Convenient Economics: The Incorporation and Implications of Convenience in Market Equilibrium Analysis

George Davis
Virginia Tech University

JEL Codes: D1, D2
Keywords: Convenience, cost shifting, food, household production, scale economies, scope economies

Abstract
Convenience is perhaps the most important “commodity” being sold in the market today, and yet there is nothing of analytical substance to be found in most undergraduate textbooks. The purpose of this article is to provide a straightforward framework for teaching students the economics of convenience, utilizing the standard tools of introductory and intermediate microeconomics. The framework is used to answer several questions related to convenience that could not be answered with the typical supply and demand framework found in undergraduate textbooks. The key analytical features of the article are provided in a complementary PowerPoint file in the online supplementary appendix.

1 Introduction
One does not have to be an economic savant to recognize that we live in an “on-demand culture” (Fromm 2019) and “convenience is the ultimate currency” (Nielsen 2018). Everywhere you look companies are extracting economic rents from the mine of convenience: home robotic devices, voice-activated devices, virtual use devices, driver assisted technologies, finger print and face recognition access, online and automatic bill pay, personalized ads based on shopping history, and even price drop notification features online. Convenience is especially prevalent in the food sector: in-store ordering kiosks, personal checkouts at grocery stores, online food shopping and delivery or pick-up service, meal kit delivery services, a continuum of pre-prepared foods, and touchscreen smart refrigerators. One study estimates that, “on average consumers are willing to pay 11 percent more for each layer of convenience in the food chain in anything from online grocery delivery to restaurant take out” (Findling 2017).

Given the ubiquity of convenience in the marketplace, one would expect our textbooks to be replete with at least chapters or sections on convenience. That is not the case. Perusing some of the top selling undergraduate microeconomics textbooks reveals there is nothing of analytical substance on convenience (via Amazon: e.g., Mankiw 2012; Sowell 2015; Krugman and Wells 2018). This significant analytical gap is important because it leads to an inability of students to understand the economics of convenience; its implications on decision making and the standard variables of interest: market prices and quantities.

For example, here are just a few questions in the food sector that are difficult to address with the standard economics found in undergraduate texts but are easily addressed with the economics of convenience covered in this article.

• Is store location more important than store prices in choosing a store?
• Why do single-headed households demand more convenience and eat out more than dual-headed households?
How do transaction assisted devices affect markets?
What is the common link between online, offline grocery markets, meal kits, and grocerants?
Why are prices in food deserts higher than in nonfood deserts?

The purpose of this article is to provide a straightforward framework for teaching students the economics of convenience utilizing the standard tools of introductory and intermediate microeconomics. The framework can be used to analyze the impact of convenience on market prices and quantities in a straightforward extension to typical supply and demand diagrams. Methodologically, a general Stigler and Becker (1977) framework is followed by incorporating convenience within a broader set of resource constraints from advanced consumer and retail supply theory. As the student should know, demand and supply analysis rests on the ideas of agents optimizing an objective(s) (i.e., utility, profit) subject to constraints. On the demand side, because there can be a preference for convenience, it needs to enter the utility function. However, time and effort are also limited resources, and given that convenience saves these resources, it also enters into constraints. On the supply side, given that consumers value time and effort, firms may seek to provide convenience and effectively shift the cost of inconvenience from the consumer to the firm. This must occur within the context of profitability (i.e., revenue and cost impacts).

The framework presented is methodologically progressive because it has “excess explanatory content” over explanations that appeal simply to preferences, behavioral biases, or irrationality (e.g., Lakatos 1993; Davis 1997). For example, the framework creates an intuitive link between classical economics and the exciting new field of neuropeconomics. Neuropeconomics “combines research from neuroscience, neurobiology, and economics [and] . . . provides parsimonious models of decision making capable of delivering qualitative behavioral predictions” (Brocas and Carrillo 2008 p. 175). By embedding several of the key concepts from neuropeconomics within an extended framework of standard tools, topics in neuropeconomics can be easily introduced and discussed with students because they become novel applications of well-known concepts.

The next sections define convenience, present the demand side, then the supply side, and then brings them together to analyze some of the previous questions. Given the target audience is a typical undergraduate course, convenience is incorporated within the context of a perfectly competitive model (supply and demand). It is certainly recognized that the appropriateness of this depends on a host of factors: the questions of interest, the market, the degree of spatial, temporal, and product aggregation, and so on. As such, the conclusions provide some discussions and guidelines for extending the analysis to imperfectly competitive settings. Finally, the key analytical features of the article are provided in a complementary PowerPoint file in the online supplementary appendix.

2 Defining Convenience in the Market

Convenience is normally defined as saving time, but it can also include saving effort, physical and mental. Two activities can require the same amount of time but very different effort levels. A 15-minute walk does not require the same effort as a 15-minute run, or shopping online for an hour does not require the same effort as shopping offline for an hour. The importance of recognizing both physical and mental effort in economic analysis has a long history, especially in the study of wealth and labor.\footnote{Early economic textbooks, such as Marshall (1920) and Taylor (1913) and more recently Becker (1985) all included discussions of both time and effort. For example, Marshall (1920, p. 76) states, “the theory of wants can claim no supremacy over the theory of efforts.” Taylor (1913, p. 1) states, “it is a fact obvious to everyone that wealth is a thing which absorbs a very large amount of our time, thought, and effort.”} Mental effort has been called “psychic cost” (e.g., Sjaastad 1962; Ingene 1984; Rosen 1986), or within the fields of behavioral economics and neuropeconomics, it is closely related to the concept of cognitive load (e.g., Sweller 1988; Camerer, Loewenstein, and Prelec 2005).

Starting from the basics, all economic transactions require four steps: (i) information acquisition (e.g., who, what, where, how much), (ii) payment acquisition (e.g., in-kind, cash, credit, electronic), (iii) good acquisition (e.g., at purchase, delivery), and (iv) possible good transformation (e.g., used as input into
producing something else). Convenience is considered “the ultimate currency” because, like a currency, it is absent or present in each of these steps. And each of these four steps requires labor, both in time and effort, by both consumers and producers, and this is how convenience enters the market. Consequently, at each step there are potential opportunities for both consumers and producers for saving time and effort. Because time and effort are valuable resources, they have associated with them implicit opportunity costs and thus value in being saved.

