An Economic Model for Pricing Digital Products

Gerald V. Post
University of the Pacific
JPost@Pacific.edu

Abstract
An economic model is created that emphasizes the characteristics of digital products. By focusing on price instead of quantity, the model examines the impact of low marginal costs and advertising revenue. The conditions for producing negative prices are explored along with the limitations of assuming linear demand curves in an e-commerce setting. The model explains observed phenomenon in Web examples including digital music, search engine competition, and social networks.

1. Introduction

Economic theory has evolved over dozens of years to produce useful methods to evaluate many aspects of business. Theories, which tend to be based on mathematical models, are powerful because they can identify relevant assumptions, precisely define a problem, and offer provable solutions. Models are sometimes criticized for being too abstract and not always providing specific actionable results. Yet, models provide critical frameworks for examining decisions and directing future research. In particular, they are vital to statistical analysis—providing the knowledge of what data to collect and how it should be analyzed.

Neoclassical economic models have long studied the issues of production and sales—leading to supply and demand curves. The foundations are well-understood, and the basic models are studied in common textbooks, such as [6], [14], and [7]. However, the traditional models, after a history of debate, emphasize the role of product quantity in making decisions. Firms and consumers are considered to be agents that respond to the market price. Consumers find their optimal allocation of product quantities based on income and prices. Firms optimize profit by determining the amount to produce given the market prices for products and inputs. Particularly in perfect competition, firms and consumers can control only the quantity—prices are determined by market interactions.

This approach makes sense in an economic world where scarce resources determine costs and increasing production of physical goods requires higher levels of resources. The traditional models are flexible and can be applied to many situations.

They produce parallel solutions to the results in this paper. However, it is easier to understand the digital world by looking at the problem with a new perspective. The resulting extensions to incorporate interesting cases in the digital world will appear less sacrilegious—such as examining negative prices.

Key distinguishing features of digital products:
1. They tend to have low or zero marginal costs—particularly for copies.
2. Individual products are somewhat unique—creating a monopoly or at least a monopolistically competitive market. For example, copyrights grant the owner of a specific song (or news article, book, and so on) an exclusive right to control the product.

Neoclassical models have the ability to analyze monopoly and monopolistically competitive markets. These methods are valuable and will play a role in this paper as well. But it is important to remember the critical aspect of these markets: The firm now has some control over price. Technically, price is ultimately determined by the consumer demand, but the firm needs to look at the consumer response to identify how price and quantity are determined.

The issue of zero or miniscule marginal costs are trickier. Here, the neoclassical models tend to fail—reaching constrained solutions (e.g., zero price), or offering little help (e.g., declining marginal costs lead to suboptimal solutions). The concept of negative prices is probably heretical. Because these concepts rarely occur in a physical world focused on production, they are harder to see within the traditional models. Yet, as the spending within the digital world increases, and new firms are created, there is a need for an economic model to provide guidance for pricing and evaluation decisions.

Research exists on several related areas in e-commerce, networks, and two-sided markets. [3] explained the value of bundling digital products with low marginal costs and positive prices. [4] examined a publishing model where consumers are willing to pay subscription prices to gain access to pricing information in advertisements. [8] looked at pricing options for database providers (such as DIALOG).
even if their total purchases are \( N_B N_S \), or possibly some fraction of that number. Implying that each purchaser will buy one item, one time from each seller. In two-sided market models, the key question lies with the adoption of a specific platform. As explained by [12] in the general case, these assumptions lead to a profit maximizing price written in terms of demand elasticity or Lerner formula:

\[
p^* = MC/(1 + 1/E_d)
\]

(1)

The two-sided market twist is that prices are symmetric for the two sides so that:

\[
p^* = p_b + p_s
\]

(2)

Any change in price on one side of the platform can be compensated by an equal but opposite change on the other side. For instance, software platforms can cut prices or subsidize developers by raising prices to buyers. Variations of the model lead to interesting results. However, quantity is generally not examined in the two-sided market models and marginal costs are rarely examined in detail.

Digital products can be examined in the two-sided market model; however, the products do not have to be sold that way, so the purpose of this paper is to build an economic model that emphasizes the characteristics of digital products. A key element of this model is to focus on price instead of quantity. Once the model is constructed, it can be used to examine several current issues, such as zero and negative prices for Web content and services.

2. Developing the Model

Economic models consist of two key elements: Businesses and Consumers. Consumers make decisions based on maximizing their “utility” of consumption subject to an income constraint. Businesses, outside the fixed-price environment of perfect competition, use the consumer model and cost models to maximize profits. The consumer model is reduced to a simple demand curve—representing consumer tradeoffs between price and quantity (incorporating additional factors as needed). It is easier to begin with the model of the firm, so that is presented in the first section. Consumer behavior is relatively traditional, but details with a few twists are covered in the second section.

