adopters decays at a rate β_0^s.

The coefficients β_1^s and β_1^d determine, respectively, how strongly supply and demand growth respond to the surplus generated for sellers and buyers. This can be seen by considering the case where buyers are persistent, so $\beta_0^d = 0$. In this case we have $\dot{d}(t)/d(t) = \beta_1^d u$, which means an increase of one dollar in per-buyer utility u leads to a β_1^d percent increase in demand side adopters per unit time. The value $1/\beta_1^d$ (resp. $1/\beta_1^d$) has units of dollars and can be interpreted as the *adoption cost*; that is, how much surplus must be provided to buyers (resp. sellers) to generate an incremental adopter on the demand (resp. supply) side.

In terms of micro-structure, one can think of this adoption growth as coming from word-of-mouth adoption of new users as in the "imitation effect" of the Bass model (Bass, 1969) (See also Kalish (1985) and Oren and Schwartz (1988)), more usage by the same population of users or some combination of both. In the first interpretation, the growth equations describe the number of adopters on each side of the market - keeping the usage frequency per adopter constant. Using the demand side as an example, in this interpretation buyers who successfully transact and pay a price p_d lower than their valuation v, have a positive experience that is either observed or communicated to potential new adopters. Hence, the adoption rate is positively related to the per-transaction surplus $(v - p_d)$ and the average number of transactions per current demand adopter $g(s, d)/d$. In the second interpretation, the growth functions describe the change in the usage frequency for a representative user, fixing the total number of users in the market. This aligns with the marketing literature in consumer loyalty; once a user has a positive experience (receives positive surplus) from purchasing a product, they are more likely to purchase it again due to brand-specific user skills, reduced quality uncertainty and reduced search costs (Wernerfelt, 1991). Under this loyalty interpretation of growth, previous successful (resp. negative) usage encourages (resp. discourages) future usage. Hence, the change in usage frequency is again positively related to the per-transaction surplus and the average number of transactions per user.

Substituting the values w and u above and simplifying, we can rewrite the supply and demand growth as

$$\dot{s}(t) = -\beta_0^s s(t) + \beta_1^s (p_s - c) g(s(t), d(t)) \tag{6}$$

$$\dot{d}(t) = -\beta_0^d d(t) + \beta_1^d (v - p_d) g(s(t), d(t)) \tag{7}$$

Note above that the growth of both supply and demand is increasing in total output $g(s(t), d(t))$.

3.3 Optimal growth formulation

We are next formulate the platform's optimal growth problem. Note that at any time t, the platform's profit rate is the price difference between the demand side and the supply side, $p_d - p_s$, multiplied by the number of successful transaction g. We assume the platform has a discount rate $\rho > 0$. The platform's objective is then to maximize its discounted aggregated profit:

$$\max_{p_d, p_s} \int_0^\infty e^{-\rho t} (p_d - p_s) g(s, d) dt \tag{8}$$

subject to the growth equations (6), (7). (There may also be some additional growth constraints which we will discuss later in detail.)

3.4 Market size-and-balance state-space reduction

We next show that this two-sided market model with two states, $s(t)$ and $d(t)$, can be reduced to an equivalent model with a single "market size" state variable. How this market size is allocated to the supply and demand side is then determined by a new "market balance" control variable.

This market size-and-balance representation is a state-space reduction, not merely a reformulation in the sense that market size is the only quantity necessary to determine both the rewards and the evolution of the system at any given time. It both simplifies and clarifies the analysis of optimal growth and balance. We next formally define the state-space reduction and then discuss the interpretation of the resulting market size and market balance variables.

3.4.1 State-space reduction

While the supply and demand stocks are separate state variables in our model, the following proposition establishes that these two states can be reduced to a single market size state provided prices are unconstrained:

Proposition 1 *If there are no constraints on prices p_s, p_d, then the state (s, d) can be reduced to a single-dimensional market size state:*

$$x \triangleq \frac{s}{\beta_1^s} + \frac{d}{\beta_1^d} \tag{9}$$

Proof. Suppose we are at state (s_0, d_0) at time τ. Then we will show that without cost we can instantaneously shift to any other state (s_1, d_1) that satisfies

$$x \equiv \frac{s_0}{\beta_1^s} + \frac{d_0}{\beta_1^d} = \frac{s_1}{\beta_1^s} + \frac{d_1}{\beta_1^d} \tag{10}$$

To see why, consider the pair of impulse price paths (price shocks) $p_s(t;\tau) = \frac{s_1 - s_0}{\beta_1^s g(s_0,d_0)}\delta(\tau - t)$ and $p_d(t;\tau) = \frac{d_0 - d_1}{\beta_1^d g(s_0,d_0)}\delta(\tau - t)$ where $\delta(\cdot)$ is the Dirac function defined by

$$\delta(x) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases} \text{ and } \int_0^\infty \delta(x)dx = 1$$

Define $s(\tau) = s(\tau^-) = s_0$. Then the stock of supply right after τ, i.e. $s(\tau^+)$ is given by

$$
\begin{aligned}
s(\tau^+) &= s_0 + \int_{\tau^-}^{\tau^+} \dot{s}(t)dt \\
&= s_0 + \int_{\tau^-}^{\tau^+} -\beta_0^s s(t) + \beta_1^s (p_s(t) - c) g(s(t), d(t)) dt \\
&= s_0 + \beta_1^s g(s_0, d_0) \int_{\tau^-}^{\tau^+} \frac{s_1 - s_0}{\beta_1^s g(s_0, d_0)} \delta(\tau - t) dt = s_0 + (s_1 - s_0) = s_1
\end{aligned}
$$

The second equality is by (6). An identical argument proves that d_1 can be instantly achieved by setting $p_d(t) = \frac{d_0 - d_1}{\beta_1^d g(s_0, d_0)} \delta(\tau - t)$. Now by (10), $\frac{s_1 - s_0}{\beta_1^s} = \frac{d_0 - d_1}{\beta_1^d}$. Hence, the total cost to the platform of the two price shocks is

$$
\int_{\tau^-}^{\tau^+} (p_s(t) - p_d(t)) g(s(t), d(t)) dt = \int_{\tau^-}^{\tau^+} \left(\frac{s_1 - s_0}{\beta_1^s g(s_0, d_0)} - \frac{d_0 - d_1}{\beta_1^d g(s_0, d_0)} \right) \delta(\tau - t) g(s(t), d(t)) dt = 0
$$

So the change from (s_0, d_0) and (s_1, d_1) is costless. □

Intuitively, the reason the state space collapses to a single market size variable is that the platform can use a cost-less price shock to arbitrarily shift the market balance at any point in time between supply and demand. These price shocks keep the total market size x fixed and are budget neutral, so they do not affect the profit of the platform. This means the platform's optimal growth problem (which we define formally below) reduces to one with a single state variable x together with a new *market balance* control variable

$$
\gamma = \frac{s}{\beta_1^s x}
$$

with $x\gamma = \frac{s}{\beta_1^s}$, $x(1-\gamma) = \frac{d}{\beta_1^d}$. The balance control γ the represents the fraction of the market size allocated to the supply versus demand side; a high value of γ (close to one) corresponds to a supply-heavy market balance, while a low value of γ (close to zero) corresponds to a demand-heavy market balance.

3.4.2 Interpretation of market size and balance

In the definition of market size x in (9), $1/\beta_1^s$ and $1/\beta_1^d$ are interpreted, respectively, as the "cost" of a unit of supply and demand and have units of dollars per unit of supply/demand. The market size x

therefore has units of dollars and can be interpreted as the "market value" or "level of investment". Indeed, one can show, using an argument similar to that in the proof of Proposition 1, that the platform can at any time choose to shock the market with extremely high demand prices and low supply prices such that it brings the market size from x down to 0 and in the process converts its user base to a cash value of x. In this sense, x can also also be thought of as the liquidation value of the market.

The market balance γ is the fraction of market value contained in the supply side while $1-\gamma$ is the fraction of market value contained in the demand side. In this sense, balance is a concept of relative value not relative quantity. For example, if suppliers are expensive to acquire and buyers are not (as measured by $1/\beta_1^s$ and $1/\beta_1^d$), then even if γ is close to one (a high concentration of total value in the supply side), this does not necessarily correspond to a high number of adopters on the supply side relative to adopters on the demand side.

Besides providing analytical simplification, this market size and balance formulation is extremely helpful in interpreting optimal growth. For one, having a scalar market size allows us to unambiguously identify one growth strategy as being "faster" than another if it produces a market size that is larger at all times (just as the concept of "distance" is fundamental to defining "speed" in physics). It also lets us identify important size thresholds. In particular, we show below that in the increasing returns case there is a critical market size such that growth must be subsidized if market size is below this critical size. And in the decreasing returns case, there is a market size at which the market becomes saturated and it is not optimal to grow beyond this size.

Likewise, the concept of balance is also quite useful and enables us to show precisely how the optimal allocation of size between supply and demand changes during the evolution of market growth. Indeed, in practice understanding market balance is of first-order practical importance when managing growth in two-sided markets like ride-sharing and our formulation highlights this important choice variable.

3.4.3 Output and growth equations as a function of market size and balance

Under this size-and-balance reformulation, the market output g can be written as a function of γ and x,

$$g(s,d) = g(\beta_1^s \gamma x, \beta_1^d (1-\gamma)x) = x^\alpha g(\beta_1^s \gamma, \beta_1^d (1-\gamma))$$

The rightmost equality is due to the fact that $g(s,d)$ is homogeneous of degree α. Define $h(\gamma) = g(\beta_1^s\gamma, \beta_1^d(1-\gamma))$. Then the market output is given by

$$g(\gamma, x) = x^\alpha h(\gamma) \tag{11}$$

The market size x has initial condition: $x(0) = \frac{s(0)}{\beta_1^s} + \frac{d(0)}{\beta_1^d} > 0$, and evolves according to the differential growth equation:

$$\dot{x} = \frac{\dot{s}}{\beta_1^s} + \frac{\dot{d}}{\beta_1^d} = -(\beta_0^s\gamma + \beta_0^d(1-\gamma))x + (v - c - \pi)g(\gamma, x) \tag{12}$$

where $\pi = p_d - p_s$ is the platform's *profit margin*. Note this implies that under this reformulation price also reduces to a one-dimensional control.

3.4.4 Recovering the original state variables and prices

The two-dimensional state $s(t)$,$d(t)$ and control $p_s(t), p_d(t)$ can be recovered from $x(t)$, $\gamma(t)$ (Proposition 14 and 15 in the Appendix). An important consequence of these propositions is that while the market size and balance decomposition assumes impulse price shocks can be used to choose the market balance at any point in time, if growth paths and balance paths are smooth, then so are prices. Moreover, we show below in Theorem 1 that the optimal balance itself is a continuous (and continuously differentiable) function of the market size $x(t)$. This means if a growth path $x(t)$ is continuous, then so is the optimal balance $\gamma(t)$, and therefore smooth growth paths imply smooth price paths.

3.5 Optimal growth reformulation

We next reformulate the optimal growth problem applying the size-and-balance reformulation:

$$\max \int_0^\infty e^{-\rho t} \pi g(\gamma, x) dt \tag{13}$$

subject to the state equation

$$\dot{x} = -(\beta_0^s\gamma + \beta_0^d(1-\gamma))x + (v - c - \pi)g(\gamma, x)$$

and the constraint on the market balance:

$$0 \leq \gamma \leq 1$$

with the initial condition:

$$x(0) > 0$$

Moreover, the growth path may be subject to additional constraints of the form

$$x(t) \in \mathcal{X}$$

where $\mathcal{X}$ is the set of all feasible growth paths with respect to those constraints.

Again, we consider this growth problem in the space of the market size trajectory $x(t)$ rather than in the space of the price policy. As noted, we show in the next section that the optimal balance at any time t is a function of the market size $x(t)$, and by Propositions 14 and 15, optimal prices in turn are uniquely determined by the market size and balance decisions. This means we can analyze and compare growth strategies based solely on analyzing their respective market size trajectories.

3.6 Definitions

Before proceeding, we introduce some useful terminology.

Definition 1 *$x(t)$ is an* increasing growth path *if $x(t_1) \geq x(t_2), \forall t_1 > t_2$. It is* strictly increasing *if $x(t_1) > x(t_2), \forall t_1 > t_2$.*

That is, a growth path is increasing if the market size is non-decreasing over time and strictly increasing if market size is strictly increasing with time. We also need:

Definition 2 *$x_1(t)$ is a* faster growth path *from $\underline{x}$ to $\bar{x}$ over $[\underline{t}, \bar{t}]$ than $x_2(t)$ if $\underline{x} = x_1(\underline{t}) = x_2(\underline{t})$, $\bar{x} = x_1(\bar{t}) = x_2(\bar{t})$, and $x_1(t) \geq x_2(t)$ at all $\underline{t} \leq t \leq \bar{t}$. It is* strictly faster *if $x_1(t) > x_2(t)$ for all $t \in (\underline{t}, \bar{t})$.*

This definition is useful when comparing two growth paths; if the market size in path 1 is at least as large as that in path 2 at all times, then path 1 is a faster growth path.

We say the market is *subsidized at time* t if $\pi(t) < 0$. That is, the platform pays the supply side a higher price than the price it charges the demand side when the market is subsidized. Growth may require subsidies. In an extreme case the platform may want to inject an instantaneous subsidy into the market to achieve rapid growth. Specifically, consider an impulse of subsidy at time t that moves the market size from $x(t^-) = x_0$ to $x(t^+) = x_1 > x_0$ of the form

$$\pi(t) = -\frac{x_1 - x_0}{g(\gamma_0, x_0)}\delta(t)$$

We call $\pi(t)$ a *subsidy shock*. The cost of the subsidy shock is

$$\int_{t^-}^{t^+} \pi(t) g(\gamma_0, x_0) dt = \int_{t^-}^{t^+} -(x_1 - x_0)\delta(t) dt = -(x_1 - x_0)$$

We call $m = x_1 - x_0$ the *magnitude* of the subsidy shock.

Definition 3 *A market is* viable *if there exists a feasible increasing growth path that generates a strictly higher total discounted profit in (13) than the path* $x(t) = x_0, \forall t$.

In other words, a market is viable if growing it strictly improves the total discounted profit.

We also define the following function which is important in our analysis:

Definition 4 $G(\gamma, x) = (v - c)g(\gamma, x) - (\rho + \beta_0^s \gamma + \beta_0^d (1 - \gamma))x$

Intuitively, $G(\gamma, x)$ is the instantaneous rate of net social welfare generated by being in state x with balance choice γ. That is, it is the rate of welfare generated by current transactions (the term $(v - c)g(\gamma, x)$) minus the loss in value due to the decline in market size x (the term proportional to x). This market value loss is the sum of the time discount rate ρ and the attrition rates per unit of supply and demand, $\beta_0^s \gamma + \beta_0^d (1 - \gamma)$, times the market size. Hence, $G(\gamma, x)$ is the net social welfare rate.