3 Demand Side
With this background, a theory that incorporates convenience both in the utility function and in the resource constraints is desirable, and Nobel Laureate Gary Becker’s (1965, 1985) Household Production Theory (HPT) is well suited for this task.

On the preference side, Becker’s HPT is based on the observation that individuals do not get utility directly from goods purchased in the market, but rather use these goods as inputs, in combination with other inputs (e.g., time and effort), to produce commodities that give direct utility (i.e., step four above). This is a very old idea dating back to Bentham (1963), who identified 15 basic pains and pleasures the individual produces (e.g., warmth, shelter, nutrition, safety, etc.).

On the constraint side, students should know the core concept of allocative efficiency: a resource is allocated for an objective efficiently (without waste) via a (cost) price system. In undergraduate classes we tend to focus only on money, but time and effort are two equally important constraints. Specifically, based on the computational view of the brain from psychology (e.g., Edelman 2008), the concept of allocating limited cognitive resources is now well established in the literature (e.g., Alonso, Brocas, and Carrillo 2014; Kool et al. 2010; Kool and Botvinick 2014). Any decision task will have associated with it a cognitive load. A high cognitive load task requires more cognitive resources than a low cognitive load task (e.g., doing your taxes vs. doing your nails). Cognitive load plays a key role in the utilization of cognitive resources and also in the dual system view of the brain. Dual system processing consists of a fast system (system 1) that uses little cognitive resources and a slow system (system 2) that uses more cognitive resources (e.g., Kahneman 2011). The basic principles of allocating a scarce resource then apply whereby perceived benefits and costs are compared. One of the main findings in this literature is that many decisions are made in the context of trying to conserve cognitive resources so there is a tendency to use the fast system for decision making if possible, ceteris paribus. Consequently, the inclusion of an effort constraint is a parsimonious and intuitive way to connect the standard toolbox to the neuroeconomics literature.

Along these lines, Becker (1965, 1985) defines the full income constraint, which can consist of money, time, and effort. Associated with the full income constraint are full prices that consists of two parts: a direct price and an indirect price. Recall in the context of a constraint, the price represents how much of a resource must be given up (the opportunity cost) to get one unit of the good or activity. The direct price is simply the price associated with the money constraint. However, for any other resource constraint, such as time and effort, there will be an indirect or shadow price as well. The student will probably recognize the idea of a full price, if not the name, if they are familiar with the economics of a negative externality, such as steel production generating pollution. In that context, the marginal social cost of pollution is an indirect cost of steel, and when added to the market price, gives the full cost of steel production. In the typical supply and demand graph, this will be shown as a shift up in the supply curve that is attributed to the marginal social cost of pollution per unit of steel produced. This idea can be generalized for distinguishing between the market price and the full price and is a powerful general construct that allows for incorporating many other types of costs within the typical supply and demand

---

2 Davis and Serrano (2016), chapter 10 provide much more detailed development and discussion of dual system decision making in a food context.

3 A closely related broad term for indirect costs not priced in the market is transaction cost, but the transaction cost literature tends to focus on the implications for industrial organization not households (see Pollak 1985).
diagram. However, the incorporation needs to be done in a non-ad hoc and theoretically consistent fashion that extracts all the potential explanatory power and applications.

More formally, let \( i \) denote the individual, \( j \) the location, and \( k \) the good. Utilizing an undergraduate version of Becker’s (1965, 1985) model, the individual \( i \) receives utility from \( K \) commodities \( Z_{ij1}, Z_{ij2}, \ldots, Z_{ijk} \). In a food context, obviously one of the commodities could be a meal. The commodities are produced by the individual using market good inputs \( Q_{ij1}, Q_{ij2}, \ldots, Q_{ijk} \), own time inputs \( T_{ij1}, T_{ij2}, \ldots, T_{ijk} \), and effort inputs \( E_{ij1}, E_{ij2}, \ldots, E_{ijk} \). In a slight generalization of Becker’s (1965) basic model, the commodity production technology has the form:

\[
Q_{ijk} = a_{ijk}Z_{ijk} \quad (1)
\]

\[
T_{ijk} = b_{ijk}Z_{ijk} \quad (2)
\]

\[
E_{ijk} = c_{ijk}Z_{ijk} \quad (3)
\]

A unit of \( Z_{ijk} \) produced requires a triplet combination of goods, time, and effort. The technology parameters \( a_{ijk}, b_{ijk}, \) and \( c_{ijk} \) for converting goods, time, and effort into commodities are individual \( (i) \), location \( (j) \), and good/commodity \( (k) \) specific. As will be shown, this generalization proves very useful in analyzing several forms of convenience. The parameters \( b_{ijk} \) and \( c_{ijk} \) capture the idea of convenience in both time and effort. Specifically, \( b_{ijk} \) is the amount of time (the quantity), and \( c_{ijk} \) is the amount of effort (the intensity) required for individual \( i \) in location \( j \) per unit of \( Z_{ijk} \) produced. So, in this “household production” context, all the standard economic intuition from production theory related to biased technology change can be utilized because a change in the technology parameters \( a_{ijk}, b_{ijk}, \) and \( c_{ijk} \) can be thought of as technological change. Consequently, a decrease in one of these parameters means that less of the input (market good, time, or effort) is required to produce the same level of the commodity, and the new technology is input “saving.” It is important to recognize that (2) and (3) refer to the total time and effort required, which may be composed of many categories that are added together, such as planning, travel, and shopping time/effort so the saving may occur in any one of these categories or several.4