3. Businesses

Even with digital products, businesses are assumed to maximize profits. (The economics literature contains many examinations of this assumption but they are not relevant to this paper.) Before considering digital product, it is helpful to summarize the basic tenets of the neoclassical model. Profits are generically defined as the difference between total revenue and total cost:

\[
\pi = TR - TC
\]

(3)

In neoclassical systems, both total revenue \((TR)\) and total cost \((TC)\) are functions of the quantity produced, which is the only control variable available to the firm. Consequently, profit can be maximized by differentiating profit with respect to quantity and solving the equation to reveal that marginal revenue \((MR)\) equals marginal cost \((MC)\):

\[
MR = MC
\]

(4)

Of course, second-order conditions impose constraints on the functional form of the profit function—notably it must be concave (or the second-derivative matrix must be negative semi-definite). It is important to remember that MR and MC are defined in terms of quantity \((MR = dTR/dQ)\). In the neoclassical monopolistic model, \(TR = P * Q\), but price is actually a function of quantity as well, so that \(TR = P(Q) * Q\). The price function is an inversion of the consumer demand curve, indicating that decreasing output \((Q)\) will result in higher prices for the product. Given specific functional forms for \(P(Q)\) and \(TC(Q)\), it is possible to solve equation (4) in terms of \(Q\) to obtain the optimal production quantity.

3.1. Digital Products and Services

In terms of digital products, businesses can more directly control price \((P)\) instead of quantity \((Q)\). Typically, content is placed online and downloaded or used by consumers. This process includes digital content such as music, video, and books—somewhat similar to traditional products; but it also includes

\[
E = p_b + p_s
\]

\[
E = p_b + p_s
\]

\[
E = p_b + p_s
\]
digital services, such as search engines, peer-to-peer interaction sites (e.g., Facebook), and any other type of digital service. Firms establish a price (often zero, and conceivably negative), and then consumers decide how much to use or purchase. The firm has relatively low marginal cost of producing more output. In particular, there are no direct labor costs since the process is automated, so the firm does not have to make a deliberate decision on each copy.

It is challenging to define what is meant by quantity in a digital environment. Ultimately, any digital product or service results in the transfer of data bytes, so quantity could be measured in terms of bytes transferred. In terms of products, this value might be aggregated as an average. For example, perhaps 2.5 MB represents one song. The basic model specifies Q as a generic quantity which can be redefined to meet the needs of specific situations.

The basic concept of profit maximization will hold, so equation (3) is useful. However, revenue to the digital firm is defined a bit differently, so the profit equation is expanded to:

\[ \pi = (P + Ad) Q(P) - TC(Q(P)) \]  

(5)

Equation (5) shows that firms can receive revenue from two sources: Prices paid by consumers and fees paid by advertisers. Observe that advertising revenue is expressed as a per-quantity rate (Ad). Firms have developed several mechanisms for charging for advertising content, but most of them can be reduced to this form. For example, the term can be expanded to specifically match common advertising methods:

\[ TR_{\text{ads}} = Ad \times CTR \times Q \text{Displayed} \]  

(6)

Equation (6) represents a typical method, where the firm earns ad revenue based on the number of ads displayed multiplied by the click-through-rate (CTR). For keyword advertising, the Ad rate might vary based on several exogenous factors, so it might be necessary to use an average value for a defined period of time. The point is that this equation can be reduced. Assume that any firm has a known ratio of ads it is willing to display so that:

\[ Ads \text{Displayed} = \gamma Q \]  

(7)

The number of ads displayed depends on the number of users of the site and the amount of content they observe or download. Thinking in terms of bytes, the firm would allocate a certain percentage of screen space or bandwidth to advertising. Combining equations (4) and (5) yields:

\[ TR_{\text{ads}} = Ad \times CTR \times \gamma Q \]  

(8)

And the three fixed terms in Equation (8) can be written as the simplified constant (Ad) used in Equation (5). This simplification is made at this point to keep the model easier to read. The Ad component can be replaced and analyzed in more detail after the basic model is developed.

Mathematically, it is straightforward to find the maximum profit in Equation (5) with respect to price (P). Remember that the function Q(P) represents the aggregate consumer demand for the digital product. Consequently, the first order condition is:

\[ \frac{\partial \pi}{\partial P} = Q + (P + Ad) \frac{\partial Q}{\partial P} - \frac{\partial TC}{\partial Q} Q \frac{\partial P}{\partial Q} = 0 \]  

(9)

The second-order conditions are not shown here, but in general they will hold for most common cases, such as negatively-sloped demand curves (\(\frac{\partial Q}{\partial P} < 0\)) and reasonable cost functions.

Notice that Equation (9) contains terms for both quantity and price, which is common. In practice, given a specific demand curve Q(P) and cost curve (\(\frac{\partial TC}{\partial Q}\)), the equation can be solved for a specific value of P*. Inserting this value into the demand curve yields the optimal Q* customers will purchase.

However, reducing the equation makes it easier to understand and evaluate. First, the equation can be rearranged in terms of the price:

\[ P^* = (MC - Ad) - \frac{\partial P}{\partial Q} Q \]  

(10)

For confirmation of the process, Equation (10) is similar to the neoclassical solution, such as that derived by [7]. In terms of interpretation, assuming Ad is zero for the moment, the equation states the optimal price will be equal to marginal cost plus an increment due to the monopoly power of the firm. Remember that \(\frac{\partial P}{\partial Q}\) is negative.