3.7 Assumptions

We impose the following regularity conditions on $h(\gamma)$:

Assumption 1 $h(\gamma)$ *is twice-differentiable and* $h(\gamma) > 0$ *on* $(0, 1)$*. Moreover,* $h(0) \geq 0$*,* $h(1) \geq 0$*, and both are finite.*

Remark 1 *$h(\gamma)$ is bounded for $\gamma \in [0, 1]$.*

Assumption 2 *$h(\gamma)$ is strictly concave in γ on $(0, 1)$, i.e. $h''(\gamma) < 0$.*

Assumption 3 *$h'(\gamma) = 0$ has a unique solution on [0,1].*

Assumption 4 *When $\alpha < 1$, $(1 - \alpha)h'(\gamma)^2 < -\alpha h(\gamma)h''(\gamma)$ for $\gamma \in (0, 1)$.*

We then have the following result on the constant elasticity of substitution (CES) function:

Lemma 1 *Suppose the output function $g(s, d)$ is the CES production function, i.e.*

$$g(s, d) = (\rho s^m + (1 - \rho)d^m)^{\frac{\alpha}{m}}, 0 \leq m < 1, 0 < a < 1$$

Then the production function satisfies Assumption 1-4 if $0 < \alpha < 1$ or $1 < \alpha < 1+\epsilon$, where $\epsilon > 0$ is sufficiently small. In particular, if $m = 0$, $g(s, d)$ follows the Cobb-Douglas function, and it satisfies Assumption 1-4 as long as $\alpha < \min\{\frac{1}{\rho}, \frac{1}{1-\rho}\}$.

4 Main Results

We next analyze the optimal growth policies for the infinite-horizon problems. We begin by showing that the optimal choice of market balance, γ, at any time t depends only on the market size $x(t)$; that is, in all cases there is a common optimal choice of market balance for any given market size. This optimal market balance reflects a fundamental trade-off between minimizing loss of market size due to attrition (durability) and maximizing adoption growth by maximizing market output. With this optimal balance result in hand, we then analyze the optimal growth paths themselves.

4.1 Optimal balance

First, note that maximizing the market output $g(\gamma, x) = x^\alpha h(\gamma)$ – which maximizes the surplus received by both sides of the market – is achieved by a balance γ such that $h'(\gamma) = 0$. We denote this balance choice as γ^* and call it the *output-maximizing balance*. Thus, $\gamma^* = (h')^{-1}(0)$. Our main result on optimal market balance is then (proof in the Appendix):

Theorem 1 (Optimal Balance) *For any level of market size $x(t) = x$, the optimal market balance $\gamma^*(x)$ is given by*

$$\gamma^*(x) = \max\Big\{\min\Big\{(h')^{-1}\big(\frac{(\beta_0^s - \beta_0^d)x^{1-\alpha}}{v-c}\big), 1\Big\}, 0\Big\} \tag{14}$$

Moreover, as the market size grows from zero to infinity, when the market exhibits decreasing returns to scale, if demand (supply) is more durable, i.e. $\beta_0^s > \beta_0^d$ ($\beta_0^s < \beta_0^d$), then $\gamma^*(x)$ is decreasing (increasing) from the output-maximizing balance γ^* to the durability-maximizing balance 0 (1); when the market exhibits increasing returns to scale, if demand (supply) is more durable, i.e. $\beta_0^s > \beta_0^d$ ($\beta_0^s < \beta_0^d$), then $\gamma^*(x)$ is increasing (decreasing) from the durability-maximizing balance 0 (1) to the output-maximizing balance γ^*.

In summary, as x increase from zero to infinity, we have

Optimal Balance $\gamma^*(x)$ as Market Size x Increases[5]

	$\alpha > 1$	$\alpha < 1$
$\beta_0^s > \beta_0^d$	$0 \to \gamma^*$	$\gamma^* \to 0$
$\beta_0^s < \beta_0^d$	$1 \to \gamma^*$	$\gamma^* \to 1$

To see the intuition behind these optimal balance results, note that there are two main components of supply and demand growth in our model (12): 1) loss in market size due to the supply and demand attrition coefficients β_0^s and β_0^d (the terms proportional to γ and $1-\gamma$ in (12)); and 2) growth in market size due to the rate of adoption (the term proportional to total market output $g(\gamma, x) = h(\gamma)x^\alpha$ in (12)). Attrition losses in market size are minimized by concentrating the market balance in the "most durable" side of the market: the supply side ($\gamma = 1$) when supply attrition is lower than demand attrition ($\beta_0^s < \beta_0^d$) and the demand side ($\gamma = 0$) when demand attrition is lower than supply attrition ($\beta_0^d < \beta_0^s$). On the other hand, adoption growth is maximized by maximizing market output, which maximizes the net utility received by both sides of the market. This maximum output is achieved by maximizing the term $h(\gamma)$ in the growth equation (12), which

[5] $\beta_0^s > \beta_0^d$: demand is more durable than supply, and vice versa; α is the total returns to scale; $0 \to \gamma^*$ means $\gamma(x)$ goes from 0 to the output-maximizing balance γ^* as x goes from 0 to ∞, and so on.

implies a balance of $\gamma = \gamma^*$. There is a tension between these two forces. And their relative importance is a function of both market size and returns to scale.

With increasing returns to scale and low market size x, market output relative to size is very small and attrition losses are the dominate factor influencing growth. Hence, the optimal balance is to concentrate investment in the most durable size of the market. As the size grows, increasing returns imply output per unit of size increases and hence for large size, adoption growth dominates attrition losses. Therefore, the optimal balance shifts to the output-maximizing balance.

With decreasing returns to scale the effects are the opposite. At low market size, output relative to market size is large and hence adoption growth dominate attrition losses and the optimal balance choice is to maximize output. But as the market grows, decreasing returns imply that output relative to market size decreases and hence attrition loss becomes the dominant factor influencing growth. Therefore, the optimal balance shifts to the durability-maximizing balance.

Importantly, this optimal balance result also implies that we can analyze growth policies by simply analyzing the market size path $x(t)$, since the optimal balance is determined by (14) once $x(t)$ is specified. Therefore, in the rest of the paper, we only consider policies that satisfy (14).

4.2 Stationary solution

We next define an important size threshold corresponding to the first-order conditions for our optimal control problem, namely:

Theorem 2 (Stationary Solution) *The stationary solution to the infinite-horizon problem (13) is*

$$x^* = \left\{ \frac{\beta_0^s \gamma^*(x^*) + \beta_0^d (1 - \gamma^*(x^*)) + \rho}{(v - c)\alpha h(\gamma^*(x^*))} \right\}^{\frac{1}{\alpha - 1}}$$

$$\gamma^*(x^*) = \max\left\{ \min\left\{ h'^{-1}\left(\frac{(\beta_0^s - \beta_0^d) {x^*}^{1-\alpha}}{v - c}\right), 1 \right\}, 0 \right\}$$

For example, when using Cobb-Douglas function $g(s,d) = As^{\alpha_s}d^{\alpha_d}$, the stationary solution is in closed-form:

$$x^* = \left\{ \frac{A(v-c)(\beta_1^s)^{\alpha_s}(\beta_1^d)^{\alpha_d}\alpha_s^{\alpha_s}\alpha_d^{\alpha_d}}{(\rho+\beta_0^d)^{1-\alpha_s}(\rho+\beta_0^s)^{1-\alpha_d}} \right\}^{\frac{1}{1-\alpha}} \{\alpha_s(\rho+\beta_0^d) + \alpha_d(\rho+\beta_0^s)\}$$
$$\gamma^*(x^*) = \frac{\alpha_s(\rho+\beta_0^d)}{\alpha_s(\rho+\beta_0^d)+\alpha_d(\rho+\beta_0^s)}$$

When $\alpha < 1$ (decreasing returns), one can show the Hamiltonian (shown in the appendix) is concave in x and hence this stationary solution is the unique steady-state solution to the infinite horizon problem (13). In this case, the stationary size x^* is a profit-maximizing market size, which we call the *saturation size*. When $\alpha > 1$ (increasing returns), the stationary solution above only characterizes a saddle point.

Therefore, a decreasing returns to scale market naturally stops growing once reaching x^*, while an increasing returns to scale market will not stop growing unless it reaches its size limit.

4.3 Optimal growth policy

We first show the optimal policy for an increasing returns to scale market with some growth constraints. We consider an upper bound on the market size, which we call the *market potential*, denoted as $\bar{x}$. We also consider an upper bound on the growth rate for a given market size, denoted as $f(x)$.

Theorem 3 *For an increasing returns to scale market, consider the infinite horizon problem (13) with the constraints*

$$\mathcal{X} = \{x(t)|x(0) = x_0, x(t) \le \bar{x}, 0 \le \dot{x} \le f(x)\}$$

where $f(x) > 0$ for $x > 0$. Let $F(t)$ be the solution to the differential equation $\dot{x} = f(x), x(0) = x_0$. Then the optimal policy is either to grow the market from x_0 to $\bar{x}$ as fast as possible, or not to grow at all, depending on whether the market potential $\bar{x}$ is sufficiently large relative to x_0. In particular: If $x^* < x_0 < \bar{x}$ or $x_0 < x^*, \bar{x} > \tilde{x}$, it is optimal to grow the market as fast as possible to $\bar{x}$, i.e.

$$x^*(t) = \begin{cases} F(t), & 0 \le t \le F^{-1}(\bar{x}) \\ \bar{x}, & t > F^{-1}(\bar{x}) \end{cases} \tag{15}$$

If $x_0 < x^*$, $x_0 < \bar{x} < \tilde{x}$, it is optimal not to grow the market, i.e. $x^*(t) = x_0, t \geq 0$, where $\tilde{x} > x_0$ is uniquely defined for any given $x_0 < x^*$ such that

$$\int_0^{F^{-1}(\tilde{x})} e^{-\rho t} G_x(\gamma^*(F(t)), F(t)) F'(t) dt = 0,$$

and x^* is the stationary market size defined in Theorem 2.

For a decreasing returns to scale market, the optimal policy is given in the following proposition:

Theorem 4 (Optimal Growth for Decreasing Returns) *Consider a market with decreasing returns to scale ($\alpha < 1$,) and an initial market size, x_0, that is below the saturation size ($x_0 < x^*$). Then for the infinite horizon problem (13) with $\mathcal{X} = \{x(t)|x(0) = x_0, x(t) \leq \bar{x}, 0 \leq \dot{x} \leq f(x)\}$, faster growth dominates slower growth getting from x_0 to $\min\{x^*, \bar{x}\}$, where x^* is the stationary market size from Theorem 2.*

In the decreasing returns case, the output per unit size is high when the market size is initially low, but as the market size increases, the output per unit size declines due to the decreasing returns. At some point, the optimal growth rate reaches zero. This stopping point is characterized by the steady-state solution given by Theorem 2 where x^* is the long-term profit-maximizing size. Once $x(t) = x^*$ is achieved, it is optimal to stay at x^*. Hence, we call x^* the market *saturation size* and the corresponding optimal balance $\gamma^*(x^*)$ the *saturation balance*.

Combining Theorem 4 and Theorem 5 together, we conclude that fast growth is generally optimal for both increasing and decreasing returns to scale markets:

Theorem 5 (Optimal Growth) *For the infinite-horizon problem (13), it is optimal to grow the market as fast as possible, given that the market is viable. Moreover, if the market has increasing returns to scale, the optimal growth path converges to the upper bound of the market size; if the market has decreasing returns to scale, the optimal growth path converges to the saturation size.*

Corresponding to the previous discussion on the stationary solution, the concavity/convexity of the objective function (13) in x leads to different long-run optimal sizes for decreasing/increasing returns to scale markets. Nevertheless, as long as there is sufficient potential in market growth, it is optimal to grow the market as fast as possible in both case. In the next section, we discuss the specific optimal growth policies in some real-world settings, and introduce conditions that identify a viable market in those settings.

4.3.1 An unconstrained market

We start with the simplest case without bounds on market size or the growth rates.

Proposition 2 *When the market exhibits increasing returns to scale, if $\mathcal{X}$ is the set of all increasing growth paths from x_0, then the optimization problem (13) is unbounded.*

In short, when there is no limiting market size in an increasing returns to scale market, the optimal growth policy results in a market size that is unbounded and generates infinite profit. As a practical matter, this result implies that real-world two-sided markets cannot indefinitely exhibit increasing returns to scale as they grow, else there would be opportunities for unbounded profit. Moreover, due to the scale effect, the output per unit size also becomes unbounded, which is also unrealistic. This is simply the result of a production model that cannot make physical sense in the limit of ever increasing market size.

A more surprising result is the decreasing returns case. Specifically, while a decreasing-returns market can be self-sufficient at low scale and could in fact grow to saturation size organically without subsidies, it is in fact still optimal to subsidize to achieve faster growth in this case. Indeed, the next result, which follows directly from Theorem 4, shows that when starting at an initial size $x_0 < x^*$, a strategy that produces faster growth from x_0 to x^* is always better.

Corollary 1 (Optimal Growth for Decreasing Returns) *If the market has decreasing returns to scale ($\alpha < 1$), the initial market size, x_0, is below the saturation size ($x_0 < x^*$) and prices are unconstrained, then the optimal growth strategy is to apply a subsidy shock of magnitude $m = x^* - x_0$ at time $t = 0$ to immediately grow the market to size x^*.*

If there is no constraint on the growth rate, the fastest grow path is simply a jump from x_0 to x^* at time 0, which can be achieved with a subsidy shock of magnitude $x^* - x_0$. As noted, this result is more surprising in the decreasing returns case because the market is profitable at low scale and therefore is able to grow from its initial size to the saturation size without subsidy. However, such organic growth comes at the cost of delaying the time to reach the saturation size and hence delaying the profits that come from operating at the long-run maximum size. As a result, subsidizing more rapid growth is still beneficial, despite not being strictly necessary.

In both the increasing and decreasing returns case, growth leads to higher profit. We conclude with the condition for a viable market in this case:

Proposition 3 *When there is no constraint on market growth or size, a decreasing returns to scale market is viable as long as $x_0 < x^*$, and an increasing returns to scale market is always viable.*

4.3.2 A market with finite market size

We next consider a market with finite market potential $\bar{x}$, but unconstrained growth rates. Moreover, to avoid trivialities, we will assume the initial market size x_0 to be sufficiently small ($x_0 < x^*$). Intuitively, a subsidy shock should still be optimal if the market is viable.

However, in contrast to the unconstrained case, an increasing returns to scale market is not always viable. Indeed, the market potential $\bar{x}$ needs to be sufficiently large so that the potential future reward can compensate for the cost of growth. In an increasing returns to scale market, growth is expensive because when the size is small, output relative to market size is low and hence the platform has to heavily incentivize growth through subsidies. If the market cannot recoup this subsidy cost by reaching a sufficient size, it may be unprofitable to grow. Specifically, define:

Definition 5 *The* critical size *is the smallest market size x such that $\dot{x} \geq 0$ and $\pi \geq 0$.*

The next proposition characterizes the critical size:

Proposition 4 *The critical size x_c is given by*

$$x_c = \left(\frac{\beta_0^s \gamma_c^\dagger + \beta_0^d (1 - \gamma_c^\dagger)}{(v - c) h(\gamma_c^\dagger)} \right)^{\frac{1}{\alpha - 1}} \tag{16}$$

$$\gamma_c^\dagger = \begin{cases} 0, & \frac{h'(0)}{h(0)} < \frac{\beta_0^s - \beta_0^d}{\beta_0^d} \\ \gamma_c, & \beta_0^s \frac{h'(1)}{h(1)} < \beta_0^s - \beta_0^d < \beta_0^d \frac{h'(0)}{h(0)} \\ 1, & \frac{h'(1)}{h(1)} > \frac{\beta_0^s - \beta_0^d}{\beta_0^s} \end{cases} \tag{17}$$

where γ_c uniquely defined by $(\beta_0^s \gamma_c + \beta_0^d (1 - \gamma_c)) \frac{h'(\gamma_c)}{h(\gamma_c)} = \beta_0^s - \beta_0^d$.