Regarding the resource constraints, there are three: (1) an expenditure (money) constraint, (2) a time constraint, and (3) an effort constraint. The full income constraint is derived from recognizing that money income comes from converting both work time and effort into money via the labor market. This implies the two main constraints are time and effort:5

\[
T_i = T_{iw} + \sum_j \sum_k T_{ijk} \quad (4)
\]

\[
E_i = E_{iw} + \sum_j \sum_k E_{ijk} \quad (5)
\]

where \( T_{iw} \) and \( E_{iw} \) is the quantity of time and effort spent in market work, respectively. The full income constraint is then:

\[
\sum_j \sum_k P_{jk} Q_{ijk} = W_i T_{iw} + R_i E_{iw} + V_i \quad (6)
\]

where \( P_{jk} \) is the market price the individual faces in location \( j \) for the \( k \)th market good, and \( W_i \) is the individual’s hourly wage rate or opportunity cost per unit of time. The variable \( R_i \) represents the dollar value for a unit of effort or the cognitive load. It is this term that links the standard toolbox with key

---

4 Becker (1965) allows for this by using vector notation for the technology constraints.
5 For the student, the notation \( \sum \) is the Greek letter for “S” and is just shorthand notation saying “S”um over all types of goods (the \( k \)s) and over all locations (the \( j \)s).
The variable $V_i$ is unearned income. Substituting (1) – (5) into (6) and rearranging yields the full income constraint expressed in full prices or:

$$Y_i = \sum_j \sum_k \Pi_{ijk} Z_{ijk}$$  \hspace{1cm} (7)

where:

$$\Pi_{ijk} = a_{ijk} P_{jk} + b_{ijk} W_i + c_{ijk} R_i$$  \hspace{1cm} (8)

is the full price of the commodity $Z_{ijk}$ and $Y_i \equiv W_i T_i + R_i E_i + V_i$. The first term of the full price ($a_{ijk} P_{jk}$) represents the direct price, and the next two terms ($b_{ijk} W_i + c_{ijk} R_i$) represent the indirect price. Numerous economic insights are already forthcoming by taking a closer look at the full price and its components in equation (8).

### 3.1 The Full Price Principle

The same or lower full price does not mean the same or lower market price and vice versa. Because the full price consists of three separate components, there is an infinite number of combinations that can lead to the same or even lower full price. A high direct (market price) component ($a_{ijk} P_{jk}$) can be offset by a low indirect component ($b_{ijk} W_i, c_{ijk} R_i$). Alternatively, the same or even lower direct component ($a_{ijk} P_{jk}$) can be offset by a higher indirect component ($b_{ijk} W_i, c_{ijk} R_i$) leading to higher full price. Knowing the value of the full price $\Pi_{ijk}$ tells you nothing about any of the values of the subcomponents ($a_{ijk} P_{jk}, b_{ijk} W_i, c_{ijk} R_i$) and vice versa. Perhaps most importantly, the triple subscript notation implies these equalities or differences can be due to individual ($i$), location ($j$), or good ($k$) equalities or differences or some combination.

### 3.2 Some General and Specific Applications of the Full Price Principle

Consider then some general and specific applications of the full price principle. Comparing across goods, the principle implies two goods in different locations can have the same full price ($\Pi_{111} = \Pi_{222}$) but different market prices ($P_{111} \neq P_{222}$) because some other components of the full price differ ($a_{ijk}, b_{ijk}, c_{ijk}, W_i, R_i$). In fact, many market options may not only have a lower indirect time price ($b_{ijk} W_i$) but also a lower indirect effort (cognitive) price ($c_{ijk} R_i$) such that an individual is willing to pay a higher direct (market) price ($a_{ijk} P_{jk}$) because the full price ($\Pi_{ijk}$) will be the same or even lower. This result is ubiquitous in the marketplace. For example, a common phenomenon that plays out every weekend all over the world is individuals go out to eat, walk into a restaurant without a reservation, and ask what is the wait time (que) for seating. If the que is too long, they go to another restaurant in hopes of a shorter que. Individuals will often be willing to pay more for the meal ($P_{222} > P_{111}$) if the que is shorter, and this is captured by the full price because the full price between the two restaurants can be equal ($\Pi_{222} = \Pi_{111}$) even though the time and effort prices differ. In a recent study De Vries, Roy, and Koster (2018) found that longer wait times relate to a longer time to customers returning, a shorter dining duration (i.e., trying to keep the full price lower).

---

6. Treating the cost of a unit of time and a unit of effort as not good/activity specific is a rather standard simplifying assumption that could be relaxed adding more complexity without a great deal more insight. The key point is the quantity of time and effort each has an implicit cost and one should not confuse the per unit cost with the quantity. Different activities certainly require different quantities of time and effort and so the expenditure per good/activity will differ.

7. This is just an application of the concept of compensating differentials from labor economics dating back to Adam Smith (Rosen 1986).

8. The astute student may recognize this as just an application of the more general algebra rule of more unknowns (six) than equations (one) being undetermined, and so nothing can be said about the values of the unknowns.

9. It is likely the student has already been exposed to this idea in a strictly spatial setting if they have been exposed to the “law of one price” where once transportation costs are taken into account, the prices of the same product from two locations are equal. This is just a generalization of that concept.
low), and higher revenue for shorter wait times. The principle applies to any case where there are substitute products that have a shorter wait time: from semi-processed ingredients versus basic ingredients in a homemade meal (Yang, Davis, and Muth 2017) or delivery versus takeout. Another example would be search costs stemming from information searches as an example of the indirect price components. From this perspective, firms with good reputations or with well-known national brands and advertising reduce search costs and thus can have a higher good price but not a higher full price (Stigler 1961; Ehrlich and Fisher 1982; Pashigian and Bowen 1994). Other examples of this are ubiquitous as well. For example, even if you have enough expertise to do your own taxes, you may pay an accountant to do them because you view the full price as cheaper from the accountant than doing them yourself because you attach a high value to your time and the cognitive effort. You may pay more for a product online because once the time and effort costs are taken into account, the full price is cheaper than offline shopping. Other examples are a lawn service that cuts your grass, in-home cleaning services, and so on. Generally stated, an individual can be willing to pay a higher direct price if the indirect price associated with the good is lower, leaving the full price the same or even lower.