Figure 1 shows the solution in geometric terms, using the neoclassical approach. It assumes a relatively small value for marginal advertising revenue and assumes that marginal costs are small and fixed. These assumptions can be modified but they are used here to reduce the clutter in the chart.
The solution in Figure 1 ($P^*, Q^*$) mirrors Equation (10). Look at the vertical axis to find the decomposition of price into MC (green line to axis), $\partial P/\partial Q$ (axis to price point), and marginal Ad revenue (MR line to MR+Ad) line. Although it is not highlighted, the solution with no ad revenue is found from the intersection of the MR and MC lines. The advertising revenue enables the firm to reduce prices, leading to an increase in sales or an increase in customers. Equation (10) indicates another way to draw the solution is to use the traditional MR curve and simply reduce the MC curve by the fixed amount of the marginal advertising revenue.

With a focus on price, Equation (10) can be reconfigured to a more interesting form. Add price to the last term to get:

$$P = (MC - Ad) - \partial P/\partial Q Q/P P$$

Recall that the price elasticity of demand is defined as the percent change in quantity divided by the percent change in price:

$$E_d = \partial Q/\partial P P/Q$$

Substitute this definition into Equation (11) as the inverse and collect the price terms to get a variation of the Lerner condition:

$$P^* = MC - Ad$$

Compare this result to two-sided market presented in equations (1) and (2). In the two-sided market, the Ad (seller) revenue appeared on the left side with the price (buyer). The difference arises because this model focuses on quantity purchased, which is a function of price but not Ad revenue. This model also assumes that the firm has some price control in the buyer market, but not in the seller (Ad) market.

Bear in mind that the elasticity is generally not a constant. Typically, it changes along the demand curve so it is also a function of price. Consequently, Equation (13) is harder to use in practice than it appears. However, elasticity is a commonly-understood term and relatively easy to measure, so it could be computed for various points on a demand curve—leading to a relatively quick method to estimate optimal prices for new Web service.

### 3.2. Implications of the Model for Advertising

The beauty of Equation (13) is that it clarifies several relationships. First, if $-1 < E_d < 0$, where demand is inelastic, then the denominator will be negative and Equation (13) will not provide an optimal price—for the simple reason that the optimal price for inelastic demand is to raise the price to infinity (or until the antitrust regulators come after you).

More realistically, particularly at low prices, $E_d < -1$ and demand is elastic. Consequently, the inverse is constrained to a limited range: $-1 < E_d < 0$; which means the overall denominator of the equation is similarly constrained: $0 < \text{Denom.} < 1$. If advertising revenue is zero, this effect is relatively clear: optimal price will be greater than marginal cost—echoing the neoclassical answer.

An important question in the digital world revolves around the issue of how advertising revenue should affect price. Clearly, an increase in ad revenue will result in a lower price, but how much lower? The general result is found by differentiating Equation (10) with respect to Ad—remembering that the slope ($\partial P/\partial Q$) might also depend on $Q$ or $P$:

$$\partial P/\partial Ad = -1 - \left[\frac{\partial (\partial P/\partial Q) \partial P Q + \partial P/\partial Q \partial Q P}{\partial P/\partial Ad}\right]$$

Note that the second term in the brackets reduces to 1. Collect the $\partial P/\partial Ad$ terms and divide to get the general answer:

$$\partial P/\partial Ad = -1/(2 + \partial (\partial P/\partial Q) \partial P Q)$$

Although it looks a little strange, Equation (15) is useful—both for firms seeking to decide what price to charge if ad revenue rates change, and for policy analysts seeking to understand the relationships between digital businesses and customers. Effectively, Equation (15) reveals the percentage of advertising revenue that is shared with customers. It answers the question: If a dollar of marginal ad revenue is received, how much should price be reduced?

Consider the easiest answer first. With a linear demand curve, the slope is fixed, so its derivative is zero—removing the complicated term in the denominator, and reducing the value to exactly one-half. This result holds for any linear demand curve, which is a common approach to estimating customer relationships. The answer to the ad revenue question will always be the same: Reduce the price by half the amount of the ad money—in others words, share the ad revenue equally with the customers.

Another interesting case arises for exponential demand curves of the form:

$$Q = ae^{bP}$$

With this functional form, $dQ/dP = bQ$, so the inverse $dP/dQ = (bQ)^{-1}$. Differentiating with respect to $P$ yields:

$$\partial dP/dQ = -b(bQ)^{-2} (bQ) = -1/Q$$

Plugging this value into Equation (15) results in $\partial P/\partial Ad = -1$. That is, if demand follows an exponential function, regardless of the parameters, all per-unit advertising money is returned to customers in the form of lower prices.

Other functional forms result in slightly different answers. For example, quadratic demand curves...
result in values from -0.4 to -0.7 depending on the specific coefficients. (Note, extreme coefficients were not tested, so these bounds and not absolutes, just values that might arise in common practice.)

The critical conclusion is that any firm wishing to find an optimal price for digital goods and services in the presence of advertising revenue needs to carefully estimate the functional form of the consumer demand relationship. The form is likely to be more important than the actual coefficients. Since many practitioners automatically choose a linear form, the good news is that it is not even necessary to estimate the coefficients. From an equilibrium price, simply reduce the price by half the amount of the advertising rate.