We next show that to grow to the critical size requires continuous subsidies:

Proposition 5 *When $\alpha > 1$, consider any market size $x < x_c$, then any increasing growth path from x to x_c must be subsidized at all times t along the path.*

This shows that when a platform's initial size is below the critical size, subsidies are the only growth option. Since no point on the growth path generates positive profit, one may expect that a market with a maximal market size below or at the critical size is not viable. The next corollary confirms this intuition:

Corollary 2 (Optimal Growth to the Critical Size) *For the infinite horizon problem (13) with $\alpha > 1$, x_0 sufficiently close to 0, and $x(t) \leq x_c$, the market is not viable, and it is optimal to keep $x(t) = x_0$.*

Growth to the critical size is a special case. Next, we analyze the optimal growth policy for a general upper bound $\bar{x}$:

Proposition 6 *Consider the infinite horizon problem (13) with $\alpha > 1$ and the constraint that $\mathcal{X} = \{x(t)|x(0) = x_0, x(t) < \bar{x}\}$. Define $\tilde{x}$ such that*

$$G(\gamma^*(x_0), x_0) = G(\gamma^*(\tilde{x}), \tilde{x}), \tilde{x} > x_0. \tag{18}$$

If the market is viable ($\bar{x} > \tilde{x}$), then it is optimal to apply a subsidy shock of magnitude $m = \bar{x} - x_0$ at time $t = 0$ to immediately grow the market to size $\bar{x}$. If the market is not viable ($\bar{x} < \tilde{x}$), then it is optimal to keep $x(t) = x_0$.

Note $\tilde{x}$ is the threshold size that determines whether an increasing returns to scale market is viable. The following lemma relates this minimal viable size to the critical size:

Lemma 2 *For an increasing returns to scale market, if the initial market size is sufficiently close to 0, then the critical market size $x_c < \tilde{x}$.*

From Proposition 6, the optimal policy to grow an increasing returns to scale market is either to grow it as fast as possible, or not to grow at all. From Proposition 6, whether the market is worth growing at all hinges on the market potential, $\bar{x}$, since $\bar{x}$ determines the long-term reward the platform can eventually collect once the growth phase completes.

When $\bar{x}$ is small, it is optimal to grow the market as slowly as possible, and the slowest growth path is to apply a subsidy shock at the terminal time. Since we are considering an infinite-horizon problem with free endpoints, the slowest path is effectively a path along which the market size never increases. The intuition is that in an increasing returns to scale market, when the market size is small, output per unit size is very low and hence attrition losses dominate surplus-driven growth. In this regime, it is very costly to maintain market size and hence the optimal policy is to maintain as small a size as long as possible. Indeed, for a small $\bar{x}$, the increase in the long-term reward for reaching $\bar{x}$ does not compensate for the cost of growing the market from x_0 to $\bar{x}$; hence, it is optimal not to grow the market at all.

Nevertheless, when $\bar{x}$ is large, meaning there is a positive reward to achieving the maximal market size, a slower growth path delays the time that the platform can collect that reward. In this case, the benefit from collecting it sooner dominates the loss from a higher subsidy cost. Therefore, for a $\bar{x}$ that is large enough, the optimal growth policy switches to a subsidy shock at time zero; that is, the fastest growth path possible.

While stylized, this result reflects the intuition in the ride-sharing industry about the merits of rapid growth. Ride-sharing is simply not economically viable at low density and low scale due to the long pickup times, which significantly increase the driver labor time per ride. To sustain such an inefficient scale through subsidies – or even slowly grow the market all the while it is at inefficient scale – is uneconomical. A large injection of funding to bring the market rapidly up to efficient scale is warranted. This is in fact the growth strategy that most ride-sharing companies have adopted.

As for the decreasing returns to scale market, since the market is efficient even at a small size and can grow organically, it is viable regardless of the value of the upper bound $\bar{x}$. Therefore, $\bar{x}$ only limits the long-run market size, but does not change the structure of the optimal policy.

Proposition 7 *Consider the infinite horizon problem (13) with* $\alpha < 1$ *and subject to the constraint that* $x(t) < \bar{x}$. *Then it is optimal to grow the market to* $\min\{\bar{x}, x^*\}$ *as fast as possible, i.e. by applying an impulse shock and bringing the market size from* x_0 *to* $\min\{\bar{x}, x^*\}$ *instantaneously.*

4.3.3 Returns to scale that change with market size

The optimal growth policies for finite and unconstrained market size can be pasted together to model a market that transitions from increasing to decreasing returns after growing beyond a given transition size $\bar{x}$ as follows: (We simply sketch out the idea since the details are straightforward.) By backward induction, first solve for the second-stage optimal growth policy by considering a decreasing returns optimal growth problem with initial condition $x(0) = \bar{x}$. The optimal value of this second-stage problem gives us the present value of the total reward for reaching the transition size $\bar{x}$. Next, solve an increasing returns to scale optimal growth problem with finite market size $\bar{x}$. The results of Proposition 6 then hold for this problem if we suitably modify the definition of viability to reflect that there is a fixed reward for reaching $\bar{x}$ equal to the present value of the optimized second-stage profit.

4.4 Finite relative growth rates

In the discussion thus, the growth rate itself was unconstrained. In reality, many factors could slow market growth rates. The first friction we consider is that there are typically decreasing marginal benefits to spending on subsidies that limit the relative growth rate one can achieve. We abstract this sort of friction as a bound on the maximal relative growth rate of the form:

$$\dot{x}(t) \leq bx(t) \tag{19}$$

In other words, we can subsidize and achieve growth according to our growth model (12), but once the relative growth rate $\dot{x}(t)/x(t)$ reaches b, further subsidies have no marginal effect on growth. We call (19) the *relative growth rate constraint.*

When (19) is binding, the market grows exponentially, i.e. $x(t) = x_0 e^{bt}$. This is the fastest growth path in subject to this constraint. Along this fastest path, the profit rate is given by

$$\pi(t)g(\gamma(t), x(t)) = (v - c)g(\gamma(t), x(t)) - (\beta_0^s \gamma(t) + \beta_0^d (1 - \gamma(t)) + b)x(t)$$

One can check that the optimal balance $\gamma^*(x(t))$ maximizes the profit rate, but does not affect the trajectory of the market size.

Again, we consider a case where the initial market size is small ($x_0 < x^*$). The optimal policy follows Theorem 3:

Proposition 8 *Consider the infinite horizon problem (13) with $\alpha > 1$ and subject to the constraints that $\mathcal{X} = \{x(t)|x(0) = x_0, \dot{x}/x \leq b, x(t) \leqslant \bar{x}\}$. If $\bar{x} > \tilde{x}_r$,where $\tilde{x}_r$ is uniquely defined by*

$$\int_0^{1/b \ln \frac{\tilde{x}_r}{x_0}} G_x(\gamma^*(x_0 e^{bt}), x_0 e^{bt}) x_0 e^{(b-\rho)t} dt = 0$$

then it is optimal to grow the market as fast as possible, and the optimal growth path is

$$x^*(t) = \begin{cases} x_0 e^{bt}, & t \leq \frac{1}{b} \ln \frac{\bar{x}}{x_0} \\ \bar{x}, & t > \frac{1}{b} \ln \frac{\bar{x}}{x_0} \end{cases}$$

Otherwise if $\bar{x} < \tilde{x}_r$, it is optimal to not grow the market at all, i.e. $x^*(t) = x_0, t \geq 0$.

For a decreasing returns to scale market, by Theorem 4, faster growth dominates slower growth from x_0 to $\min\{\bar{x}, x^*\}$. Therefore, the fastest growth path is still optimal. One interesting observation is that, in this case, the optimal policy may be profit-taking at a very early stage, then gradually shift to subsidizing, and eventually switch back to profit-taking at the end of market expansion. This is because the maximal growth rate $bx(t)$ is small at the beginning; the growth cannot further speed up even with subsidy. After the growth rate gradually picks up, subsidy is needed to expedite growth. In the long run, the market generates positive profit at the minimum of the saturation size x^* or the market potential $\bar{x}$.

The optimal policy follows Theorem 4:

Proposition 9 *Consider the infinite horizon problem (13) with $\alpha < 1$ and subject to the conditions that $\mathcal{X} = \{x(t)|x(0) = x_0, \dot{x}/x \leq b, x(t) \leqslant \bar{x}\}$. Then the optimal growth path is*

$$x^*(t) = \begin{cases} x_0 e^{bt}, & t \leq \frac{1}{b} \ln \frac{\min\{\bar{x}, x^*\}}{x_0} \\ \min\{\bar{x}, x^*\}, & t > \frac{1}{b} \ln \frac{\min\{\bar{x}, x^*\}}{x_0} \end{cases}$$

4.5 Fund raising limits

In the base model, we implicitly assumed that there is an abundant amount of subsidy budget available within a short period of time. In reality, subsidies must be funded and typically there are constraints on the ability to raise funds that are a function of the current market size. We analyze a stylized model of such a constraint next.

At time t, the profit rate is $\pi g(\gamma, x)$. If $\pi < 0$, it means the market is subsidized, and the platform is losing money at the rate of $\pi g(\gamma, x)$. Consider the following constraint:

$$- \pi(t)g(\gamma(t), x(t)) \leq mx(t) \tag{20}$$

Constraint (20) requires that the loss from subsidizing the two-sided market at time t should be at most a fraction m of the market size at that time; the larger the market, the larger the subsidy budget. This reflects the fact that in reality, subsidy budgets come from rounds of investment with investors willing to provide money as firms prove that they can grow their target markets. Thus, a larger market size justifies more funding. Hence, we model this using a multiple m of the market size $x(t)$ as the upper bound for the amount of funding available at a given time t. We call (20) the *fund-raising budget* and m the *funding multiple*.

By the state equation (12), $-\pi g(x, \gamma) \leq mx$ is equivalent to $\dot{x} \leq (v - c)g(\gamma, x) - (\beta_0^s\gamma + \beta_0^d(1 - \gamma) - m)x$. We require the funding multiple m not be too small (e.g. $m > max\{\beta_0^s, \beta_0^d\}$) such that growing to $\bar{x}$ from x_0 is feasible.

In contrast to the relative growth rate constraint, when the fund-raising budget (20) is binding, the market balance does not directly affect the profit rate $mx(t)$, but affects how fast the market grows. Thus, the optimal balance from Theorem 1 remains optimal here since it maximizes the growth rate for any given market size x. From this perspective, the fund-raising budget (20) can also be thought as an upper bound on the growth rate.

We show that the optimal policy has a similar structure as in the previous case:

Proposition 10 *Consider the infinite horizon problem (13) with $\alpha > 1$ and subject to the conditions that $\mathcal{X} = \{x(t)|x(0) = x_0, -\pi g(x, \gamma) \leq mx, x(t) \leq \bar{x}\}$, where $m > \max\{\beta_0^s, \beta_0^d\}$. Let $F(t)$ be* *the solution to the differential equation*

$$\dot{x} = (v-c)g(\gamma^*(x), x) - (\beta_0^s\gamma^*(x) + \beta_0^d(1-\gamma^*(x)) - m)x, x(0) = x_0 \tag{21}$$

If $\bar{x} > \tilde{x}_b$, it is optimal to grow the market as fast as possible to $\bar{x}$, i.e.

$$x^*(t) = \begin{cases} F(t), & 0 \leq t \leq F^{-1}(\bar{x}) \\ \bar{x}, & t > F^{-1}(\bar{x}) \end{cases} \tag{22}$$

If $\bar{x} < \tilde{x}_b$, it is optimal not to grow the market, i.e. $x^*(t) = x_0, t \geq 0$, where $\tilde{x}_b > x_0$ is uniquely defined such that

$$\int_0^{F^{-1}(\tilde{x}_b)} e^{-\rho t} G_x(\gamma^*(F(t)), F(t))F'(t)dt = 0 \tag{23}$$

Note that as the funding multiple m goes to infinity, $\dot{x}$ goes to infinity, and $F^{-1}(\bar{x})$ approaches 0. Then the optimal path (22) in Proposition 10 gradually approaches the optimal path in Proposition 6, i.e. $x(t)$ jumps from x_0 to $\bar{x}$ at time 0 and stays at $\bar{x}$. Recall that without a bound on the growth rate, an increasing returns to scale market is viable if the market size is above $\tilde{x}$. The following lemma shows that, the threshold size $\tilde{x}$ without the growth rate constraint is lower than that with the fund-raising constraint.

Lemma 3 *$\tilde{x}_b > \tilde{x}$, where $\tilde{x}_b$ is defined by (23) and $\tilde{x}$ is defined by (18).*

This comparison reveals that when the growth rate is bounded, a larger market potential $\bar{x}$ is required for the market to be profitable. Combining with the results in Proposition 10, this implies that the funding multiple m and the market potential $\bar{x}$ complements each other; if sufficient funds cannot be raised fast enough, then even if the market has potential profitability (i.e. $\bar{x} > \tilde{x}$), there is not enough momentum to jump start the growth. On the flip side, if the market limit $\bar{x}$ is small, even if the budget is abundant and allows impulse-like growth, it does not help because the potential profitability is too small. The intuition is that early growth is expensive in an increasing returns to scale market due to the low output-to-size ratio. Slow growth means that the market is stuck in an inefficient phase for even longer, and hence a large long-term reward is required to compensate for the high growth cost.

A decreasing returns to scale market, however, is self-sustaining even early on. Organic growth without subsidy is not only feasible but also generates a positive profit. Hence, the fund-raising budget (20) only affects how fast we can grow, but it does not change the decision about whether or not to grow:

Proposition 11 *Consider the infinite horizon problem (13) with $\alpha < 1$ and subject to the conditions that $\mathcal{X} = \{x(t)|x(0) = x_0, -\pi g(x,\gamma) \leq mx, x(t) \leq \bar{x}\}$, where $m > \max\{\beta_0^s, \beta_0^d\}$. $F(t)$ is the solution to the differential equation*

$$\dot{x} = (v-c)g(\gamma, x) - (\beta_0^s\gamma + \beta_0^d(1-\gamma) - m)x, x(0) = x_0 \tag{24}$$

Then the optimal growth path is

$$x^*(t) = \begin{cases} F(t), & t \leq F^{-1}(\min\{x^*, \bar{x}\}) \\ \min\{x^*, \bar{x}\}, & t > F^{-1}(\min\{x^*, \bar{x}\}) \end{cases}$$

4.6 Matching function

In the previous discussion, we required the production function $g(s, d)$ to satisfy certain smoothness conditions (Assumption 1). And we focused our attention on strictly increasing and decreasing returns to scale markets. A natural question then is: what growth policy is optimal in a constant returns to scale market?

The perfect matching function, $g(s, d) = A\min\{s, d\}$, is constant returns to scale in (s, d). It characterizes a frictionless market in which supply and demand are matched instantaneously as long as both inputs are positive. Although this function does not satisfy the smoothness conditions in Assumption 1, its simplicity allows us to derive the optimal policy directly without Theorem 1 and 5. Hence, we use this function as an example to characterize optimal growth for a constant returns to scale market.

Under the market size-and-balance reformulation, the production function can be written as $g(\gamma, x) = A\min\{\beta_1^s\gamma, \beta_1^d(1-\gamma)\}x$. As before, we show the optimal growth policies in the space of (γ, x).