As the principle suggests, the logic works in reverse as well. Two market goods can have the same direct market prices ($P_{11} = P_{22}$) but different full prices ($\Pi_{11} \neq \Pi_{22}$) because some other components of the full price differ ($a_{i1k}, b_{ijk}, c_{ijk}, W_i, R_i$). For example, a grocery store across the street from your house versus one a mile away may have the same prices and indeed may be part of the same chain, but the one closer to home will have the lower full price simply because of the lower time cost. Marshall and Pires (2017) find that store convenience is a more important determinant of store choice than prices, lending support to the importance of full prices over good prices.

Although this is a useful result for comparing different goods or locations, it is perhaps even more useful for helping explain differences across individuals within the same household because the technology parameters are not only location and good specific, they are also individual specific. Consider then the case of a dual-headed household, where individual one is more productive than individual two in producing the commodity, say a meal ($b_{1jk} < b_{2jk}$). Even if all other elements of the component prices are the same, individual one will have a lower full price than individual two ($\Pi_{1jk} < \Pi_{2jk}$), and thus if the household is minimizing cost of production, individual one will produce the meal. In this context, the household may still consume a meal produced at home because individual one has a production technology that makes it cheaper than eating food away from home, *ceteris paribus*. Thus “no matter how the members divide family resources between the two members, each member agrees to choose the most efficient shopper [producer] for each of the goods that the family purchases . . . The efficient solution requires the member with the lowest minimized full price be the shopper” (Pashigian and Bowen 1994, p. 39). This is essentially just an example of the insights that production efficiencies can achieve by division of labor, as stated in the first sentence of *The Wealth of Nations* (Smith 2010), and provides insights on household organizational structure (Pollak 1985).

Note what this would imply for the lack of intrahousehold “trade” opportunities for single-headed households. In 1960, about 5 percent of households had only one person, and 9 percent of children lived in single-headed households. By 2018, these numbers were 30 percent and 27 percent, respectively (U.S. Census Bureau 2019). A single-headed household can only compare their full price of a meal at home to a meal prepared away from home, not to a perhaps cheaper full price from a partner. The theory would predict therefore that, *ceteris paribus*, single households would demand more convenience, spend less time at in-home food production, and eat out of the home more frequently, which is what has been found in the literature (e.g., Dave et al. 2009; Anekwe and Zeballos 2019; Byron 2019; and You and Davis 2019).

### 3.3 Demand Function and Curve

Proceeding to the demand function for the market good, first note that optimization of the utility function subject to this full income constraint leads to the demand functions for the commodities of the general form:
The parenthetical sign under each variable indicates the direction of the relationship between the variable and the quantity demanded, so as the full price of commodity one \( P_{i11} \) increases (decreases) the quantity demand of commodity one \( Z_{i11} \) decreases (increases), ceteris paribus. The question marks under the other full prices \( P_{ijk} \) indicates the directional relationship will depend on if other commodities are substitutes (+ sign) or complements (- sign) and under the income \( Y_i \), if the commodity is a normal (+ sign) or an inferior good (- sign).

However, the main question of interest is how does convenience affect the market demand for the market good? Remember the underlying framework is household “production,” and therefore, the market good is an input used in production of the commodity so the market good demand is derived demand. Furthermore, the interest is in how the different components of the full price affect the market demand for the market good, so we can proceed as follows. First, substitute the full price for good one from (8) into the individual demand function (9) and substitute the result into (1). Next, recall the market demand is an aggregation of individual demands, so drop the \( i \) subscript such that the variables will be market level variables. Finally, just let the bold variable \( O^D \) represent a list of all the other variables (a vector) not related to the full price of good one and the list would now include other full prices, income, population, and perhaps other factors, such as seasonal variables. The market demand for good one in location one can then be written in general form as:

\[
Q_{11}^D = Q_{11}(P_{11}, W, R, a_{11}, b_{11}, c_{11}, O^D): \text{ Market One Derived Demand Function (10)}
\]

As seen from (8), all the components of the full price \( (P_{11}, W, R, a_{11}, b_{11}, c_{11}) \) will tend to increase the full price and, given the law of demand in the full price, anything that increases the full price will decrease the demand for the commodity and thus decrease the derived demand for the market input. This is the reason all the component variables have a parenthetical negative sign.

As Stigler and Becker (1977, p. 89) point out, a movement along the commodity demand function is captured by a shift in the market good demand function. Why? Recall, demand curves show the relationship between own price and quantity demanded. A movement along the demand curve shows the relationship between the own price \( P_{11} \) and the quantity demanded \( Q_{11} \). The change in some other variables, other than own price, is then captured by a shift in the demand curve or simply stated a change in demand. For example, if the time it takes to purchase a good increases, say waiting in line, the demand for the market good will decrease or shift to the left, ceteris paribus. Moreover, if any of the other component variables \( (W, R, a_1, b_1, c_1) \) increase (decrease), the demand curve for the market good will decrease or shift in (increase, shift out), ceteris paribus.

4 Supply Side

On the supply side, especially retail supply, there are two important interrelated concepts that are associated with convenience economics: (i) economies of scale and scope and (ii) cost shifting.

4.1 Economies of Scale and Scope

Most students should be familiar with the concept of economies of scale. Economies of scale occur when cost per unit (average cost) decreases as the operation is scaled up or output increases. Economies of scale can occur for multiple reasons. It may be because of spreading out a fixed cost, such as a machine. For example, once a printing press is bought, the total cost per unit to print 40 papers is much higher than to print 4,000 because the main additional cost is the paper and ink. Alternatively, it may be because of
efficiency gains in repetition and specialization of laborers in their tasks or cost reduction associated with
buying or processing bulk orders of inputs. Economies of scope exists when the average cost of producing
two or more goods together in one place is less than the cost of producing them at separate locations.
This can be because of specialized “knowhow” within the firm that can be utilized in all goods sold, or
there is shareable input across the goods (Teece 1980; Panzar and Willig 1981), such as a production
facility or simply some managerial or labor expertise. A car or guitar manufacturer may produce many
different models that have a lot of the same common elements (e.g., engine size or guitar neck). A
stocking or checkout clerk, and all the associated mechanization, can stock or checkout a can of soup just
as easily as they can stock or check out a can of beans. Thus, the underlying total cost is relatively
constant, but the average cost per item is decreasing because the number of items is increasing.