3.3 Implications of the Model for Pricing

Clearly, prices are affected by the advertising revenue as shown in the previous section. However, the effects become even more interesting as the advertising rates increase, or the marginal costs decline. Some people expressed surprise when Microsoft announced in May 2008 that it would offer a cash back program for customers of its Live Search site (http://www.microsoft.com/presspass/press/2008/may08/05-21LiveSearchcashbackPR.mspx). Technically, the program generates a payment only when the consumer purchases a product from a vendor, so it is not quite a simple subsidy for using the system. But it is similar.

Equation (13) shows the optimal price is a function of MC – Ad revenue (weighted by the price elasticity). If the marginal ad revenue (expressed in the same terms as marginal cost such as value per quantity or bytes) is greater than the marginal cost, then the optimal price can be negative—or a subsidy to customers.

With digital products and services, marginal costs tend to be relatively low. Although fixed costs might be high (and tiered), the cost to providing service to an additional customer, or one more download is relatively low. As discussed in the next section, MC is probably not zero and decreases with an increase in scale. This situation is a marked contrast to physical goods. Consequently, any substantial marginal advertising revenue can lead to a negative price.

Equation (13) highlights an interesting twist when prices drop below zero. Elasticity of demand turns positive—a condition outside the realm of traditional economics. Note that the slope of the demand curve remains negative, but it is now multiplied by a negative value for price. However, the interpretation of elasticity remains the same—it measures the responsiveness of quantity to changes in price. Business managers still need to know how much quantity will increase as price is decreased. Consequently, demand curves exhibiting $0 < Ed < 1$ can still be called inelastic. A large percentage decrease in price results in a smaller percentage increase in quantity.

This interpretation is important, because for linear demand curves, the Ed cannot exceed one. Write a linear demand curve as:

$$Q = aP + b$$

Evaluate the limit of Ed as $P$ approaches negative infinity:

$$\lim_{P \to -\infty} Ed = \lim_{P \to -\infty} \frac{\partial Q/\partial P}{P} = a \lim_{P \to -\infty} \frac{P(aP + b)}{aP + b}$$

$$= a \lim_{P \to -\infty} \frac{1}{1 + b/P} \to a(1/a) = 1$$

A similar process for quadratic equations yields a limit of 2. In fact, for any demand function that is a polynomial in $P$, it is straightforward to show that the limit on the elasticity is the value of the highest power (3 for cubic, 5 for quintic, and so on). For an exponential demand curve ($Q = a e^{bP}$), the elasticity reduces to just $bP$, which approaches positive infinity as $P$ approaches negative infinity because $b$ is negative.

The value of the elasticity is important even for negative values of price because it indicates the customer responsiveness to the subsidy. Again, the model shows that the choice of functional form is more important than actually estimating the subsidy. If the demand is presumed to be linear, the demand will be inelastic for subsidies; indicating that relatively large subsidies will result in only small increases in quantity downloaded, or little time spent online clicking on ads. At a minimum, to estimate the effect of a subsidy, researchers should try to estimate quadratic equations or possibly exponential equations. But exponential demand responses could be rare—even in terms of subsidies.

3.4. Implications for Application

It is tempting to assume that marginal costs for digital products are zero. In actuality, although the costs are typically dominated by high fixed costs, it is likely that marginal costs are at least slightly above zero. Table 1 shows the typical costs incurred by most firms creating digital products and services.

As usual, the key is to identify exactly which costs apply to a given situation and then find a way to estimate them. Because fixed costs fall out of the marginal analysis, it is sometimes tempting to claim
all costs are fixed. However, in the digital world, the
data transmission cost is one that should probably be
treated as a marginal cost. Most network providers
charge a fixed cost per month, plus a per-gigabyte
charge above a minimum level. In most cases, it is
possible to change tiers on a monthly basis—making
even the “fixed” costs relatively flexible. Treating
these as constant average costs makes them
equivalent to fixed marginal costs. Similarly,
processing, storage, and backup/security costs are
generally tiered. These systems are typically
configured in a scalable system that can be expanded
relatively quickly. In particular, larger systems are
typically configured with some type of scale
mechanism; effectively converting fixed/average
costs into marginal costs.

<table>
<thead>
<tr>
<th>Cost Source</th>
<th>Fixed or Marginal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software</td>
<td></td>
</tr>
<tr>
<td>development</td>
<td>Fixed</td>
</tr>
<tr>
<td>Software</td>
<td></td>
</tr>
<tr>
<td>maintenance</td>
<td>Fixed (monthly)</td>
</tr>
<tr>
<td>Security</td>
<td>Fixed, but somewhat dependent on usage, such as the cost of backups.</td>
</tr>
<tr>
<td>Scale</td>
<td>Usually defined by tiers so fixed once a tier is chosen, but some room for marginal costs within scale (e.g., new servers in a server farm). At higher levels of quantity/bytes, more costs become marginal.</td>
</tr>
<tr>
<td>Data transfer</td>
<td>Fixed and marginal. Most hosting systems have a base data transfer rate and charge by the GB for additional transfers.</td>
</tr>
<tr>
<td>Content</td>
<td>Typically in the form of royalties, where content owners receive a percentage of revenue or a fixed marginal rate.</td>
</tr>
<tr>
<td>Marketing</td>
<td>A base fixed charge, but likely to increase as a firm tries to reach more potential customers.</td>
</tr>
</tbody>
</table>

Table 1: Typical costs with digital data.