Proposition 12 (Optimal Balance for a Perfect Matching Market) *In a perfect matching* *market, if*

$$-(v-c)A\beta_1^d < \beta_0^s - \beta_0^d < (v-c)A\beta_1^s \tag{25}$$

then the optimal balance is the output-maximizing balance, i.e.

$$\gamma^*(x) = \frac{\beta_1^d}{\beta_1^d + \beta_1^s} \tag{26}$$

Otherwise, the optimal balance is the durability maximizing balance, i.e.

$$\gamma^*(x) = \mathbb{1}_{\beta_0^s < \beta_0^d} \tag{27}$$

The fact that the optimal balance is constant is not surprising. Recall that by Theorem 1, the optimal balance is a monotone function in market size, and the direction of the change depends on the interplay between attrition and adoption. For a constant returns to scale market like the perfect matching market, however, the relation between attrition and adoption does not change with market size. Whichever force is stronger will always be stronger, and the optimal balance just maximizes that force. If condition (25) holds, adoption is stronger, and thus the optimal balance stays at the output-maximizing balance (26); if not, attrition is stronger, then the optimal balance stays at the durability maximizing balance (27).

Another way to understand (25) is from the perspective of redistributing the surpluses. Rewrite the condition as $A(v-c) > \max\left\{\frac{\beta_0^s-\beta_0^d}{\beta_1^s}, \frac{\beta_0^d-\beta_0^s}{\beta_1^d}\right\}$. $A(v-c)$ here is the maximal profit margin per transaction without external subsidy; $\frac{\beta_0^d-\beta_0^s}{\beta_1^d}$ $(\frac{\beta_0^s-\beta_0^d}{\beta_1^s})$ is the monetary value required to increase demand (supply) to meet supply (demand). If (25) does not hold, it means the market cannot maintain $s = d$ without subsidy. Such a market cannot be viable, because unlike an increasing returns to scale market, the output per unit size is fixed; growth will always require subsidy, and there is no future reward. Later we will show that (25) is indeed a necessary condition for a perfect matching market to be viable.

Next, we show that the optimal growth policy is similar to an increasing returns to scale market:

Proposition 13 (Optimal Growth Policy for a Perfect Matching Market) *In a perfect matching market, it is optimal to grow the market to its upper bound as fast as possible given that the market is viable. If not, it is optimal not to grow the market at all. Moreover, a market is viable* *if its parameters satisfy*

$$\rho < \frac{(v-c)A\beta_1^s\beta_1^d - (\beta_0^s\beta_1^d + \beta_0^d\beta_1^s)}{\beta_1^s + \beta_1^d} \tag{28}$$

Lemma 4 *Condition (28) implies condition (25).*

Hence, a viable market requires that adoption dominates attrition. Moreover, the discount factor ρ should be sufficiently small.

4.7 Summary of main results

Combining the result on the optimality of fast growth in an increasing, decreasing, and constant returns to scale market, our analysis suggests that subsidizing during all phases of market growth is warranted – not only during the initial phase of increasing returns in order to reach break-even size, but also later as markets mature and experience decreasing returns. Subsidizing growth in mature markets is desirable in order to reach the limiting saturation size more quickly. While we do not dismiss the fact that these findings might not be robust to changes to our particular specification of market output and supply and demand growth, the results nevertheless provide intriguing theoretical evidence that there may be more to the intense focus on subsidized growth in industries like ride-sharing than meets the eye. It suggests that highly subsidized growth is plausibly an economically optimal growth strategy and not just the result of "irrational enthusiasm" on the part of company founders and private-market investors.

5 Numerical example

To illustrate the results of our model, we next apply it to a stylized numerical example of a ride-sharing market with a Cobb-Douglas production function, $g(s,d) = As^{\alpha_s}d^{\alpha_d}$. Table 1 shows the parameters used in the numerical example. The unit of time is years, transactions are trips, and the stock of supply and demand is measured in terms of the number of riders and drivers on the platform. Riders have a value of \$18.30 per trip and drivers have a cost of \$3.75 per trip. Drivers have a mean lifetime on the platform of 2 years while riders have a mean lifetime of 5 years, so demand is more durable than supply. In terms of elasticity of growth to surplus, an increase of one dollar in the surplus for riders results in a 10% increase (per year) in the number of riders, while

Table 1: Parameters used in numerical examples

Parameter	Value	Notes
v	18.30 \$/trip	
c	3.75 \$/trip	
β_0^s	0.5 /year	2 year avg. life
β_0^d	0.2 /year	5 year avg. life
β_1^s	0.067 drivers/dollar	\$15 per driver adoption cost
β_1^d	0.100 riders/dollar	\$10 per rider adoption cost
ρ	0.15	15% per year
α_s	0.800	Incr. returns
α_d	0.400	
α	1.200	
A	0.05	
α_s	0.300	Decr. returns
α_d	0.286	
α	0.586	
A	$1,000$	

an increase in one dollar of surplus for drivers results in a increase of 6.7% increase (per year) in the number of drivers. This in turn implies an adoption cost of \$10 per rider and \$15 per driver.

Figures 1 and 2 illustrate the growth paths and total profits resulting from alternative (non-optimal) subsidy policies subject to fund-raising budget constraints. Figure 1 shows the growth paths and total discounted profit for the increasing returns case, comparing a high-subsidy policy of $m = 1$, a low-subsidy policy of $m = 0.2$, and an unsubsidized policy of $m = 0$, for the case where the market potential is $\bar{x} = \$100\text{M}$. The initial market size is $x_0 = \$1\text{M}$. Note in this case the high-subsidy policy yields a total discounted profit of $\$26,404,464$ and the low-subsidy policy yields a total discounted profit of $-\$782,162$, while the total discounted profit from the zero-subsidy policy is zero. Note the zero-subsidy policy is not able to reach the critical market size, which is $x_c = \$8,334,002$, and thus also cannot reach the market potential $\bar{x}$. For the low-subsidy policy, although the market is able to reach the market potential, the total discounted profit goes negative due to the prolonged inefficient phase. For comparison, the optimal discounted profit of the subsidy shock policy is $\$1,000,000 - \$100,000,000 + \$23,028,440/0.15 = \$54,522,933$, where $\$23,028,440$ is the profit rate when the market size is at it potential, $\bar{x}$ and 0.15 is the discount rate.

Figure 2 show a similar comparison of policies for the decreasing returns case, comparing a high-subsidy policy of $m = 1$ to a zero-subsidy policy of $m = 0$. The initial market size is $x_0 = 100$.

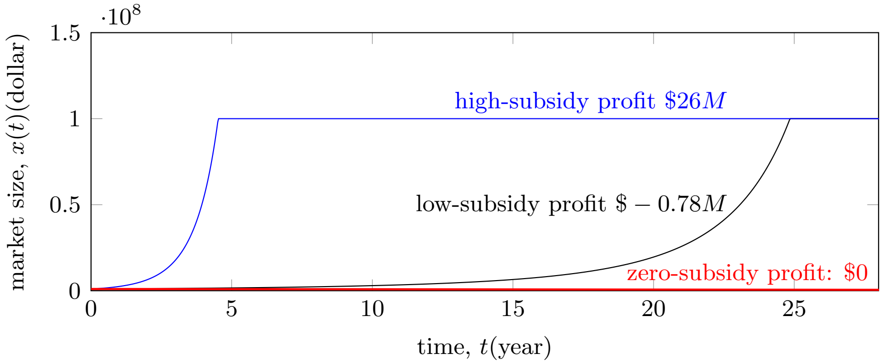

Figure 1: Policies and profits under fund-raising budgets (increasing returns to scale)
Note. High-subsidy policy: $m = 1$; low-subsidy policy: $m = 0.2$; zero-subsidy policy: $m = 0$; $\bar{x} = \$100M$; $x_0 = \$1M$

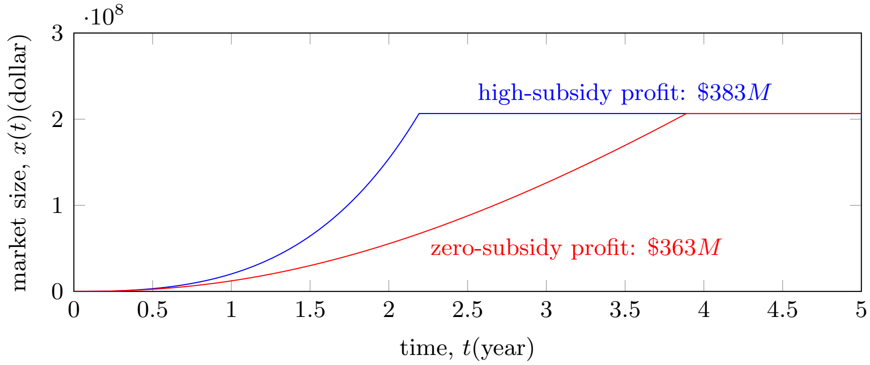

Figure 2: Policies and profits under fund-raising budgets (decreasing returns to scale)
Note. High-subsidy policy: $m = 1$; zero-subsidy policy: $m = 0$; $x_0 = 100$;

Here, the total discounted profit from the subsidy policy is $\$383,682,517$, while the total discounted profit from the zero-subsidy policy is $\$363,378,728$. Note in this case, both policies are able to reach the saturation size, which is $x^* = \$206,543,541$. However, the discounted total profit of the subsidized policy is greater. For comparison, the optimal discounted profit of the subsidy shock policy is $\$100 - \$206,543,541 + \$97,805,254/0.15 = \$445,491,589$, in which $\$97,805,254$ is the optimal profit rate and 0.15 is the discount rate.

Figures 3 and 4 show how optimal market balance changes as market size evolves in the numerical examples for increasing and decreasing returns to scale, respectively. Since demand is more durable than supply in this example, the durability-maximizing balance is $\gamma = 0$. The output-maximizing balance is $\alpha_s/(\alpha_s + \alpha_d)$, which is equal to 0.667 in the increasing returns case and 0.512 in the decreasing returns case.

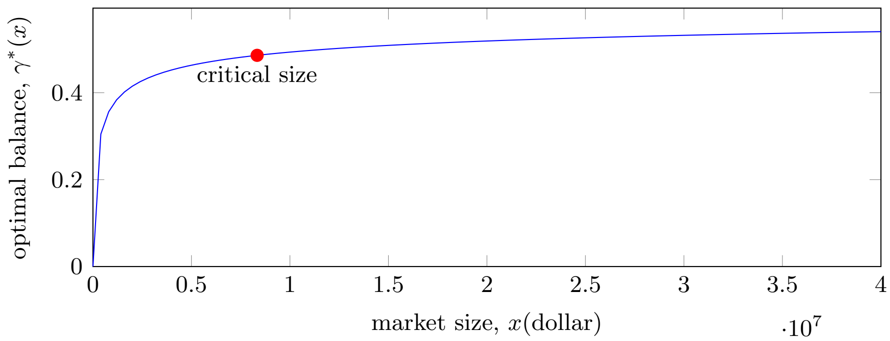

Figure 3: Evolution of optimal balance in the change of market size (increasing returns)

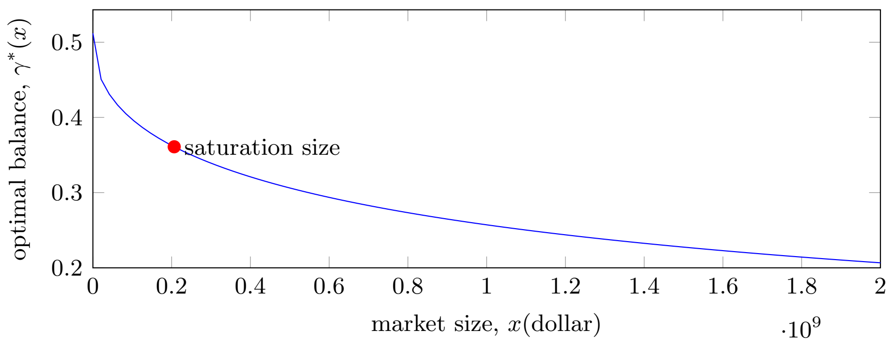

Figure 4: Evolution of optimal balance in the change of market size (decreasing returns)

As is shown in Figure 3, in the increasing returns case the market balance starts out at zero (the durability-maximizing balance) and increases toward the output-maximizing balance as size increases, albeit slowly once the size approaches the critical size $x_c = \$8,334,002$. Market balance at the critical size is approximately $\gamma = 0.49$.

In contrast, as shown in Figure 4, market balance in the decreasing returns case starts out at the output-maximizing balance of $\gamma = 0.512$ when the market size is small and decreases toward zero (the durability-maximizing balance) as the market grows. Figure 4 shows the optimal balance is approximately $\gamma = 0.37$ when the market reaches the saturation size of $x^* = \$206,543,541$.

6 Conclusion

Our model and results provide a theoretical framework to understand optimal growth in two-sided markets. The state-space reduction to a scalar market size together with a market balance control provide both conceptual clarity and analytical tractability that should prove useful when analyzing other variations of the growth problem.

Our results on optimal balance provide useful insights into the factors that should be considered when deciding how to invest in the supply or demand sides of the market. In particular, it highlights the importance of the attrition (durability) of supply and demand, the adoption cost of generating supply and demand, and the elasticity of output produced by supply and demand – and how these three forces evolve in relative importance as markets grow.

We also characterize important size thresholds: the critical size at which a market becomes self-sustaining (increasing returns) and the saturation size at which optimal growth reaches its limit (decreasing returns). It would be interesting to map the predictions of our model empirically to determine what critical and saturation market sizes look like in real-world markets.

Lastly, we showed that in general faster growth is better and that subsidy shocks are optimal, both in early phases of growth when returns to scale are increasing and in later phases of growth when returns to scale are decreasing. The findings provide theoretical evidence of the value of bold subsidized growth strategies in two-sided markets – even when platforms have a monopoly and do not face competitive entry threats. It would be interesting to see if these extreme subsidy strategies hold up under other modeling assumptions and to what extent there is empirical evidence supporting their optimality.

Extending our theory of optimal growth to competing two-sided platforms is another worthwhile extension. It is important to understand to what extent the results here are robust to environments in which two or more platforms compete. Toward that end, we have work in process on an oligopoly version of our growth model, which is forthcoming.

Appendix

Proof of Lemma 1

Proof. By the definition of h,

$$h(\gamma) = (\rho\gamma^m(\beta_1^s)^m + (1-\rho)(1-\gamma)^m(\beta_1^d)^m)^{\frac{\alpha}{m}}$$

The smoothness can be checked by taking the derivative of $h(\gamma)$. Moreover, given that $m \geq 0$, $h(\gamma)$ is defined on $\gamma = 0$ and $\gamma = 1$. Hence, Assumption 1 is confirmed.