4.2 Cost Shifting
Retail supply theory provides a very intuitive way to handle convenience utilizing all the standard tools
(e.g., Betancourt 2004; Bronnenberg 2018). In the basic supply and demand framework, prices and
quantities are common variables to both the producers and consumers in decision making. Retail supply
theory effectively extends this analysis to include the household technology parameters. The key is to
recognize that the retailer can affect the technology parameters in the household production technology
\((a_{ij}, b_{ijk}, c_{ikj})\) by providing “distribution services” that are designed to change these technology
parameters. This is known as cost shifting in the retailing literature because firms effectively take on
some of the costs the individual would normally incur in the production of the commodity (e.g.,
Betancourt 2004, p. 8). Cost shifting can occur in any of the four basic transactions, ranging from simply
providing the consumer some information, to delivering a good, to a central buying location, to more
processing to make the good closer to a commodity. In the food sector, food delivery, bagged salads, meal
kits, or any pre-prepared meal are all examples where the retailer incurs some of the cost the consumer
would normally incur, in an effort to hopefully increase profits.

The cost of distribution services has long been recognized as implicit in the standard supply
analysis, but are often overlooked.\textsuperscript{10} Recall the price on the supply curve is the minimum price required
to bring the good to the consumer and that must include all costs. Within a supply and demand diagram,
the market clearing price and quantity occur where the good is sold, not just produced.

Cost shifting can be thought of as a form of technological change, which is often categorized as one
of two types: (i) technology push or (ii) demand pull. Technology push is driven by an internal innovation
of the firm designed to reduce the cost of production or distribution with no direct impact on demand.
For example, reducing the number of checkout clerks by installing more self-checkout scanners, after
paying for the scanners, would decrease labor costs and thus would be a technological push change.
Alternatively, demand pull technology change occurs because of a perceived profit opportunity through a
potential increase in demand and may be associated with an increase in cost (Kamien and Schwartz 1982,
chapter 2). In the present context, a demand-pull innovation is a form of induced innovation. Induced
innovation occurs when the high cost of a factor of production induces an innovation to reduce the use of
that factor (Hicks 1932). In the present context, individuals’ high time and effort cost induces firms to
produce new time- and effort-saving technologies for individuals. Installing an in-store bakery and hiring
bakers would be a demand push technological change.\textsuperscript{11}

\textsuperscript{10} Marshall (1920), in his principles book, discussed the difference between the costs of production versus the cost of
“acquiring” a market (Marshall 1920, p. 239). Chamberlain (1962) spends an entire chapter (chapter 5) discussing the
difference between production costs versus selling costs, but as he discusses, the standard approach is to just consider selling
costs as part of production cost as a simplifying assumption.

\textsuperscript{11} More specifically, the self-scanners would be considered a process innovation and the bakery a product innovation. As
Kamien and Schwartz (1982, p. 2) state, “likewise we shall not distinguish between process innovations and product
innovations. Process innovations are technical advances that reduce the cost of producing existing products, whereas product
innovations involve development of new or improved products. Equivalently, the former may be defined as upward shifts in
Formally, retail firms produce joint products: the explicit good and the implicit distributional services (Betancourt and Gautschi 1988; Betancourt 2004). Using a specification similar to that given by Ehrlich and Fisher (1982) and Pashigian and Bowen (1994), the household technology parameters can be made functions of three types of variables: (i) firm cost shifting services $f_{ijk}$ that may be individual, location, and good specific, ranging from something as simple as delivery services to personalized ads, (ii) public good services $g$, such as a public transportation to get to a grocery store, and (iii) individual, location, good specific capital $h_{ijk}$, which may be physical capital, such as a car, but could also be human capital, such as education level or route knowledge to a store. Thus, using function notation we would have:

$$a_{ijk} = a_{ijk}(f_{ijk}^a, g^a, h_{ijk}^a)$$

$$b_{ijk} = b_{ijk}(f_{ijk}^b, g^b, h_{ijk}^b)$$

$$c_{ijk} = c_{ijk}(f_{ijk}^c, g^c, h_{ijk}^c)$$

The superscript letter signifies possible different targeted variables for each of the inputs (good, time, and effort). The parenthetical negative sign indicates that increases in these variables would decrease the individual consumer’s technology parameters, which would in turn reduce the full price via equation (8) and cause an increase in demand (shift out in the demand curve).

Given that firms are construed as producing both the market good $Q_{jk}$ and distribution services $f_{ijk}$, the firm’s cost functions and thus supply curves have to reflect this multiproduct nature. Recall in the standard single good setting, the firm’s supply curve is its marginal cost curve above the minimum of the average variable cost curve (short run) and the marginal cost depends on the quantity produced (movements along marginal cost) and input prices (shifts in the marginal cost). The multiproduct extension is straightforward, but there are multiple ways to write the multiproduct supply function that are theoretically consistent (Beat the and Taylor 1985, chapter 5). In the present context, the most transparent approach is to use a conditional supply function. In a multiproduct setting that allows for scale and scope economies, a conditional supply function will express the quantity supplied of one good as a function of its output price, the price of inputs used in its production, the quantity of the other goods produced (distribution services), and indicators of operation scale and scope.

The scale, scope, and cost shifting effects on the firm’s supply curve will depend on how scale, scope, and production of the distribution services affects the marginal cost per unit of the market good sold. Increases in scale and scope would be expected to decrease marginal cost and thus shift the marginal cost curve out as these are technological push factors. Alternatively, increases in distribution service could be marginal cost increasing (e.g., hiring more service laborers), neutral (e.g., adopting a technology that only affects average cost, such as a Wi-Fi connection), or decreasing (e.g., creating more self-checkout lanes, decreasing labor cost) as these could be either technological push or demand pull factors.