To see the overall model, consider a sample linear demand curve: \( Q = aP + b \), where \( a = -5 \), and \( b = 100 \). Treat MC as fixed at 2. The solution is similar to that shown in Figure 1. Table 2 shows the optimal price values as the marginal advertising revenue rises above zero.

<table>
<thead>
<tr>
<th>Ad</th>
<th>P*</th>
<th>Q*</th>
<th>Ed</th>
<th>Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>11</td>
<td>45</td>
<td>-1.222</td>
<td>405</td>
</tr>
<tr>
<td>5</td>
<td>8.5</td>
<td>57.5</td>
<td>-0.739</td>
<td>661</td>
</tr>
</tbody>
</table>

Table 2: Sample linear demand solutions.

Notice that when the advertising rate is greater than marginal cost, demand becomes inelastic. This change makes the denominator in Equation (13) negative, so the overall price is still greater than zero, even though advertising is higher than the marginal cost. Because the demand is inelastic, the firm would need to make substantial price cuts to see large increases in quantity sold. With this example, price does not become negative until the advertising rate rises above 22—eleven times greater than MC. On the other hand, the original \( P^* \) when advertising is zero is relatively high, so it takes a high advertising rate to offset that initial price. In comparison, a similar quadratic example drops from a price of 10 down to a negative value when the advertising rate hits 14. The result is a consequence of the demand curve remaining elastic at low prices in the quadratic model.

Of course, ultimately the results depend on the specific form of the demand curve and its estimated coefficients. A very real question that remains is how consumers will react to negative prices. The examples to this point have assumed that the demand curve itself will respond similarly to negative and positive values of prices. Conceivably, some group of consumers might see the subsidy as a way to generate income and rapidly expand their use of the system. Answering this question requires looking at the consumer decision in more detail.

4. Consumers

One big question regarding demand curves is what happens to quantity demanded as price drops to zero, or below. Will customers demand an unlimited (or extreme) amount at a low or subsidized price? In a simple (2-commodity) model, this outcome is possible. The good news is that it would presumably drive up advertising revenue. But, experience with some Web services already indicates that even a zero prices demand is limited. Clearly, a more detailed model is needed to understand the limiting factors on consumer demand.

Economists model consumer behavior with a generic utility function. The simplest form contains two products. Evaluating digital strategies requires a
few more elements. Begin by considering two digital products: Q and Qs. As with the firm, these quantities can often be thought of in terms of quantity of bytes transferred. In most examples, the two products will be considered substitutes or competitors and the two can be compared using various assumptions. However, the consumer will also have a reference product Q0 that represents a collection of traditional goods. For the most part, this product is simply a reference item used as a relative measure because prices and demand for traditional products are already established. A utility function can be written based on these three fundamental products. The standard economic presumption is that consumers act to maximize their utility:

\[ U(Q, Q_s, Q_0, T(Q, Q_s), N) \]  
subject to their basic income (I) constraint:

\[ PQ + P_s Q_s + P_0 Q_0 \leq I - FC \]  
\[ FC \] represents their fixed cost of paying for Internet access. In most situations, it simply reduces the income available for marginal purchases, but it is listed here to remind the reader that it could play a role in some decisions—particularly mobile applications or those that require huge bandwidth.

The T function represents time required by the user to utilize the quantity. For example, some systems might be easier to use, resulting in less time spent obtaining the data. Similarly, the time function can represent search costs. For instance, music services such as iTunes make it easy for customers to find and obtain music relatively quickly. In contrast, with extensive searches, users might be able to find specific music elsewhere for free or reduced costs (possibly in violation of copyright laws). The time function is needed to compare these situations. In general, increases in time will reduce utility, so \( \partial U / \partial T < 0 \). Additionally, it is assumed that users have a time constraint:

\[ T(Q) + T(Q_s) \leq T^m \]  
The N variable/function represents any network effects where the value of a service increases with the number of other users. If the researcher needs to compare systems with different sets of users, N could be split into values for each of the services. The other implication of networks is that the marginal utilities will be different (\( \partial U / \partial Q \neq \partial U / \partial Q_s \)).

Most users also face a constraint on the amount of data that can be transferred within a given time period. With relatively slow connections (even 1-2 mbps is limiting), the total data that can be transferred can be expressed as a constraint:

\[ Q + Q_s \leq G \]  
\[ G \] As last-mile network speeds improve, this constraint may no longer be binding, but it should still be considered.

A final constraint is relevant to the digital world—particularly when negative prices (subsidies) are considered. Remember that in this model firms control prices and individuals determine the quantity to consume based on the price. If a firm uses a negative price, it most likely will impose limits on the payout—simply as protection from a few individuals who might try to “game” the system. The constraint to the user can be written:

\[ PQ \geq C \]  
In a traditional system, C would be zero—a constraint that is more typically written by requiring prices to be non-negative. In this more general model, C can be some negative value representing the maximum subsidy that a firm is willing to pay for a given period of time. If the competitor is also granting subsidies, a similar constraint could exist.