For Assumption 4, Denote $l(\gamma) = \rho\gamma^m(\beta_1^s)^m + (1-\rho)(1-\gamma)^m(\beta_1^d)^m$. $l(\gamma) > 0$. Then

$$h'(\gamma) = \frac{\alpha}{m} l(\gamma)^{\frac{\alpha}{m}-1} l'(\gamma) \tag{29}$$

$$h''(\gamma) = \frac{\alpha}{m}\left(\frac{\alpha}{m} - 1\right) l(\gamma)^{\frac{\alpha}{m}-2} l'(\gamma)^2 + \frac{\alpha}{m} l(\gamma)^{\frac{\alpha}{m}-1} l''(\gamma) \tag{30}$$

Then

$$\begin{aligned}
&(1-\alpha)h'(\gamma)^2 + \alpha h(\gamma)h''(\gamma) \\
&= \frac{\alpha^2}{m^2} l(\gamma)^{\frac{2\alpha}{m}-2} l'(\gamma)^2 - \frac{\alpha^3}{m^2} l(\gamma)^{\frac{2\alpha}{m}-2} l'(\gamma)^2 + l(\gamma)^{\frac{2\alpha}{m}-2}\left(\frac{\alpha^3}{m^2} - \frac{\alpha^2}{m}\right) l'(\gamma)^2 + \frac{\alpha^2}{m} l(\gamma)^{\frac{2\alpha}{m}-1} l''(\gamma) \\
&= \frac{\alpha^2}{m^2} l(\gamma)^{\frac{2\alpha}{m}-2} l'(\gamma)^2 + l(\gamma)^{\frac{2\alpha}{m}-2}\left(-\frac{\alpha^2}{m}\right) l'(\gamma)^2 + \frac{\alpha^2}{m} l(\gamma)^{\frac{2\alpha}{m}-1} l''(\gamma) \\
&= \frac{\alpha^2}{m} l(\gamma)^{\frac{2\alpha}{m}-2}\left\{\left(\frac{1}{m} - 1\right) l'(\gamma)^2 + l(\gamma) l''(\gamma)\right\}
\end{aligned}$$

To check the sign of $\left(\frac{1}{m} - 1\right) l'(\gamma)^2 + l(\gamma)l''(\gamma)$,

$$l'(\gamma) = \rho(\beta_1^s)^m m\gamma^{m-1} - (1-\rho)(\beta_1^d)^m m(1-\gamma)^{m-1}$$

$$l''(\gamma) = \rho(\beta_1^s)^m m(m-1)\gamma^{m-2} + (1-\rho)(\beta_1^d)^m m(m-1)(1-\gamma)^{m-2}$$

By some algebraic manipulation,

$$\left(\frac{1}{m} - 1\right) l'(\gamma)^2 + l(\gamma)l''(\gamma) = (m-1)m\rho(1-\rho)(\beta_1^s)^m(\beta_1^d)^m\gamma^{m-2}(1-\gamma)^{m-2}$$

Then

$$(1-\alpha)h'(\gamma)^2 + \alpha h(\gamma)h''(\gamma) = \alpha^2 l(\gamma)^{\frac{2\alpha}{m}-2}(m-1)\rho(1-\rho)(\beta_1^s)^m(\beta_1^d)^m\gamma^{m-2}(1-\gamma)^{m-2}$$

Since $m < 1$, $(1-\alpha)h'(\gamma)^2 + \alpha h(\gamma)h''(\gamma) < 0$. Hence Assumption 4 is confirmed.

For Assumption 3, setting $h'(\gamma) = 0$ gives

$$\gamma = \frac{1}{1 + \left(\frac{\rho}{1-\rho}(\frac{\beta_1^s}{\beta_1^d})^m\right)^{\frac{1}{m-1}}}$$

Since $\left(\frac{\rho}{1-\rho}(\frac{\beta_1^s}{\beta_1^d})^m\right)^{\frac{1}{m-1}} > 0$, the solution of γ is always in the range of $(0,1)$.

For Assumption 2, we have shown in (30) that

$$h''(\gamma) = \frac{\alpha}{m} l(\gamma)^{\frac{\alpha}{m}-2}\left((\frac{\alpha}{m}-1)l'(\gamma)^2 + l(\gamma)l''(\gamma)\right)$$

Since $l(\gamma)$ does not contain any term related to α, for $m > 0$, one can show that $(\frac{\alpha}{m}-1)l'(\gamma)^2 + l(\gamma)l''(\gamma)$ is increasing and continuous in α. If $\alpha < 1$,

$$(\frac{\alpha}{m}-1)l'(\gamma)^2 + l(\gamma)l''(\gamma) < (\frac{1}{m}-1)l'(\gamma)^2 + l(\gamma)l''(\gamma) < 0$$

If $\alpha > 1$, by continuity, there always exists $\epsilon > 0$ that is sufficiently small, such that

$$(\frac{1}{m}-1)l'(\gamma)^2 + l(\gamma)l''(\gamma) < (\frac{1+\epsilon}{m}-1)l'(\gamma)^2 + l(\gamma)l''(\gamma) < 0$$

Therefore, if $1 < \alpha \le 1+\epsilon$, $h''(\gamma) < 0$ still holds.

For $m = 0$, $h(\gamma) = \gamma^{\alpha\rho}(1-\gamma)^{\alpha(1-\rho)}$, which is the Cobb-Douglas function. The proof for Assumption 1 to 4 still hold by taking the limit of $m \to 0$. In particular, for Assumption 2, it can be shown that $h''(\gamma) < 0$ for any $\alpha < \min\{1/\rho, 1/(1-\rho)\}$, given $m = 0$:

$$\begin{aligned}
&\lim_{m\to 0} \frac{\alpha}{m} l'(\gamma) = \lim_{m\to 0} (\frac{\alpha}{m} - 1) l'(\gamma) = \alpha(\rho\gamma^{-1} - (1-\rho)(1-\gamma)^{-1}) \\
&\lim_{m\to 0} \frac{\alpha}{m} l''(\gamma) = \alpha\{-\rho\gamma^{-1} - (1-\rho)(1-\gamma)^2\} \\
&\lim_{m\to 0} l(\gamma)^{\frac{\alpha}{m}-2} = \gamma^{\alpha\rho}(1-\gamma)^{\alpha(1-\rho)}
\end{aligned}$$

Hence,

$$\begin{aligned}
\lim_{m\to 0} h''(\gamma) &= \gamma^{\alpha\rho}(1-\gamma)^{\alpha(1-\rho)} \left((\frac{\alpha\rho}{\gamma} - \frac{\alpha(1-\rho)}{1-\gamma})^2 - \frac{\alpha\rho}{\gamma^2} - \frac{\alpha(1-\rho)}{(1-\gamma)^2} \right) && (31) \\
&= \gamma^{\alpha\rho}(1-\gamma)^{\alpha(1-\rho)} \left(\frac{\alpha\rho(\alpha\rho - 1)}{\gamma^2} + \frac{\alpha(1-\rho)(\alpha(1-\rho)-1)}{(1-\gamma)^2} - \frac{2\rho(1-\rho)\alpha^2}{\gamma(1-\gamma)} \right) && (32)
\end{aligned}$$

Since $\alpha\rho < 1$, $\alpha(1-\rho) < 1$, (31) is negative. Hence, $h''(\gamma) < 0$. □

Proof of Lemma 2

Proof. We first show that $G(\gamma^*(x_c), x_c) < 0$. By Theorem 1, $\gamma^*(x_c) = \gamma_c$, and

$$-(\beta_0^s \gamma_c + \beta_0^d (1-\gamma_c)) x_c + (v-c) h(\gamma_c) x_c^\alpha = 0$$

Therefore, by Lemma 5,

$$G(\gamma^*(x_c), x_c) = -\rho x_c < 0$$

Since $\alpha > 1$,

$$\lim_{x\to 0} G(\gamma^*(x), x) = 0$$

By Lemma 6, $\gamma^*(x)$ is continuous. By Assumption 1, $h(\gamma)$ is continuous. Hence, $G(\gamma^*(x), x)$ is continuous too. Then there always exists an $x_0 > 0$ sufficiently small such that

$$G(\gamma^*(x_c), x_c) < G(\gamma^*(x_0), x_0) = G(\tilde{x})$$

Then for such an x_0, it must be true that $x_c < \tilde{x})$, since for any $x > \tilde{x}$, $G(\gamma^*(x), x))$ is increasing in x by the definition of $\tilde{x}$. □

Proof of Lemma 3

Proof. Integrating by parts and (18) gives

$$\begin{aligned}
&\int_0^{F^{-1}(\tilde{x})} e^{-\rho t} G_x(\gamma^*(F(t)), F(t))F'(t)dt \\
&= \rho \int_0^{F^{-1}(\tilde{x})} e^{-\rho t} G(\gamma^*(F(t)), F(t)) - G(\gamma^*(x_0), x_0) + e^{-\rho F^{-1}(\tilde{x})} G(\gamma^*(\tilde{x}), \tilde{x}) \\
&= \int_0^{F^{-1}(\tilde{x})} \rho e^{-\rho t} \left(G(\gamma^*(F(t)), F(t)) - G(\gamma^*(x_0), x_0) \right) dt
\end{aligned}$$

Since $G(\gamma^*(x_0), x_0) = G(\gamma^*(\tilde{x}), \tilde{x})$, by convexity of G, $G(\gamma^*(x), x) < \gamma^*(x_0), x_0)$ for all $x \in [x_0, \tilde{x}]$. Hence,

$$\int_0^{F^{-1}(\tilde{x})} e^{-\rho t} G_x(\gamma^*(F(t)), F(t))F'(t)dt < 0$$

Then it is implied that

$$\int_0^{F^{-1}(\tilde{x})} e^{-\rho t} G(\gamma^*(F(t)), F(t)) + \frac{e^{-\rho F^{-1}(\tilde{x})}}{\rho} G(\gamma^*(\tilde{x}), \tilde{x}) < \frac{1}{\rho} G(\gamma^*(x_0), x_0)$$

Hence, $\tilde{x} < \tilde{x}_b$. □

Proof of Lemma 4

Proof. Since $\rho > 0$, (28) implies that $(v - c)A\beta_1^s\beta_1^d > \beta_0^s\beta_1^d + \beta_0^d\beta_1^s$. Then

$$(v - c)A\beta_1^s\beta_1^d > \beta_0^s\beta_1^d + \beta_0^d\beta_1^s > \max\{\beta_0^s\beta_1^d, \beta_0^d\beta_1^s\} > \max\{(\beta_0^s - \beta_0^d)\beta_1^d, (\beta_0^d - \beta_0^s)\beta_1^s\}$$

Dividing each term by $\beta_1^s\beta_1^d$ and rearranging the terms give (25). □

Lemma 5 *For any finite trajectory $x(t)$, the following equation holds:*

$$\int_0^T e^{-\rho t} \pi g(\gamma, x)dt = \int_0^T e^{-\rho t} G(\gamma, x)dt - e^{-\rho T} x(T) + x(0)$$

Proof. By Definition 4,

$$G(\gamma, x) = (v - c)g(\gamma, x) - (\rho + \beta_0^s\gamma + \beta_0^d(1 - \gamma))x$$

By the formulation of $\dot{x}$,

$$\pi g(\gamma, x) = (v - c)g(\gamma, x) - (\beta_0^s \gamma + \beta_0^d (1 - \gamma))x - \dot{x}$$

Then

$$\int_0^T e^{-\rho t} \pi g(\gamma, x) dt = \int_0^T ((v - c)g(\gamma, x) - (\beta_0^s \gamma + \beta_0^d (1 - \gamma))x) e^{-\rho t} dt - \int_0^T \dot{x} e^{-\rho t} dt$$

Integration by parts gives

$$\int_0^T \dot{x} e^{-\rho t} dt = \int_0^T x(t) \rho e^{-\rho t} dt + e^{-\rho T} x(T) - x(0)$$

Therefore,

$$\int_0^T e^{-\rho t} \pi g(\gamma, x) dt = \int_0^T ((v - c)g(\gamma, x) - (\rho + \beta_0^s \gamma + \beta_0^d (1 - \gamma))x) e^{-\rho t} dt - e^{-\rho T} x(T) + x(0)$$

□

Lemma 6 *Under Assumption 2 and 4, $G(\gamma, x)$ is strictly concave in γ. Moreover, if $\alpha < 1$, then $\max_\gamma G(\gamma, x)$ is strictly concave in x and reaches maximum at $x = x^*$; if $\alpha > 1$, then $\max_\gamma G(\gamma, x)$ is strictly convex in x and reaches global minimum at $x = x^*$.*

Proof. By the concavity of $h(\gamma)$ (Assumption 2),

$$\frac{\partial^2 G(\gamma, x)}{\partial \gamma^2} = (v - c)h''(\gamma)x^\alpha < 0$$

Then $\arg\max_\gamma G(\gamma, x)$ can be obtained by setting

$$\frac{\partial G(\gamma, x)}{\partial \gamma} = -(\beta_0^s - \beta_0^d)x + (v - c)h'(\gamma)x^\alpha = 0$$

Note that $h'(\gamma)$ is bounded by $h'(0)$ and $h'(1)$, due to $\gamma \in [0, 1]$. Thus, the optimal γ can be solved by

$$\gamma^*(x) = \max\left\{\min\{(h')^{-1}(\frac{(\beta_0^s - \beta_0^d)x^{1-\alpha}}{v-c}), 1\}, 0\right\} \tag{33}$$

Then for those x that $\gamma^*(x)$ has an interior solution,

$$\frac{d\gamma^*(x)}{dx} = \frac{(\beta_0^s - \beta_0^d)(1-\alpha)}{(v-c)h''(\gamma)x^\alpha}$$

Since $h''(\gamma) < 0$,

$$\mathrm{sgn}(\frac{d\gamma^*(x)}{dx}) = \mathrm{sgn}((\alpha - 1)(\beta_0^s - \beta_0^d)) \tag{34}$$

By the envelope theorem,

$$\frac{\partial G(\gamma^*(x), x)}{\partial x} = \frac{\partial G(\gamma, x)}{\partial x}\Big|_{\gamma=\gamma^*(x)} = -(\rho + \beta_0^s\gamma^*(x) + \beta_0^d(1-\gamma^*(x))) + (v-c)\alpha h(\gamma^*(x))x^{\alpha-1}$$

When (33) has an interior solution, plugging in (33)gives

$$\frac{\partial G(\gamma^*(x), x)}{\partial x} = \alpha(\beta_0^s - \beta_0^d)\frac{h(\gamma^*(x))}{h'(\gamma^*(x))} - (\beta_0^s\gamma^*(x) + \beta_0^d(1-\gamma^*(x)) + \rho)$$

Then

$$\frac{\partial G^2(\gamma^*(x), x)}{\partial x^2} = (\beta_0^s - \beta_0^d)\left(\frac{\alpha h'(\gamma)^2 - \alpha h(\gamma)h''(\gamma)}{h'(\gamma)^2} - 1\right)\frac{d\gamma^*(x)}{dx}$$

By (34), $\mathrm{sgn}(\frac{\partial G^2(\gamma^*(x),x)}{\partial x^2}) = \mathrm{sgn}((\alpha-1)((\alpha-1)h'(\gamma)^2 - \alpha h(\gamma)h''(\gamma)))$. When (33) does not have an interior solution, $\mathrm{sgn}(\frac{\partial G^2(\gamma^*(x),x)}{\partial x^2}) = \mathrm{sgn}(\alpha - 1)$.

Therefore, if $\alpha > 1$, $G(\gamma^*(x), x)$ is strictly convex; if $\alpha < 1$, $G(\gamma^*(x), x)$ is strictly concave under Assumption 4.

Setting $\frac{\partial G(\gamma^*(x),x)}{\partial x} = 0$ gives

$$x = \left(\frac{\rho + \beta_0^s\gamma^*(x) + \beta_0^d(1-\gamma^*(x))}{(v-c)\alpha h(\gamma^*(x))}\right)^{\frac{1}{\alpha-1}}$$

The solution to the above equation is γ^*, and the corresponding market size x is x^*. To check the existence of x^*, it can be verified that $\lim_{x\to 0} G_x(\gamma^*(x), x))$ and $\lim_{x\to+\infty} G_x(\gamma^*(x), x)$ have opposite signs. Since $G_x(\gamma^*(x), x)$ is monotone and continuous, $G_x(\gamma^*(x), x) = 0$ must have a unique solution.