### 4.3 Supply Function and Curve

Similar to the market level demand, the market level supply for good one in location one is an aggregation of individual supplies, so dropping the $i$ subscript denotes market-level variables. In terms of right hand side variables, supply will obviously be a function of the market price of good one in location...
one \((P_{11})\), a scale and scope indicator for firms in location one \((L_1, C_1)\), the distribution factors for good one in location one \((f^{a}_{11}, f^{b}_{11}, f^{c}_{11})\), and other factors, such as input prices, number of producers, and seasonal factors, all subsumed in the vector \(O^S\). The market level supply function for good one in location one then becomes:

\[
Q_{11}^S = Q_{11}(P_{11}, L_1, C_1, f^{a}_{11}, f^{b}_{11}, f^{c}_{11}, O^S): \text{Market One Supply Function (14)}
\]

The parenthetical question marks as before indicate that the direction of the relationship between the variable and the quantity supplied could be zero, negative, or positive, depending on the type of distributional service provided or other variable. In terms of the market supply curve, a change in the price \(P_{11}\) is captured by a movement along the supply curve and a change in any other variable \((L_1, C_1, f^{a}_{11}, f^{b}_{11}, f^{c}_{11}, O^S)\) will cause a shift in the supply curve with the direction of the shift being determined by the sign under the variable.

5 Graphical Equilibrium Analysis of Some Topical Questions

With both the demand and the supply sides developed, they can be brought together to analyze a few of the questions posed at the beginning of the paper in the typical fashion found in any microeconomics textbook. Before proceeding, the student should be reminded of a few caveats about graphical supply and demand analysis to head off some typical questions. First, there is an art in applying models, and the question dictates the choice. In most undergraduate texts, the questions are often rather general, and therefore, the good is usually some aggregate (e.g., food, food away from home, pizza), and the geographical or temporal aspects of the market may or may not be defined. Second, supply and demand diagrams are qualitative tools, not quantitative tools. They only provide directional insights, not magnitude insights. The size of the shift of a curve could be small or large, depending on the amount the underlying variable changes and the sensitivity of the market to the change. Third, the slopes of the curves could be very flat (very elastic) or very steep (very inelastic), depending on the good, and thus the magnitude of the changes in price and quantity will also depend on these slopes (elasticities). Fourth, in its simplest form, the above analysis relates to demand and supply for a final good (at the consumer level), though the concept of utility is broad enough to include profit, and it could be adapted to any level in the supply chain. Finally, as always, the ceteris paribus clause applies, meaning unless stated otherwise, we are conceptually holding all other factors constant, but in reality, multiple factors are usually changing, and the demand and supply curve will shift accordingly.

5.1 Cost Shifting Affecting Demand Only . . . Perhaps

How do device-assisted transactions affect demand and supply? On the demand side, device-assisted transactions lower consumer’s search costs, time costs, and cognitive load such that the technology parameters \(b\) and \(c\) will be lower or decrease. GPS location–activated searches and sale or price reduction notifications reduce search costs. Voice-activated devices shave off seconds in a myriad of ways, ranging from voice-activated text typing, to thermostat or light settings, to ordering products online. In restaurant markets, apps, such as Trip Advisor, allows one to locate restaurants and see their ranking based on customer reviews, along with pricing information. In the grocery market, there is a dizzying array of apps designed to reduce the cost of all aspects of meal planning, nutrition assessment, and time in the grocery store. These apps integrate several meal production activities in one app, allowing you to search for and save recipes, create shopping lists from recipes, get weekly sales notifications, check for coupons or discounts, comparison item shop, simply scan a bar code of existing items you need to purchase to add to a shopping list, and get an in-store navigation map (Klecker 2019). Thus, on the demand side, device-assisted transactions lower the full price and lead to a higher derived demand for the good.
On the supply side, apps are especially appealing to firms because they exploit a technology standardization, like a public good, that can be leveraged at a relatively low cost to provide greater service (Pantano and Viassone 2014). Customers own and pay for the information delivery device and its operation (the phone and data), they are already familiar with how to use the device, and firms need only provide the resources needed to develop and maintain the app. Depending on the app and its maintenance, this cost may affect marginal cost, but for initial simplicity it is assumed to only affect average cost, not marginal cost, and thus leaves the (short run) supply curve unaffected. Figure 1 then shows a graph of the market for the good (e.g., groceries) where there is a demand curve where the device-assisted transactions are not available (D⁰) and then a higher demand (D¹) where these services are available and, ceteris paribus, more of the good would be sold and for a higher price, but again magnitudes will depend on slope and shift magnitudes. Of course, if these costs of providing these services affected marginal cost, then the supply curve with these services would be higher and the price effect higher and the quantity effect attenuated.

![Figure 1. Supply Neutral Cost Shifting Increasing Demand](image)

**Application: Transaction-Assisted Devices.** Cost shifting that does not affect the short-run supply curve, such as transaction-assisting devices, leading to higher demand (D¹ > D⁰) because of lower time and effort costs (b₁ < b₀, c₁ < c₀), and thus lower full price.