Utility defined in Equation (20) can be maximized, assuming the functional form meets the standard second-order conditions. The first-order conditions are:

\[ \partial U / \partial Q = \partial T / \partial Q = \mu_1 P - \mu_2 \partial T / \partial Q + \mu_3 Q + \mu_4 P = 0 \]  
\[ \partial U / \partial Q_s = \partial T / \partial Q_s = \mu_1 P_s - \mu_2 \partial T / \partial Q_s + \mu_3 Q_s = 0 \]  
\[ \partial U / \partial Q_0 = \mu_1 P_0 = 0 \]  
The additional Kuhn-Tucker conditions must also hold, including the constraints specified by Equations (19), (20), (21), and (22). Note that the multipliers (\( \mu \)) are numbered 1 through 4 from those constraints. Additionally, these multipliers must be non-negative, and the dual conditions must hold (\( \mu \times \text{constraint} = 0 \)).

At first glance, the equations in (23) do not mean much. They are slightly easier to understand if you subtract the second equation from the first and collect the terms:

\[ (\partial U / \partial T - \mu_2) (\partial T / \partial Q - \partial T / \partial Q_s) - \mu_3 (P - P_s) + \mu_4 P = \mu_3 (Q - Q_s) \]  
Equation (26) is a useful starting point for several analyses—which begin with various assumptions. A few examples are presented in the following sections.

4.1. Digital Music Example

To illustrate the value of the model, consider the value of a music service such as iTunes. The same model and assumptions would apply to similar content such as textbooks. P and Q will represent the price and quantity to the consumer. Ps and Qs are the relevant data for a substitute—such as searching the Web for free or reduced-price versions of the desired music. This description points out the key assumptions: (a) \( P > P_s \), and \( \partial T / \partial Q < \partial T / \partial Q_s \), which

A final constraint is relevant to the digital world—particularly when negative prices (subsidies) are considered. Remember that in this model firms control prices and individuals determine the quantity to consume based on the price. If a firm uses a negative price, it most likely will impose limits on the payout—simply as protection from a few individuals who might try to “game” the system. The constraint to the user can be written:

\[ PQ \geq C \]  
In a traditional system, C would be zero—a constraint that is more typically written by requiring prices to be non-negative. In this more general model, C can be some negative value representing the maximum subsidy that a firm is willing to pay for a given period of time. If the competitor is also granting subsidies, a similar constraint could exist.

Utility defined in Equation (20) can be maximized, assuming the functional form meets the standard second-order conditions. The first-order conditions are:

\[ \partial U / \partial Q = \partial T / \partial Q = \mu_1 P - \mu_2 \partial T / \partial Q + \mu_3 Q + \mu_4 P = 0 \]  
\[ \partial U / \partial Q_s = \partial T / \partial Q_s = \mu_1 P_s - \mu_2 \partial T / \partial Q_s + \mu_3 Q_s = 0 \]  
\[ \partial U / \partial Q_0 = \mu_1 P_0 = 0 \]  
The additional Kuhn-Tucker conditions must also hold, including the constraints specified by Equations (19), (20), (21), and (22). Note that the multipliers (\( \mu \)) are numbered 1 through 4 from those constraints. Additionally, these multipliers must be non-negative, and the dual conditions must hold (\( \mu \times \text{constraint} = 0 \)).

At first glance, the equations in (23) do not mean much. They are slightly easier to understand if you subtract the second equation from the first and collect the terms:

\[ (\partial U / \partial T - \mu_2) (\partial T / \partial Q - \partial T / \partial Q_s) - \mu_3 (P - P_s) + \mu_4 P = \mu_3 (Q - Q_s) \]  
Equation (26) is a useful starting point for several analyses—which begin with various assumptions. A few examples are presented in the following sections.

4.1. Digital Music Example

To illustrate the value of the model, consider the value of a music service such as iTunes. The same model and assumptions would apply to similar content such as textbooks. P and Q will represent the price and quantity to the consumer. Ps and Qs are the relevant data for a substitute—such as searching the Web for free or reduced-price versions of the desired music. This description points out the key assumptions: (a) \( P > P_s \), and \( \partial T / \partial Q < \partial T / \partial Q_s \), which

A final constraint is relevant to the digital world—particularly when negative prices (subsidies) are consid
they search online for free music? In terms of the will customers actually pay for the service, or will the proposed service will be greater than the quantity on the left side of Equation (26).