Therefore, x^* is the global minimum (maximum) for $\max_\gamma G(\gamma, x)$ under $\alpha > 1$ ($\alpha < 1$). $\square$

Lemma 7 *Any increasing growth path from x_0 to $\bar{x}$ is weakly dominated by*

$$x(t; t_i) = \begin{cases} x_0, & t \leq t_i \\ F(t - t_i), & t_i < t \leq t_i + F^{-1}(\bar{x}) \\ \bar{x}, & t_i + F^{-1}(\bar{x}) < t \leq T \end{cases} \tag{35}$$

where $0 \leq t_i \leq T - F^{-1}(\bar{x})$.

Proof. If $x_0 < \bar{x} < x^*$, by Theorem 6, it is optimal to grow the market as slow as possible. In this case, the slowest growth path is given by

$$x_s(t) = x(t; T - F^{-1}(\bar{x})) \tag{36}$$

To see why, suppose there is another increasing growth path $y(t)$ from x_0 to $\bar{x}$ over $[0, T]$ that is admissible and not faster than $x_s(t)$ given here. By definition, there exists a time point $t' \in [0, T]$ such that $y(t') < x_s(t')$. if $t' \leqslant T - F^{-1}(\bar{x})$, then $y(t') < x_s(t') = x_0$. This can't be true because $y(t)$ is an increasing growth path, and thus $y(t) \geqslant x_0$. If $t' > T - F^{-1}(\bar{x})$, then $y(t') < F(t' - T + F^{-1}(\bar{x}))$. Then $F^{-1}(y(t')) < t' - T + F^{-1}(\bar{x})$. The shortest time it takes to grow from $y(t')$ to $\bar{x}$ follows

$$F^{-1}(\bar{x}) - F^{-1}(y(t')) > T - t'$$

Therefore, $y(t)$ cannot reach $\bar{x}$ before or at $t = T$. Contradiction.

Similarly, if $x_0 > x^*$, again by Theorem 6, it is optimal to grow the market as fast as possible. The fastest growth path is given by

$$x_f(t) = x(t; 0) \tag{37}$$

The proof is similar to that for the slowest growth path (36).

For $x_0 < x^* < \bar{x}$, again consider an increasing growth path $y(t)$ from x_0 to $\bar{x}$ over $[0, T]$ that is admissible but doesn't satisfy (35). Since the growth rate is bounded, $y(t)$ is continuous, and thus must cross x^*. Denote the time $y(t) = x^*$ as $t_{y=x^*}$. We construct a the following growth path:

$$x_{s-f}(t) = x(t; t_{y=x^*}) \tag{38}$$

One can check that this growth path is the slowest from x_0 to x^* over $[0, t_{y=x^*}]$ and the fastest from x^* to $\bar{x}$ over $[t_{y=x^*}, T]$ using similar arguments for proving (36). $\square$

Proof of Proposition 2

Proof. We will construct a feasible policy with unbounded value. Denote the initial market size as x_0. Fix $\gamma = 0.5$ and $\pi = 0.5(v - c)$. Apply an impulse to instantly increase the market size from $x(0)$ to $\tilde{x}$, where $\tilde{x} > (\frac{\beta_0^s + \beta_0^d}{(v-c)h(0.5)})^{\frac{1}{\alpha-1}}$. Then we can directly obtain the expression of $x(t)$ by integrating the differential equation:

$$\dot{x} = -0.5(\beta_0^s + \beta_0^d)x + 0.5(v - c)h(0.5)x^\alpha$$

which gives

$$x(t) = \left(\frac{0.5(\beta_0^s + \beta_0^d)}{0.5(v - c)h(0.5) - C_0 e^{0.5(\beta_0^s + \beta_0^d)(\alpha - 1)t}} \right)^{\frac{1}{\alpha - 1}}$$

$C_0 = 0.5(v - c)h(0.5) - 0.5(\beta_0^s + \beta_0^d)\tilde{x}^{1-\alpha}$. Then for this policy, x goes to infinity as t approaches $\frac{1}{0.5(\alpha-1)(\beta_0^s+\beta_0^d)} \ln(0.5(v - c)h(0.5)/C_0)$. Since the integrand becomes unbounded in a finite amount of time, the discounted objective value is unbounded. $\square$

Proof of Proposition 3

Proof. For $\alpha > 1$, by Proposition 2, there exists a feasible growth path that leads to unbounded profit, while keeping $x(t) = x_0$ generates finite profit. Thus, an increasing returns to scale market is viable.

For $\alpha < 1$, if $x_0 < x^*$, then by Lemma 6, $G(\gamma^*(x), x) > G(\gamma^*(x_0), x_0)$ for all $x_0 < x \leq x^*$. Hence, by Lemma 5, any increasing growth path from x_0 to x^* generates a higher profit than $x(t) = x_0$. $\square$

Proof of Proposition 4

Proof. Prove by contradiction. Suppose there exists some $\tilde{x} < x_c$ such that $\dot{x} \geq 0$ and $\pi \geq 0$. When the market size is $\tilde{x}$ and $\dot{x} \geq 0$, by (12),

$$-(\beta_0^s\gamma + \beta_0^d(1-\gamma))\tilde{x} + (v - c - \pi)h(\gamma)\tilde{x}^\alpha \geq 0$$

Since $h(\gamma) > 0$,

$$\pi \leq v - c - \frac{\beta_0^s\gamma + \beta_0^d(1-\gamma)}{h(\gamma)\tilde{x}^{\alpha-1}}$$

Maximize the right-hand side over γ on $[0, 1]$:

$$\left(v - c - \frac{\beta_0^s\gamma + \beta_0^d(1-\gamma)}{h(\gamma)\tilde{x}^{\alpha-1}}\right)_\gamma = \frac{-(\beta_0^s - \beta_0^d)h(\gamma) + (\beta_0^s\gamma + \beta_0^d(1-\gamma))h'(\gamma)}{h(\gamma)^2\tilde{x}^{\alpha-1}} \tag{39}$$

Setting the numerator of (39) to 0 gives

$$\frac{h'(\gamma)}{h(\gamma)}(\beta_0^s\gamma + \beta_0^d(1-\gamma)) = \beta_0^s - \beta_0^d \tag{40}$$

When (17) has a solution on $[0, 1]$, it is the expression for γ_c. Moreover, (39) is positive for $\gamma < \gamma_c$ and negative for $\gamma > \gamma_c$. To see why, the numerator in (39) is decreasing in γ. This can be seen by checking its first-order derivative:

$$-(\beta_0^s - \beta_0^d)h'(\gamma) + (\beta_0^s - \beta_0^d)h'(\gamma) + (\beta_0^s\gamma + \beta_0^d(1-\gamma))h''(\gamma) = (\beta_0^s\gamma + \beta_0^d(1-\gamma))h''(\gamma) < 0$$

The last step is by Assumption 2. Therefore, $\gamma = \gamma_c$ is the global maximizer. When (17) does not have a solution on $[0, 1]$, it means that the maximizer is a boundary solution. Extending the definition of γ_c to include the boundary solutions gives by

$$\gamma_c^\dagger = \begin{cases} 0, & \frac{h'(0)}{h(0)} < \frac{\beta_0^s - \beta_0^d}{\beta_0^d} \\ \gamma_c, & \beta_0^s\frac{h'(1)}{h(1)} < \beta_0^s - \beta_0^d < \beta_0^d\frac{h'(0)}{h(0)} \\ 1, & \frac{h'(1)}{h(1)} > \frac{\beta_0^s - \beta_0^d}{\beta_0^s} \end{cases}$$

This means

$$v - c - \frac{\beta_0^s\gamma + \beta_0^d(1-\gamma)}{h(\gamma)\tilde{x}^{\alpha-1}} \leq v - c - \frac{\beta_0^s\gamma_c^\dagger + \beta_0^d(1-\gamma_c^\dagger)}{h(\gamma_c^\dagger)\tilde{x}^{\alpha-1}}$$

Since $\alpha > 1$ and $\tilde{x} < x_c$,

$$v - c - \frac{\beta_0^s \gamma_c^\dagger + \beta_0^d (1 - \gamma_c^\dagger)}{h(\gamma_c^\dagger)\tilde{x}^{\alpha-1}} < v - c - \frac{\beta_0^s \gamma_c^\dagger + \beta_0^d (1 - \gamma_c^\dagger)}{h(\gamma_c^\dagger){x_c}^{\alpha-1}}$$

By (16), the right-hand side equals to 0. Then

$$\pi \leq v - c - \frac{\beta_0^s \gamma + \beta_0^d (1 - \gamma)}{h(\gamma)\tilde{x}^{\alpha-1}} \leq v - c - \frac{\beta_0^s \gamma_c^\dagger + \beta_0^d (1 - \gamma_c^\dagger)}{h(\gamma_c^\dagger)\tilde{x}^{\alpha-1}} < v - c - \frac{\beta_0^s \gamma_c^\dagger + \beta_0^d (1 - \gamma_c^\dagger)}{h(\gamma_c^\dagger){x_c}^{\alpha-1}} = 0$$

Therefore, such an $\tilde{x}$ does not exist. □

Proof of Proposition 5

Proof. By the proof of proposition 4, for any $x < x_c$, $\dot{x} \geq 0$ implies that $\pi < 0$. □

Proof of Proposition 6

Proof. By Lemma 5, for the infinite horizon problem, the objective function is equivalent to

$$\int_0^\infty e^{-\rho t} G(\gamma, x) dt + x(0)$$

By Theorem 1, $\gamma(t) = \gamma^*(x(t))$. By the convexity of $G(\gamma^*(x), x)$ by Lemma 6, $\arg\max_{x \leqslant \bar{x}} G(\gamma^*(x), x)$ is either x_0 or $\bar{x}$. If $G(\gamma^*(\bar{x}), \bar{x}) > G(\gamma^*(x_0), x_0)$, then a jump from x_0 to $\bar{x}$ is optimal; if not, $x(t) = x_0$ is optimal.

By Lemma 6, $G(\gamma^*(x), x)$ is continuous. $\lim_{x \to 0} G(\gamma^*(x), x) \to 0$. $\lim_{x \to \infty} G(\gamma^*(x), x) \to \infty$. Then for $x_0 < x^*$, $\tilde{x}$ defined in Proposition 6 must always exist. Moreover, if $\bar{x} > \tilde{x}$, $G(\gamma^*(\bar{x}), \bar{x}) > G(\gamma^*(x_0), x_0)$; if $x_0 < \bar{x} < \tilde{x}$, $G(\gamma^*(\bar{x}), \bar{x}) < G(\gamma^*(x_0), x_0)$. □

Proof of Proposition 7

Proof. This is a direct result of Theorem 4. □

Proof of Proposition 8, 9, 10, 11

Proof. Proposition 8 and 10 are direct results of Theorem 3; Proposition 9 and 11 are direct results of Theorem 4. □

Proof of Proposition 12 and Proposition 13

Proof. First, Proposition 1 still holds because it only requires no constraint on p_s, p_d. Lemma 6 holds too because it only requires g to be integrable. Hence,

$$G(\gamma,x) = (v-c)A\min\{\beta_1^s\gamma, \beta_1^d(1-\gamma)\}x - (\rho + \beta_0^s\gamma + \beta_0^d(1-\gamma))x$$
$$= \begin{cases} (((v-c)A\beta_1^s - (\beta_0^s - \beta_0^d))\gamma - (\rho+\beta_0^d))x, & \gamma < \frac{\beta_1^d}{\beta_1^d+\beta_1^s} \\ ((-(v-c)A\beta_1^d - (\beta_0^s - \beta_0^d))\gamma + (v-c)A\beta_1^d - (\rho+\beta_0^d))x, & \gamma \geqslant \frac{\beta_1^d}{\beta_1^d+\beta_1^s} \end{cases}$$

There are three cases:

(i) $\beta_0^s - \beta_0^d \geqslant (v-c)A\beta_1^s$. For a given $x > 0$, $G(\gamma,x)$ is decreasing in γ. $G(\gamma,x) \leqslant G(0,x) = -(\rho+\beta_0^d)x$. The optimal balance is then $\gamma^*(x) = 0$, and the optimal policy is to keep $x(t) = x_0$. The market is not viable.

(ii) $\beta_0^s - \beta_0^d \leqslant -(v-c)A\beta_1^d$. For a given $x > 0$, $G(\gamma,x)$ is increasing in γ. $G(\gamma,x) \leqslant G(1,x) = -(\rho+\beta_0^s)x$. The optimal balance is then $\gamma^*(x) = 1$, and the optimal policy is again to keep $x(t) = x_0$. The market is not viable.

(iii) $-(v-c)A\beta_1^d < \beta_0^s - \beta_0^d < (v-c)A\beta_1^s$ For a given $x > 0$, $G(\gamma,x)$is increasing in γ in the first piece and decreasing in γ in the second piece. Hence, $\gamma^*(x) = \frac{\beta_1^d}{\beta_1^d+\beta_1^s}$. $G\left(\frac{\beta_1^d}{\beta_1^d+\beta_1^s}, x\right) = \frac{(v-c)A\beta_1^s\beta_1^d - \beta_1^s(\rho+\beta_0^d) - \beta_1^d(\rho+\beta_0^s)}{\beta_1^s+\beta_1^d}x$. $h(\gamma^*(x)) = \frac{(v-c)A\beta_1^s\beta_1^d - \beta_1^s(\rho+\beta_0^d) - \beta_1^d(\rho+\beta_0^s)}{\beta_1^s+\beta_1^d}$.