How are online, offline grocery shopping, meal kits, and the emergence of grocerants all connected? “The U.S. e-grocery market had a share of 3 percent of total sales in 2016. The market share was expected to grow to 10 percent by 2020” (Statista 2019). However, a recently published article in *The Atlantic* has the title, “Why People Still Don’t Buy Groceries Online” (Semuels 2019). The title implies it is a demand side problem. It is actually a supply side problem. Four key product attributes come into play in analyzing the economics of online versus offline purchases: (i) the *quality heterogeneity* of individual products
within the order, (ii) the perishability of the products within the order, (iii) the number of products in the order, and (iv) the packing arrangement of the order. Orders where these attributes are not that important are better suited for online purchases and delivery than products where these attributes are important. For example, electronics, household goods, and even clothes are mass-produced with very uniform quality, are not perishable, and can be combined in a single order with little concern for packing arrangement. On the other hand, there is a great deal of possible heterogeneity for a grocery item (e.g., green vs. ripe bananas), many grocery items are highly perishable (e.g., ice cream), the number of possible product combinations from a grocery store can quickly approach infinity, and the packing arrangement of groceries is very important (e.g., bread needs to be placed in the top of the bag). Thus, the more important these attributes are, the costlier it will be to process orders and deliver the orders. Stated succinctly in economic terms, the more important these attributes, the more labor intensive and less capital intensive is the online business model and economies of scale are difficult to achieve, especially in delivery. The basic economics of transportation costs indicate it is much cheaper per trip if a large truck can be sent to a densely populated area to deliver groceries than sending many smaller trucks to widely dispersed customers. What one would expect to see is that online grocery shopping and delivery would be more prevalent and potentially more profitable in densely populated areas, and this is indeed the case. For example, the Amazon Fresh delivery service announced last year it was suspending service in some areas while still providing services in cities such as New York, Chicago, and Boston (Semuels 2019). An intermediate business model that is more cost effective is to scrap the delivery service in markets where that cost is high, but still do a “click and collect” where the customer can shop online and then go to the store and pick up the order that was filled by store employees.

The economics of meal kits are similar. Just a few years ago meal kits were the new rage in the food sector (e.g., Blue Apron, Home Chef, Plated); not anymore. “Few business models are as unprofitable as those of meal-kit companies” (Ladd 2018). Why? On the supply side, the cost of meal kit delivery faces all the logistical hurdles of delivering grocery orders mentioned above, but with the added labor (and capital!) costs associated with designing meals, purchasing ingredients, preparing ingredients, packing meal kits, and marketing their brand. Thus, it is an even more labor-intensive service and thus would demand a much higher price to cover this extra cost. However, on the demand side, there are also economies of scope for the consumer associated with an offline store, meaning the full average price per item purchased can be lower when all items can be purchased in one sitting or location (i.e., one-stop shopping), without having to switch gears, perhaps figuratively and literally, from shopping in one place to go to another to get different products. And furthermore, there is effectively no barrier to entry preventing grocery stores from offering meal kits. Given the economies of scope advantage of grocery stores over meal kit companies, it makes sense to make the meal kit just another category line in the grocery store at a lower price point, and indeed, this is what has happened as many meal-kit businesses have now either signed agreements or been bought by traditional retail outlets (e.g., Albertsons and Plated, Kroger and Home Chef) such that meal kits are now sold in grocery stores.

The economies of scope of cost shifting within the grocery store also help explain the emergence of the sit-down restaurant in the grocery store or the “grocerant” (Meyer 2017). Grocery stores already have in the store ease of access to many of the inputs needed for operating a restaurant area, and thus the additional costs are not that high. Consequently, the restaurant can be thought of as the addition of another category line in the convenience spectrum from basic ingredients, to semi-prepared foods, to meal kits, to ready-to-eat take out, to sit down meals. In fact, one could predict that grocery stores will continue to look for economic opportunities all along the convenience spectrum, perhaps even going in the opposite direction by growing food within the store and letting the individual “pick their own,” as is being explored by Kroger (Browne 2019). Thus, by offering products along the entire convenience spectrum, they are also able to attract all the consumers along this spectrum as well.

Figure 2 demonstrates all of these market observations. The 0 superscript could denote the online grocery market conditions with a higher per unit cost and price \( P^0 \). The 1 superscript could denote the offline grocery market with lower per unit cost and price \( P^1 \) because of scale and scope economies and a
higher demand resulting mainly from consumer economies of scope (one-stop shopping). The 2 superscript could denote the offline grocerant market, which relative to a grocery market would have higher labor cost (ceteris paribus) because of diseconomies of scale in consumption relative to offline markets ($D^2 > D^0$). Grocerant markets are expected to have higher costs than regular offline grocery because of providing more services ($S^2 < S^1$), but this cost shifting would be expected to lead to higher demand ($D^2 > D^1$). Thus, we would expect to see higher prices and lower quantities for online grocery markets relative to offline grocery markets ($P^1_0 > P^1_1, Q^1_1 > Q^1_0$) and higher prices and lower quantities for grocerant markets relative to offline grocery markets ($P^1_1 > P^1_2, Q^1_2 > Q^1_1$), ceteris paribus. Note the general qualitative conclusion does not change if it is believed online has a higher demand than offline (i.e., switch the two demands).

What does convenience have to do with food deserts? Over the last decade, there has been much concern about food deserts, which is defined generally as an area devoid of a supermarket (Walker, Keane, and Burke 2010). Analyses have focused on demand side characteristics of households and compared them between food desert and nonfood desert areas. The most important finding is that food deserts are located in low-income areas or stated conversely, nonfood deserts are located in higher income areas. In addition, one of the supposed puzzles is that food deserts tend to have higher prices than nonfood deserts (Powell et al. 2007; Dutko, Ver Ploeg, and Farrigan 2012). This is only a puzzle if one ignores the economics of the supply side and specifically the store location decision. The economics—demand and supply—of convenience helps explain this observation.