One question to answer is whether the demand for the proposed service will be greater than the quantity demanded of the substitute service. In other words, will customers actually pay for the service, or will they search online for free music? In terms of the model, the question is whether $Q - Q_s > 0$. This condition will be true if the left side of Equation (6) is greater than zero, which is true if:

$$\frac{\partial U}{\partial T} - \mu_1 \left( \frac{\partial T Q}{\partial T} - \frac{\partial T Q s}{\partial T} \right) > \mu_1 (P - P_s)$$

Recall that $\partial U/\partial T < 0$, so the first term on the left is negative. The second term on the left is also negative because of the assumption that the substitute/free service has higher search and usability costs. To simplify the equation, assume that the consumer is not facing a binding time constraint, so $\mu_2$ is zero. Now, apply a traditional microeconomics trick and divide Equation (27) by the third equation in Equation (25) to get a reference:

$$\frac{\partial U}{\partial Q} \left( \frac{\partial T Q}{\partial T} - \frac{\partial T Q s}{\partial T} \right) > \frac{(P - P_s)}{P_0}$$

The ratio of marginal utilities is called the marginal rate of substitution (MRS). Equation (28) is interpreted as stating that a person will buy songs from the service if the relative value of the time saved is greater than the relative price differential. This conclusion might seem somewhat obvious (now that iTunes has been shown to be successful). However, it is helpful to have a model that actually proves this conclusion. Further, if the relative value of time can be measured for various people, the model could be used to identify target markets. That is, P can be higher (or the amount purchased) for users who place a relatively high value on their time. Additionally, the model shows that price (P) could be increased if the search cost differential could be increased. That is, if the main service can be made faster and easier to use, or legal actions are taken to make it harder to find low-cost, generally illegally shared files, then consumers will increase purchases from the service (or allow the price to be increased).

4.2. Web Service Example

The model can also compare Web services, where the focus is on providing a service instead of a specific product. Search engines are a common example. Here, price and usability become critical. If the two services are equivalent in both price and performance, the consumer will be indifferent between them. So, assume that a company wishes to start (or expand) a new service to compete with an existing firm. The new company chooses to do so by charging a lower price—even negative if the competitor has a zero price. That is, $P < 0 \leq P_s$. Two situations arise at this point. The new service could be equally efficient to the competitor or it could be less efficient ($\partial T Q / \partial T Q_s > \partial T / \partial T Q_s$).

Consider the simpler equality case first, and Equation (26) reduces to:

$$0 = \mu_1 (P - P_s) + \mu_4 (Q - Q_s) - \mu_4 P$$

Again, the goal is to determine the sign of $(Q - Q_s)$, so rearrange and focus on it:

$$(Q - Q_s) = (-\mu_1/\mu_4) (P - P_s) + (\mu_4/\mu_4) P$$

If $(P_s > P \geq 0)$, the difference in prices in Equation (30) is negative, but multiplied by a negative factor, making the entire term positive. The last term is also positive (or zero), so Q is always larger than $Q_s$. In other words, if the competitor is charging a price for a service, and the new firm can provide the same level of service usability at a lower price, consumers will switch to the new service.

However, if $P < 0$, the last term in the equation becomes negative—potentially offsetting the gains from the price differential, depending largely on the value of $\mu_4$. This result seems somewhat counterintuitive, but it is rational. If the competitor price is zero, the startup firm can only undercut the price by going below zero. If the firm imposes a limit on the amount of subsidy that a consumer can receive, the value of the price differential is mitigated. In words, if a firm tries to capture market share by offering subsidies to undercut a competitor, the ultimate success depends on the value of the limits imposed on the subsidy.

The more realistic situation arises when the new service is less efficient than the competitor’s. In this case, for the left side of Equation (26) to be positive, it can be rewritten based on the signs of the various terms to:

$$\mu_1 (P - P_s) > - (\partial U/\partial T - \mu_2) \left( \partial T Q / \partial T Q_s - \partial T / \partial T Q_s \right) - P$$

Again divide each side by the reference real-world values (which must be positive) to remove the $\mu_1$ term and obtain:

$$\frac{(P - P_s)}{P_0} > - \frac{(\partial U/\partial T - \mu_2)}{\partial U/\partial Q} \left( \frac{\partial T Q}{\partial T} - \frac{\partial T Q s}{\partial T} \right) - \mu_4 (\partial U/\partial Q_0) P$$

For simplicity, if the time and subsidy constraints are not binding, Equation (32) can be interpreted as stating that the new firm (offering a subsidized price) will capture more sales if the relative price differential (subsidy) is greater than the service inefficiency multiplied by the MRS of time versus other consumption. Or, stated similarly, if the marginal value (utility of time lost divided by price)
is less than the marginal value of other consumption (utility of consumption divided by price). Most of the results ultimately depend on the marginal utility of time (relative to the marginal utility of other consumption). Highly time-sensitive individuals will require greater price differentials to use a particular service. Conversely, a system that is more efficient (easier to use, faster results, more relevant answers, and so on) has a substantial advantage over competitors. In practice, a firm will want to estimate the time sensitivity ($\partial U / \partial T$) of its potential customers as well as the perceived difference in time efficiency between its system and the competitor’s systems ($\partial T / \partial Q - \partial T / \partial Q_s$).

As a newcomer, the firm charges a lower price ($P < P_s$). Green’s conjoint analysis [5], and many related papers, provide methods to utility part-worths.