If (28) holds, $h(\gamma^*(x)) > 0$, $G^*(\gamma^*(x),x)$ is increasing in x. Hence, faster growth is better than slower growth. Otherwise, $h(\gamma^*(x)) \leq 0$, $G^*(\gamma^*(x),x)$ is decreasing in x. Hence, it is optimal to keep $x(t) = x_0$. The market is not viable

□

Proposition 14 *Consider a trajectory of size $x(t)$ and balance $\gamma(t)$ that are continuously differentiable. Then it can be uniquely determined that the stock of supply is*

$$s(t) = \beta_1^s\gamma(t)x(t) \tag{41}$$

the stock of demand is

$$d(t) = \beta_1^d(1-\gamma(t))x(t) \tag{42}$$

the price paid to the seller is

$$p_s(t) = c + \frac{\gamma'(t)x(t) + \gamma(t)x'(t) + \beta_0^s\gamma(t)x(t)}{g(\gamma(t), x(t))} \tag{43}$$

and the price charged to the buyer is

$$p_d(t) = v - \frac{-\gamma'(t)x(t) + (1-\gamma(t))x'(t) + \beta_0^d(1-\gamma(t))x(t)}{g(\gamma(t), x(t))} \tag{44}$$

Proof. First, note that (42) and (41) are directly given by the definition of $\gamma(t)$. Next, by (6),

$$p_s = c + \frac{s' + \beta_0^s s}{\beta_1^s g(s,d)}$$

Since $s(t) = \beta_1^s\gamma(t)x(t)$ and $\gamma(t), x(t)$ are both differentiable, by the chain rule we have

$$s' = \beta_1^s(\gamma'(t)x(t) + \gamma(t)x'(t)) \tag{45}$$

Substituting s' in the expression of p_s with (45) gives $p_s(t)$. The expression of $p_d(t)$ can be obtained using similar steps by (7) and (42). □

Proposition 15 *Suppose there are finite number of jumps in the trajectory of $\gamma(t), x(t)$. Then for t around such a discontinuous time point t_0, the price paid to the seller is*

$$p_s(t) = \frac{\gamma(t_0^+)x(t_0^+) - \gamma(t_0^-)x(t_0^-)}{g(\gamma(t_0^-), x(t_0^-))}\delta(t_0 - t)$$

and the price charged to the buyer is

$$p_d(t) = \frac{(1-\gamma(t_0^-))x(t_0^-) - (1-\gamma(t_0^+))x(t_0^+)}{g(\gamma(t_0^-), x(t_0^-))}\delta(t_0 - t)$$

Proof. By proposition 14, around time t_0,

$$s(t_0^-) = \beta_1^s\gamma(t_0^-)x(t_0^-), s(t_0^+) = \beta_1^s\gamma(t_0^+)x(t_0^+)$$

We have shown in the proof of Proposition 1 that a price shock on the supply side $p_s(t) = \frac{s_1 - s_0}{\beta_1^s g(s_0, d_0)}\delta(\tau - t)$ can instantly shift the stock of supply from s_0 to s_1 at time τ. As a result, a price shock given below will shift the stock of supply from $s(t_0^-)$ to $s(t_0^+)$:

$$
\begin{aligned}
p_s(t) &= \frac{\beta_1^s \gamma(t_0^+) x(t_0^+) - \beta_1^s \gamma(t_0^-) x(t_0^-)}{\beta_1^s g(\gamma(t_0^-), x(t_0^-))}\delta(t_0 - t) \\
&= \frac{\gamma(t_0^+) x(t_0^+) - \gamma(t_0^-) x(t_0^-)}{g(\gamma(t_0^-), x(t_0^-))}\delta(t_0 - t)
\end{aligned}
$$

The price shock for the demand side can be shown using identical analysis. $\square$

Proposition 16 *Consider the production function to be $g(s,d) = As^{\alpha_s}d^{\alpha_d}$. For $\alpha < 1$, x^* is increasing in β_1^s, β_1^d, decreasing in β_0^s, β_0^d and decreasing in ρ. In words, the saturation size is decreasing in supply and demand cost, increasing in supply and demand durability, and decreasing in the time discount factor.*

The saturation balance $\gamma^*(x^*)$ in Theorem 2 is $\gamma^*(x^*) = \frac{\alpha_s(\rho+\beta_0^d)}{\alpha_s(\rho+\beta_0^d)+\alpha_d(\rho+\beta_0^s)}$ Note this balance is driven by 1) the elasticity of output from supply and demand, α_s, α_d; 2) the durability of supply and demand, $1/\beta_0^s$, $1/\beta_0^d$; and 3) the discount rate ρ. Note α_i determines the short-term return from the resources on each side of the market, while $1/\beta_0^i$ determines how long the resource lasts. Therefore, when ρ is large, the short-term reward is more valued, and $\gamma^*(x^*) \to \frac{\alpha_s}{\alpha_s+\alpha_d}$, the output-maximizing balance. When ρ is small, the long-term reward is more valued, and $\gamma^*(x^*) \to \frac{\alpha_s/\beta_0^s}{\alpha_s/\beta_0^s+\alpha_d/\beta_0^d}$, a balance which favors the more durable side of the market. The optimal steady-state balance reflects this trade-off between short-term and long-term gain.

Proof. The expression for the saturation size can be extended as

$$
x^* = \{A(v-c)(\beta_1^s)^{\alpha_s}(\beta_1^d)^{\alpha_d}\alpha_s^{\alpha_s}\alpha_d^{\alpha_d}\}^{\frac{1}{1-\alpha}} \left\{ \frac{\alpha_s}{(\rho+\beta_0^d)^{\frac{\alpha_d}{1-\alpha}}(\rho+\beta_0^s)^{\frac{1-\alpha_d}{1-\alpha}}} + \frac{\alpha_d}{(\rho+\beta_0^d)^{\frac{1-\alpha_s}{1-\alpha}}(\rho+\beta_0^s)^{\frac{\alpha_s}{1-\alpha}}} \right\}
$$

Since $\alpha < 1$, $\frac{\alpha_s}{(\rho+\beta_0^d)^{\frac{\alpha_d}{1-\alpha}}(\rho+\beta_0^s)^{\frac{1-\alpha_d}{1-\alpha}}} + \frac{\alpha_d}{(\rho+\beta_0^d)^{\frac{1-\alpha_s}{1-\alpha}}(\rho+\beta_0^s)^{\frac{\alpha_s}{1-\alpha}}}$ is decreasing in ρ, β_0^d and β_0^s. $(\beta_1^s)^{\alpha_s}(\beta_1^d)^{\alpha_d}$ is increasing in β_1^d and β_1^s. $\square$

Proof of Theorem 1

Proof. By Lemma 5, the infinite-horizon problem (13) can be written as

$$\int_0^\infty e^{-\rho t} G(\gamma, x) dt + x(0) \tag{46}$$

Since the selection of γ only affects the term $G(\gamma, x)$, by Lemma 6 and Theorem 1, it is optimal toy set $\gamma = \arg\max G(\gamma, x)$. Steps for obtaining $\gamma^*(x)$ and its monotonicity can be found in the proof of Lemma 6.

To see the limit of γ^*, consider $\beta_0^s < \beta_0^d$ and $\alpha < 1$. $\lim_{x\to 0} h'^{-1}(\frac{(\beta_0^s-\beta_0^d)x^{1-\alpha}}{v-c}) = h'^{-1}(0) = \gamma^*$. By Assumption 3, $\min(1, \gamma^*) = \gamma^*$. For $x \to +\infty$, $\lim_{x\to+\infty} h'^{-1}(\frac{(\beta_0^s-\beta_0^d)x^{1-\alpha}}{v-c}) = h'^{-1}(-\infty)$. if $h'(1)$ is bounded, then $h'(1) > -\infty$, $h'^{-1}(h'(1)) < h'^{-1}(-\infty)$, and $\min(1, h'^{-1}(-\infty)) = 1$. If $h'(1)$ is unbounded, by Assumption 2 and 3, $h'(1) \to -\infty$. Then $h'^{-1}(-\infty)$ just equals to 1. The other three cases can be checked similarly. □

Proof of Theorem 2

Proof. The current value Hamiltonian to infinite-horizon problem (13) is

$$H(x, \pi, \gamma, \psi) = \pi g(\gamma, x) + \psi \left(-(\beta_0^s \gamma + \beta_0^d (1-\gamma))x + (v - c - \pi) g(\gamma, x)\right)$$

By the maximum principle, a stationary solution must satisfy

$$H_\pi = 0, H(x^*, \pi^*, \gamma^*, \psi) > H(x^*, \pi^*, \gamma, \psi), \psi' = \rho\psi - H_x = 0, x' = 0$$

By $H_\pi = 0$ and $g(\gamma, x) = h(\gamma)x^\alpha$,

$$(1-\psi)\pi h(\gamma) x^\alpha = 0$$

Since $x > 0$, $0 < \gamma < 1$, $h(\gamma) \neq 0$,

$$\psi = 1$$

which satisfies the limiting transversality condition

$$\lim_{T\to\infty} e^{-\rho T} \psi(T) = 0$$

To maximize $H(x^*, \pi^*, \gamma, \psi)$, set $H_\gamma = 0$ assuming it has a solution on $[0, 1]$:

$$(\pi - \psi\pi + v - c)h'(\gamma)x^\alpha - \psi(\beta_0^s - \beta_0^d)x = 0$$

Substitute by $\psi = 1$ and rearrange the terms,

$$(v - c)h'(\gamma) = (\beta_0^s - \beta_0^d)x^{1-\alpha} \tag{47}$$

When (47) does not have a solution on $[0, 1]$, the optimal γ is on the boundary; whether $\gamma = 1$ or 0 depends on its gradient. One can check that the balance that maximizes H is given by

$$\gamma^*(x) = \max\left\{ \min\left\{ h'^{-1}\left(\frac{(\beta_0^s - \beta_0^d)x^{1-\alpha}}{v - c}\right), 1\right\}, 0\right\}$$

By $\psi' = \rho\psi - H_x$ and $\psi = 1$,

$$0 = \rho - H_x$$

which gives

$$\rho = -(\beta_0^s\gamma^*(x) + \beta_0^d(1 - \gamma^*(x))) + (v - c)\alpha h(\gamma^*(x))x^{\alpha-1} \tag{48}$$

Both x and φ are constants, so the last two equations are satisfied. Combining (47) and (48) gives the stationary solution.

We also need to check whether H is jointly concave in (γ, x) for sufficiency.

$$H(x, \pi, \gamma^*(x), \psi) = \{(1 - \psi)\pi + \psi(v - c)\}h(\gamma^*(x))x^\alpha - \psi(\beta_0^s\gamma^*(x) + \beta_0^d(1 - \gamma^*(x)))x$$

$$H_x(x, \pi, \gamma^*(x), \psi) = \{(1 - \psi)\pi + \psi(v - c)\}h(\gamma^*(x))\alpha x^{\alpha-1} - \psi(\beta_0^s\gamma^*(x) + \beta_0^d(1 - \gamma^*(x)))$$

$$H_{xx}(x, \pi, \gamma^*(x), \psi) = ((1 - \psi)\pi + \psi(v - c))\alpha\left(h'(\gamma)\frac{d\gamma^*(x)}{dx} + h(\gamma)(\alpha - 1)x^{\alpha-2}\right) - \psi(\beta_0^s - \beta_0^d)\frac{d\gamma^*(x)}{dx}$$

When $\gamma^*(x)$ is a boundary solution, i.e. $\gamma^*(x) = 1$ or 0, H_{xx} can be simplified as

$$H_{xx}(x, \pi, \gamma^*(x), \psi) = ((1 - \psi)\pi + \psi(v - c))\alpha h(\gamma^*(x))(\alpha - 1)x^{\alpha-2}$$

When $\gamma^*(x)$ is an interior solution, by $H_\gamma = 0$,

$$((1-\psi)\pi + \psi(v-c))h'(\gamma)x^{\alpha-1} = \psi(\beta_0^s - \beta_0^d)$$

Moreover,

$$\frac{d\gamma^*(x)}{dx} = \frac{\psi(\beta_0^s - \beta_0^d)(1-\alpha)}{((1-\psi)\pi + \psi(v-c))h''(\gamma)x^\alpha} = \frac{h'(\gamma)(1-\alpha)}{h''(\gamma)x}$$

Hence,

$$H_{xx}(x,\pi,\gamma^*(x),\psi) = \frac{(\alpha-1)\psi^2(\beta_0^s - \beta_0^d)^2}{((1-\psi)\pi + \psi(v-c))h''(\gamma)x^\alpha}\left\{\frac{\alpha h(\gamma)h''(\gamma)}{h'(\gamma)^2} + 1 - \alpha\right\}$$

By the maximum principle, $H_\pi = 0$, which implies that $\psi(t) = 1$. Therefore, when $\gamma^*(x)$ is an interior solution,

$$H_{xx}(x,\pi,\gamma^*(x),\psi) = \frac{(\alpha-1)(\beta_0^s - \beta_0^d)^2}{(v-c)h''(\gamma)x^\alpha}\left\{\frac{\alpha h(\gamma)h''(\gamma)}{h'(\gamma)^2} + 1 - \alpha\right\}$$

When $\gamma^*(x)$ is a boundary solution,

$$H_{xx}(x,\pi,\gamma^*(x),\psi) = (v-c)\alpha h(\gamma^*(x))(\alpha-1)x^{\alpha-2}$$

If $\alpha < 1$, under Assumption 2 and 4, $h''(\gamma) < 0$, $\frac{\alpha h(\gamma)h''(\gamma)}{h'(\gamma)^2} + 1 - \alpha < 0$, which implies that $H_{xx} < 0$. Therefore, Hamiltonian is jointly concave in (γ, x). If $\alpha > 1$, $\frac{\alpha h(\gamma)h''(\gamma)}{h'(\gamma)^2} < 0$, $1-\alpha < 0$, and again by Assumption 2, $h''(\gamma) < 0$. Therefore, $H_{xx} > 0$. This implies that the Hamiltonian $H(x,\pi,\gamma^*(x),\psi)$ attains its minimum at $x = x^*$. Thus, the stationary solution characterizes a saddle point. □

Proof of Theorem 3

Proof. By Lemma 5, the infinite-horizon problem (13) can be written as

$$\int_0^\infty e^{-\rho t}G(\gamma, x)dt + x(0) \tag{49}$$

Since the selection of γ only affects the term $G(\gamma, x)$, by Lemma 6 and Theorem 1, it is optimal to set $\gamma = \gamma^*(x(t))$. Then (49) is just a function of $x(t)$. We show that the solution has the property of a *most rapid approach path* (see Spence and Starrett (1975)). The next part is also similar to the steps taken in Spence and Starrett (1975):

(1) Any path from x_0 to $\bar{x} > x_0$ is feasible. This is true by having $f(x) > 0$ for all x.

(2) The optimal growth path either (a) stays at x_0 forever, or (b) goes to $\bar{x}$. This can be shown by contradiction. Suppose (a) and (b) are both not optimal; Then the optimal increasing growth path must grow to a market size y such that $x_0 < y < \bar{x}$. If $G(\gamma^*(y), y) \leq G(\gamma^*(x_0), x_0)$, by convexity, $G(\gamma^*(x), x) < G(\gamma^*(x_0), x_0)$ for any x that $x_0 < x < y$. Then any path from x_0 to y is clearly dominated by (a); If $G(\gamma^*(y), y) > G(\gamma^*(x_0), x_0)$, by Lemma 6, it must be true that $G(\gamma^*(x), x) > G(\gamma^*(y), y)$ for any x that $y < x \leq \bar{x}$. But this implies that a growth path from x_0 to y is strictly dominated by (b). Contradiction.

(3) By Lemma 7, a candidate for an optimal growth path from x_0 to $\bar{x}$ must have the following property:

$$x(t;t_i) = \begin{cases} x_0, & 0 \leq t \leq t_i \\ F(t-t_i), & t_i < t \leq t_i + F^{-1}(\bar{x}) \\ \bar{x}, & t_i + F^{-1}(\bar{x}) < t \end{cases} \tag{50}$$

We show that the optimal t_i is either 0 or $+\infty$. (49) can be further expanded as

$$\begin{aligned} J = & \int_{t_i}^{t_i+F^{-1}(\bar{x})} e^{-\rho t} G(\gamma^*(F(t-t_i)), F(t-t_i))dt + G(\gamma^*(x_0), x_0)\frac{1-e^{-\rho t_i}}{\rho} + G(\gamma^*(\bar{x}), \bar{x})\frac{e^{-\rho(t_i+F^{-1}(\bar{x}))}}{\rho} \\ = & e^{-\rho t_i}\int_0^{F^{-1}(\bar{x})} e^{-\rho t} G(\gamma^*(F(t)), F(t))dt + G(\gamma^*(x_0), x_0)\frac{1-e^{-\rho t_i}}{\rho} + G(\gamma^*(\bar{x}), \bar{x})\frac{e^{-\rho(t_i+F^{-1}(\bar{x}))}}{\rho} \end{aligned}$$

The second equality is by the change of variable. Taking derivative of J over t_i gives:

$$\begin{aligned} \frac{\partial J}{\partial t_i} &= -e^{-\rho t_i}\left\{\rho\int_0^{F^{-1}(\bar{x})} e^{-\rho t} G(\gamma^*(F(t)), F(t))dt - G(\gamma^*(x_0), x_0) + e^{-\rho F^{-1}(\bar{x})} G(\gamma^*(\bar{x}), \bar{x})\right\} \\ &= -e^{-\rho t_i}\int_0^{F^{-1}(\bar{x})} e^{-\rho t} G_x(\gamma^*(F(t)), F(t))F'(t)dt \end{aligned}$$

The second equality is obtained by integration by parts. Hence, the sign of $\frac{\partial J}{\partial t_i}$ does not change in t_i. In particular, if $\int_0^{F^{-1}(\bar{x})} e^{-\rho t} G_x(\gamma^*(F(t)), F(t))F'(t)dt > 0$, $\frac{\partial J}{\partial t_i} < 0$, the growth path $x(t;0)$ is optimal; otherwise, the growth path $x(t;+\infty)$ is optimal.