Applications: Online vs. Offline Groceries, Grocerants. Online grocery markets with delivery are denoted by 0, offline grocery markets are denoted by 1, and grocerant markets are denoted by 2. Online grocery markets with delivery are expected to have higher costs than offline grocery markets such that $S^0 < S^1$ and perhaps lower demand because of diseconomies of scale in consumption relative to offline markets ($D^1 > D^0$). Grocerant markets are expected to have higher costs than regular offline grocery because of providing more services ($S^2 < S^1$), but this cost shifting would be expected to lead to higher demand ($D^2 > D^1$). Thus, we would expect to see higher prices and lower quantities for online grocery markets relative to offline grocery markets ($P^0_1 > P^0_1, Q^0_1 > Q^0_1$) and higher prices and lower quantities for grocerant markets relative to offline grocery markets ($P^1_1 > P^1_2, Q^1_2 > Q^1_1$), ceteris paribus. Note the general qualitative conclusion does not change if it is believed online has a higher demand than offline (i.e., switch the two demands).
Profit margins in the grocery industry are some of the smallest in any industry (1–3 percent; Forbes 2016), and this is again where economies of scope, scale, and cost shifting intersect as there are built-in incentives to increase the size of grocery stores and services (Ellickson 2016). On the revenue side, though profit margins on a whole for grocery stores are small, products that contain more cost shifting have higher margins, such as ready-to-cook or ready-to-eat items found in the deli or bakery sections (Johnson 2019). The individuals who are willing to pay higher prices for these service-embedded goods are going to be those with higher incomes and higher time costs. Thus grocery firms have incentives to locate where incomes are higher and the opportunity cost of time is higher. The theory then suggests we would expect to see firms locating stores in higher income areas and providing a continuum of services that reduce the full price, ranging from more products in a single location to a wider distribution of the types of products (basic ingredients to the in-store restaurant). Pashigian, Peltzman, and Sun (2003) provide evidence that grocery stores have responded to higher time cost of households by hiring more in-store labor (providing more services, such as the bakery or deli) and locating in places that are more convenient for individuals with a higher time cost. Thus, this is a case where one needs to remember the ceteris paribus condition in the graphical analysis, because grocery firms have simultaneously been exploiting scale and scope economies, which would shift out the supply curve, but also providing more labor-intensive services, which would tend to shift the supply curve back. This then helps explain how it is possible to simultaneously observe low-income areas facing higher prices and less services.

Figure 3 demonstrates this case where now the 0 superscript refers to the market in food desert areas with prices and quantities \( P_0 \) and \( Q_0 \), respectively, as a result of lower values of the scale, scope, and distribution services variables (i.e., \( L, C, f^b, f^c \)). The 1 superscript denotes the market in the nonfood desert area with lower costs because of scale and scope economies but also higher demand because of more services with corresponding lower prices and higher quantities \( P_1 \) and \( Q_1 \), respectively.\(^{12}\)

### 6 Conclusions and Extensions

Convenience is perhaps the most important “commodity” being sold in the market today, and yet there is nothing of analytical substance to be found in most undergraduate textbooks. This article fills this important gap in a straightforward manner by incorporating convenience in the typical supply and demand framework using the standard tools of introductory and intermediate microeconomics. This was achieved on the demand side by using Becker’s (1965, 1985) household production theory to include time and effort technology and resource constraints leading to full prices that consist of the direct market price plus indirect time and effort prices. On the supply side, retail supply and distribution theory (Betancourt 2004) allowed for an interaction of scale and scope economies and cost shifting services that led to suppliers providing services that not only affect supply but also demand via the direct effect on the indirect time and effort prices, which in turn affects the direct market prices and quantities as well. The framework was used to answer several questions related to convenience that could not be answered with the standard supply and demand framework that does not explicitly account for convenience.

The framework could be employed in analyzing numerous other questions as well. For example, why are advertisers willing to pay 2.7 times more for behaviorally targeted ads than nontargeted ads, as suggested by one study (Beales 2010)? Or why, according to a report in Forbes, do “70 percent of advertisers currently work with influencers, and 40 percent plan to increase influencer budgets in the

\(^{12}\)This graph is consistent with what has been found empirically, but it implies that the outward supply shifts are greater than the outward demand shifts. This situation does not have to be the case and would vary by market and good. This is just a case of the more general principle of demand and supply: if supply and demand shift in the same direction, we can only be certain about the direction of the quantity change. Price may increase or decrease depending on the magnitude of the shifts. Alternatively, if supply and demand shift in opposite directions, then we can only be certain about the direction of the price change.
coming months” (Davis 2019)? What is the economic commonality in the development of a faster charging or longer lasting battery for a handheld device or an electric car?

Perfect competition was assumed because that is the entry point for undergraduates being taught market equilibrium analysis for the first time. As alluded to, there certainly may be applications where an imperfectly competitive market model would be more appropriate. The analytics for imperfect competition extensions are rather straightforward (e.g., monopoly, duopoly, monopolistic competition). The key is to capture all the main components within the imperfectly competitive model. On the revenue (demand) side, the key is to work with a derived demand function for the good expressed in terms of full prices, not just market prices. The full prices are functions of the household technology parameters, which in turn would be functions of the firm’s distribution services. On the cost (supply) side, the key is to have a cost function that contains scale, scope, and distribution service components. Thus, the firm (or firms) then chooses not only price (or quantity) of the market good but also levels of distribution services and perhaps even scale and scope variables as well. This quickly could get complicated if one wants to

**Figure 3. Economies of Scale, Scope, and Cost Shifting Affecting Supply and Demand**

**Applications: Food Deserts vs. Nonfood Deserts.** Food desert grocery markets are denoted by 0 and nonfood desert grocery markets are denoted by 1. Grocery stores in nonfood deserts are expected to be larger, provide a larger variety of products, and provide more services, thus benefit from economies of scale and scope, even with more cost-shifting services, such that $S^0 < S^1$. Nonfood deserts are expected to have higher demand because of more services reducing time and effort cost, but also higher income, so $D^1 > D^0$. Thus, if the economies of scale and scope outweigh the cost-shifting effects, then food deserts will have higher prices and lower quantities than nonfood deserts ($P^0_1 > P^1_1, Q^1_1 > Q^0_1$).
consider response functions of other firms. However, none of this would tend to change the fundamental equi-marginal intuition, which is simply that firms are willing to bear some of the burden of shifting time and effort costs from consumers to firms (cost shifting) if the marginal revenue exceeds the marginal cost, and this will have implications for the direct prices and quantities of goods sold in the market.
References


Browne, M. 2019. ”Kroger to Offer In-Store Living Produce Farms.” Supermarket News, November 19.


Pashigian, B., and B. Bowen. 1994. “The Rising Cost of Time of Females, the Growth of National Brands and the Supply of Retail...