4.3. Social Network Example

The basic model can also be used to evaluate firms that rely on a network effect to attract new customers and usage. Network effects exist when the number of other users is common examples—and many struggle to reach a critical mass of users. Adding the network effect means that the utility of the quantity consumed is filtered by the number of other users. Treating the network ($N$) explicitly (it could be simply embedded into $\partial U / \partial Q$) replaces the main first-order condition in Equation (24) with:

$$\frac{\partial U}{\partial Q} + \frac{\partial U}{\partial T} \cdot \mu_1 P - \mu_2 \frac{\partial T}{\partial Q} - \mu_3 Q + \mu_4 P = 0$$

For simplicity, assume that the services are relatively similar, specifically that the time issues (usability) are equivalent to remove the time effect relatively similar, specifically that the time issues (usability) are equivalent to remove the time effect. Again, subtract the equivalent expression for $Q$, from Equation (33) to get:

$$\frac{\partial U}{\partial N} \left(\frac{\partial Q}{\partial Q} - \frac{\partial N}{\partial Q_s}\right) - \mu_1 (P - P_s) - \mu_2 (Q - Q_s) = 0$$

Success depends on whether $(Q - Q_s)$ is greater than zero, which is true if Equation (34) is rearranged:

$$\frac{\partial U}{\partial N} \left(\frac{\partial Q}{\partial Q} - \frac{\partial N}{\partial Q_s}\right) > \mu_1 (P - P_s)$$

Replacing $\mu_1$ with its marginal utility valuation results in:

$$\frac{\partial U}{\partial Q} > \frac{\partial N}{\partial Q_s} \frac{(P - P_s)}{\partial N}$$

Assume that the new company has a smaller network than the existing firms, so $\partial N / \partial Q < \partial N / \partial Q_s$. As a newcomer, the firm charges a lower price ($P < P_s$). Because both sides of the inequality are negative, dividing by -1 reverses the inequality. Hence, to convince consumers to use the new system, the relative price differential must be greater than the MRS for network value times the network shortfall.

If the network effect is important, the price differential might have to be substantial. Again, if the competitor is already charging a zero price, the newcomer would have to resort to subsidies to attract customers. On the other hand, the assumption of similar networks can be dropped. In which case, a newcomer with a more time-efficient application would not need as low of a price to attract customers.

4.4. Infinite Demand Question

This approach can be used to analyze a variety of different digital applications. In all cases, price is going to be critical—which means it is quite likely that new competitors might have to resort to negative prices (subsidies) to gain market share. (Note that better, more efficient systems are a more effective answer, but are not always possible.) The section on business decisions shows that advertising revenue makes it possible to undercut prices—even below zero. But, at negative prices, will consumers flood a system and require an infinite (huge) amount of bandwidth? In most cases, the answer is “No.” Particularly when applications have different time effects, consumers will have different sensitivities to time effects, so some will shift quickly based on price, while others will require larger price differentials. Assuming a reasonable distribution of time preferences, overall demand should be imperfectly responsive to prices.

Further, individual consumers face additional constraints. Overall time and bandwidth considerations limit the amount of data that a single individual can consume in a specified period of time. Nonetheless, firms experimenting with price subsidies should always define a constraint on the total amount that can be paid to an individual. Any system will face at least some consumers for whom time and byte constraints are minimal, and who place a low value on time efficiencies; resulting in excessive increases in quantity demanded for even small reductions in price. Although ad revenue might increase because of the increased byte-flow, it also might be constrained because of uniqueness constraints or because the heavy users might ignore ads, reducing the click-through rate.

5. Summary and Limitations

The model presented in this paper is capable of evaluating and explaining several issues in the digital domain that affect businesses and consumers. Several
issues of digital products and services are unique in the economic world, including the possibility of zero marginal costs, negative prices, and network effects.

Even in general form, the model explains and proves behavior in characteristic examples including negative prices for digital services, responses to network effects, and pricing of digital products such as online music. The paper shows how the model imposes constraints on various applications, such as limitations inherent in linear demand curves.

The model can be applied to evaluate many business situations. It is straightforward to apply real data to estimate the specified functions and parameters to reach conclusions in specific situations.

The model differs from existing research in a few key areas. These differences arise because of different assumptions, which lead to somewhat different results and conclusions. In particular, revenue from “advertisers” has a different effect than in typical models of the two-sided market approach which emphasizes network effects. And this new model assumes the Web service company has minimal economic power in the advertising market.

The important point is that the models differ because of the assumptions, largely because they were designed to examine different types of effects. There is always a need for different theoretical models. Ultimately, the model is chosen that best fits the conditions of the underlying problem.

For instance, digital services (e.g., search engines) fit better with this new model than with the two-sided market models because buyers will repeatedly use the service—specifically for the service instead of the advertisements. Hence, quantity needs to be a critical element in the model.

Because it is based on standard economic theory, and it aligns with other models, this one can be extended in several directions. As indicated in section 4, network effects can be incorporated. Likewise, the model could be extended closer to the two-sided market model, where the quantity effects of this model are merged with the dual network effects.

As emphasized by [1], advertising can be handled in different ways depending on the underlying function of the advertising. Beyond extending to the two-sided market with its imperfectly competitive advertising market, a question remains about how consumers value advertising. In a model where ads are considered informative, consumers might place a positive value on seeing an ad the first time, but declining or even negative utility for additional ads.

6. References