(4) It can be checked that $x(t; +\infty)$ generates the same profit (49) as $x(t) = x_0, \forall t$. So it is sufficient to compare $x(t; 0)$ and $x(t; +\infty)$. Define

$$S(\bar{x}) = \int_0^{F^{-1}(\bar{x})} e^{-\rho t} G_x(\gamma^*(F(t)), F(t)) F'(t) dt \tag{51}$$

We show that for any $x_0 < x^*$, there exists an $\tilde{x}$ such that $S(\bar{x}) > 0$ for any $\bar{x} > \tilde{x}$, and $S(\bar{x}) < 0$ for any $\bar{x} < \tilde{x}$:

$$S'(\bar{x}) = \frac{dF^{-1}(\bar{x})}{d\bar{x}} e^{-\rho F^{-1}(\bar{x})} G_x(\gamma^*(\bar{x}), \bar{x}) F'(F^{-1}(\bar{x})) = e^{-\rho F^{-1}(\bar{x})} G_x(\gamma^*(\bar{x}), \bar{x})$$

The derivative is obtained following Leibniz integral rule. Then by Lemma 6, for $\bar{x} > x^* (\bar{x} < x^*)$, $S'(\bar{x}) > 0 (S'(\bar{x}) < 0)$. Moreover, $S(x_0) = 0$. Therefore, if $x_0 < x^*$, then $S(\bar{x}) = 0$ must have a unique positive root. Denote it as $\tilde{x}$. Then $S(\bar{x}) < 0$ for $\bar{x} < \tilde{x}$ and $S(\bar{x}) > 0$ for $\bar{x} > \tilde{x}$.

If $x_0 > x^*$, then $G_x(\gamma^*(F(t)), F(t)) > 0$ on $[x_0, \bar{x}]$. Hence, $S(\bar{x}) > 0$ as well.

Hence, given that $x_0 < x^*$, for $\bar{x} > \tilde{x}$, $S(\bar{x}) > 0$, $\frac{\partial J}{\partial t_i} < 0$, the optimal policy is to grow as fast as possible, and the optimal growth path is $x(t; 0)$; for $\bar{x} < \tilde{x}$, $S(\bar{x}) < 0$, $\frac{\partial J}{\partial t_i} > 0$, the optimal policy is to grow as slow as possible, and the optimal growth path is $x(t) = x_0, \forall t$. Given that $x_0 > x^*$, it is optimal to grow as fast as possible, and the optimal growth path is $x(t; 0)$. □

Proof of Theorem 4

Proof. By Theorem 2, in a decreasing returns to scale market, the long-run optimal size is the saturation size x^*. By Theorem 6, faster growth dominates slower growth below x^*. Hence, the optimal growth policy is to grow to the saturation size as fast as possible. □

Proof of Theorem 5

Proof. This is a combination of Theorem 3 and Theorem 4. □

Theorem 6 (Fast vs. Slow) *Consider a set of increasing growth paths from $\underline{x}$ to $\bar{x}$ and the fixed endpoint problem*

$$\max \int_0^{t_0} e^{-\rho t} \pi g(\gamma^*(x), x) dt$$

$$s.t. \quad x(0) = \underline{x}, \; x(t_0) = \bar{x}$$

Then in a decreasing returns to scale market, if $0 < \underline{x} < \bar{x} \leq x^*$, faster growth dominates slower *growth; and if $x^* \leq \underline{x} < \bar{x}$, slower growth dominates faster growth. Conversely, in an increasing returns to scale market, if $0 < \underline{x} < \bar{x} \leq x^*$, then slower growth dominates faster growth; if $x^* \leq \underline{x} < \bar{x}$, then faster growth dominates slower growth. To summarize,*

	$\alpha > 1$	$\alpha < 1$
$\underline{x} > x^*$	*faster is better*	*slower is better*
$\bar{x} < x^*$	*slower is better*	*faster is better*

This result shows that the stationary point in Theorem 2 defines a threshold between growth strategies over an interval – determining when it is optimal to grow fast and when it is optimal to grow slow. We next apply this result to analyze the optimal growth strategies overall.

Proof. By Lemma 5 in the Appendix,

$$\int_0^{t_0} e^{-\rho t}\pi g(\gamma^*(x), x)dt = \int_0^{t_0} G(\gamma^*(x), x)e^{-\rho t}dt - e^{-\rho t_0}\bar{x} + \underline{x}$$

Let $x_1(t)$ be a faster growth path from $\underline{x}$ to $\bar{x}$ over $[0, t_0]$ than $x_2(t)$. Let the corresponding pricing policies be $\pi_1(t)$ and $\pi_2(t)$. Then

$$\int_0^{t_0} e^{-\rho t}\pi_1 g(\gamma^*(x_1), x_1)dt - \int_0^{t_0} e^{-\rho t}\pi_2 g(\gamma^*(x_2), x_2)dt$$

$$= \int_0^{t_0} (G(\gamma^*(x_1), x_1) - G(\gamma^*(x_2), x_2))e^{-\rho t}dt \tag{52}$$

Setting $\frac{dG(\gamma^*(x),x)}{dx} = 0$ gives

$$-(\rho + \beta_0^s\gamma^*(x) + \beta_0^d(1 - \gamma^*(x))) + (v - c)\alpha h(\gamma^*(x))x^{\alpha-1} = 0$$

This is the same equation as (48). Therefore, $x = x^*$ is the solution to $\frac{dG(\gamma^*(x),x)}{dx} = 0$.

By Lemma 6 in the Appendix, when $\alpha > 1$, $G(\gamma^*(x), x)$ is convex in x. $x = x^*$ is the global minimum. $G(\gamma^*(x), x)$ is strictly decreasing for $x < x^*$ and strictly increasing for $x > x^*$.

If $\bar{x} < x^*$, then $x_2(t) < x_1(t) < x^*$, and

$$G(\gamma^*(x_1), x_1) < G(\gamma^*(x_2), x_2)$$

for all $0 \leqslant t \leqslant t_0$. As a result, (52) is negative. This implies that slower growth paths dominate faster growth paths.

Similarly, if $x^* < \underline{x}$, then $x^* < x_2(t) < x_1(t)$, and

$$G(\gamma^*(x_1), x_1) > G(\gamma^*(x_2), x_2)$$

for all $0 \leqslant t \leqslant t_0$. As a result, (52) is positive. This implies that faster growth paths dominate slower growth paths.

For $\alpha < 1$, under Assumption 4, $G(\gamma^*(x), x)$ is concave in x. So the monotonicity flips for $x > x^*$ and $x < x^*$. Using a similar analysis as for $\alpha > 1$ completes the proof. □

An illustrative example of supply and demand surpluses In the analytical model, we look at optimal balance as a measure for the market value of the supply relative to the total market value. Here we take a similar approach and calculate the ratio of the supply side surplus relative to the total surplus, which we call the *balance of surplus.* The formula is given by

$$\frac{(p_s - c)g(s,d)}{(p_s - c)g(s,d) + (v - p_d)g(s,d)} = \frac{\Delta(\gamma x)/\Delta t + \beta_0^s \gamma x}{\Delta x/\Delta t + \beta_0^s \gamma x + \beta_0^d (1-\gamma)x}$$

The above expression can be obtained by Proposition 14. Hence, if the trajectory of $x(t)$ is fixed, γ is also fixed at $\gamma^*(x(t))$. The trajectory of the balance of surplus can then be computed as a function of $x(t)$. Here we provide two examples.

For the increasing returns to scale market, we use the optimal market size trajectory under the high-subsidy policy $m = 1$ (shown in Figure 1). The trajectory of the balance of surplus is shown in Figure 5. The balance of surplus is increasing before the market reaches its potential at $t = 4.5$ years. Compared with Figure 3, it shows that the balance of surplus is increasing as the market grows, similar to the optimal balance.

For the decreasing returns to scale market, we use the optimal market size trajectory under the high-subsidy policy $m = 1$ (shown in Figure 2). The balance of surplus is decreasing before the

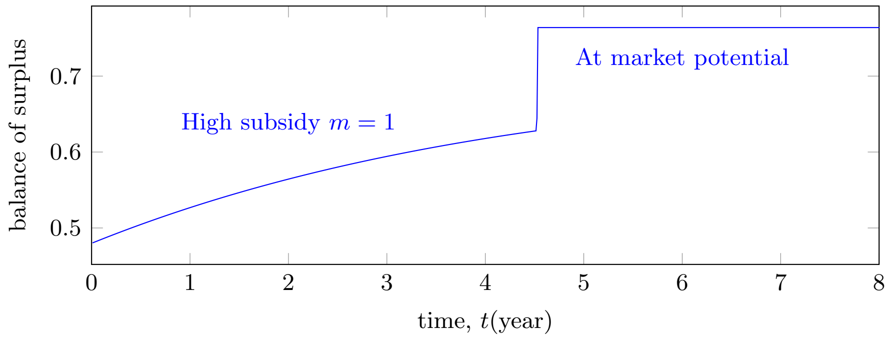

Figure 5: Evolution of balance of surplus in the change of optimal market size trajectory (increasing returns)
Note. The market size trajectory $x(t)$ is the optimal policy under high subsidy $m = 1$

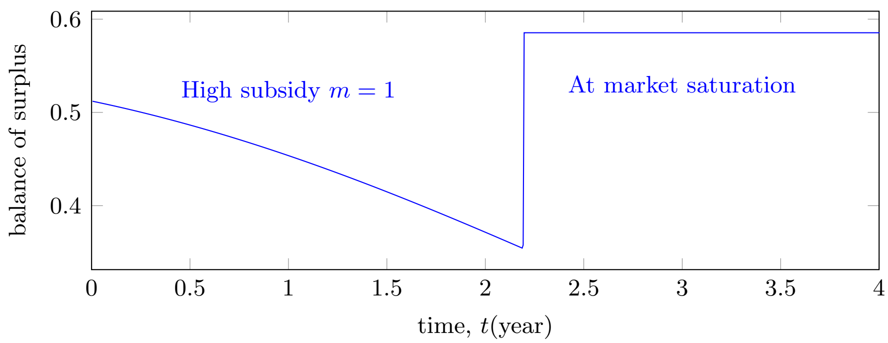

Figure 6: Evolution of balance of surplus in the change of optimal market size trajectory (decreasing returns)

market reaches the saturation size at $t = 2.2$ years. It also shows a similar trend as as the optimal balance in Figure 4.

Hence, although it is intractable to analytically show the connection between the optimal balance and the balance of surplus, numerical examples suggest that a larger market balance implies that the total surplus of supply and demand is more concentrated on the supply side, and a smaller market balance implies that the total surplus of supply and demand is more concentrated on the demand side.

References

Acemoglu, Daron. 2009. *Introduction to Modern Economic Growth*. Princeton University Press, Princeton New Jersey.

Armstrong, Mark. 2006. Competition in two-sided markets. *The RAND Journal of Economics* **37**(3) 668–691.

Bai, Jiaru, Kut C So, Christopher S Tang, Xiqun Chen, Hai Wang. 2018. Coordinating supply and demand on an on-demand service platform with impatient customers. *Manufacturing & Service Operations Management* .

Banerjee, Siddhartha, Ramesh Johari, Carlos Riquelme. 2015. Pricing in ride-sharing platforms: A queueing-theoretic approach. *Proceedings of the Sixteenth ACM Conference on Economics and Computation*. ACM, 639–639.

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

Bimpikis, Kostas, Ozan Candogan, Daniela Saban. 2019. Spatial pricing in ride-sharing networks. *Operations Research* .

Cachon, Gerard P, Kaitlin M Daniels, Ruben Lobel. 2017. The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management* **19**(3) 368–384.

Caillaud, Bernard, Bruno Jullien. 2003. Chicken & egg: Competition among intermediation service providers. *RAND journal of Economics* 309–328.

Castillo, Juan Camilo, Dan Knoepfle, Glen Weyl. 2017. Surge pricing solves the wild goose chase. *Proceedings of the 2017 ACM Conference on Economics and Computation*. ACM, 241–242.

Chen, Andrew. 2015. URL `https://andrewchen.co/ubers-virtuous-cycle-5-important-reads-about-uber/`

Cullen, Zoë, Chiara Farronato. 2014. Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms. *Job Market Paper* .

Dhebar, Anirudh, Shmuel S Oren. 1986. Dynamic nonlinear pricing in networks with interdependent demand. *Operations Research* **34**(3) 384–394.

Diamond, Peter A. 1982. Aggregate demand management in search equilibrium. *Journal of political Economy* **90**(5) 881–894.

Gurvich, Itai, Martin Lariviere, Antonio Moreno. 2019. Operations in the on-demand economy: Staffing services with self-scheduling capacity. *Sharing Economy*. Springer, 249–278.

Hagiu, Andrei. 2004. Two-sided platforms: Pricing and social efficiency. *Available at SSRN 621461* .

Hu, Ming, Yun Zhou. 2017. Price, wage and fixed commission in on-demand matching .

Kabra, Ashish, Elena Belavina, Karan Girotra. 2016. Designing promotions to scale marketplaces. Ph.D. thesis, Working paper, INSEAD, Fontainebleau, France.

Kaldor, Nicholas. 1957. A model of economic growth. *The economic journal* **67**(268) 591–624.

Kalish, Shlomo. 1985. A new product adoption model with price, advertising, and uncertainty. *Management science* **31**(12) 1569–1585.

Li, Jun, Serguei Netessine. 2019. Higher market thickness reduces matching rate in online platforms: Evidence from a quasiexperiment. *Management Science* .

Nikzad, Afshin. 2017. Thickness and competition in ride-sharing markets. *Available at SSRN 3065672* .

Oren, Shmuel S, Rick G Schwartz. 1988. Diffusion of new products in risk-sensitive markets. *Journal of Forecasting* **7**(4) 273–287.

Rochet, Jean-Charles, Jean Tirole. 2003. Platform competition in two-sided markets. *Journal of the european economic association* **1**(4) 990–1029.

Rochet, Jean-Charles, Jean Tirole. 2006. Two-sided markets: a progress report. *The RAND journal of economics* **37**(3) 645–667.

Solow, Robert M. 1956. A contribution to the theory of economic growth. *The quarterly journal of economics* **70**(1) 65–94.

Spence, Michael, David Starrett. 1975. Most rapid approach paths in accumulation problems. *International Economic Review* 388–403.

Taylor, Terry A. 2018. On-demand service platforms. *Manufacturing & Service Operations Management* .

Vidale, ML, HB Wolfe. 1957. An operations-research study of sales response to advertising. *Operations research* **5**(3) 370–381.

Wernerfelt, Birger. 1991. Brand loyalty and market equilibrium. *Marketing Science* **10**(3) 229–245.

Weyl, E Glen. 2010. A price theory of multi-sided platforms. *American Economic Review* **100**(4) 1642–72.

Acknowledgment.

The authors gratefully acknowledge the helpful feedback of several academic and industry colleagues on this work, especially Karan Girotra. Keith Chen also contributed to early ideas about this topic while he and the second author were colleagues together at Uber. We also acknowledge both Cornell Tech and Lyft for their support in making this research possible.