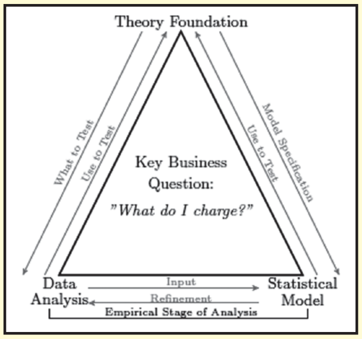
Pricing Analytics: Models and Advanced Quantitative Techniques for Product Pricing



This article is the first chapter from the author's latest book, *Pricing Analytics: Models and Advanced Quantitative Techniques for Product Pricing*. The author has made this preview exclusively available to PPS members, and the full book is available at https://bit.ly/2NiqqxO. This chapter explores the challenges businesses face when determining which pricing strategies will be most effective in their markets, as well as methods for conducting pricing research and understanding price elasticity in relation to their customers. Walter R. Paczkowski, Ph.D. is Chief Data Scientist at Data Analytics Corp. He will be leading a one-day workshop - Deep Data Analytics for Pricing: Uses, Issues, and Solutions – on October 24th at the PPS Fall Conference in Dallas. He can be reached at walt@dataanalyticscorp.com.

Each day, business managers answer a number of questions: legal, personnel, regulatory, and the list goes on. There is one question they must answer – the *Key Business Question* – that ultimately drives the business: "*What do I charge*?" They need a number to put on the tag that goes on their product so a consumer can decide whether or not to buy the product. That price, that number, has to be the "best" one to capture the most consumers and make the most sales. Otherwise, the business manager will be out of a job and the business may no longer exist. A lot rests on that price, that number.

What to charge, how to price the product, is the most difficult of the business questions to handle because it is often unclear how much people are willing and able to pay for a product. This holds true for existing products as well as new products. For the former, the current market price may not be the best one. It may be too low so money is "left on the table" (that is, higher revenue could be earned without changing the cost structure for the product) or it may be too high so that sales are needlessly lost. For the case of a new product, there is often a lack of information regarding where to begin setting the price and so new product pricing offers unique challenges.

Marketing professionals base their marketing plans on the "*P*s" of marketing, a clever mnemonic for the key tools available for marketing.[1] The *P*s are:

- ***Product:*** The item to be sold as defined by its characteristics and features; that is, its attributes.
- ***Place:*** Where the product is to be sold.
- ***Position:*** The product space location where the product will lie relative to competitor products (e.g., as a premium or a bargain product).
- ***Promotion:*** How potential consumers of the product will learn about it (i.e., how they will gain the information needed to make a purchase decision).
- ***Price:*** How much consumers will be asked to pay for the product.

Pricing is different from the other *P*s in a number of respects. The price, or a change in the price, is the only marketing factor that goes directly to the bottom-line of the income statement. Change the price and the revenue side of the income statement changes immediately. Also, the price can be changed almost instantly. Within a matter of minutes, a store can raise or lower its price merely by posting a sign or, in our increasingly technological world, by changing a statement in a computer program that controls an electronic price display such as the ones gas stations now use.

The other marketing *P*s also affect the bottom-line, but not so directly and immediately. They require time to implement and "spread the word." Take promotion, for example. Once it is decided that a new promotional campaign is needed, perhaps a new in-store display or a flyer mailed to consumers, then a lengthy process is set in motion to develop and deploy the new promotion. The development is not cost free. Creative designers and writers are needed who are usually high-priced professionals. In addition, the optimal mix of magazines, newspapers, TV and radio spots, and in-store displays has to be chosen, which also requires time and expensive market research. So, in addition to potentially improving the bottom-line, a promotion may also hurt it by increasing the cost side of the accounts. Any increase in net income from increased sales due to the promotional campaign could be partially, if not completely, offset by the costs of the campaign. A price change, on the other hand, only affects the revenue side.[2]

This chapter is divided into eight sections. This first outlines issues associated with answering the *Key Business Question* by using one of two classes of pricing strategies. Section 2 introduces the importance of a price effect measure, basically the price elasticity. Section 3 discusses approaches to pricing research, while the following two sections, 4 and 5, discuss each in depth. The use of a simulator to study the results of quantitative research is mentioned in Section 6, while Section 7 highlights the role of elasticities. Section 8 is a summary (not included here).

CONTINUED ON NEXT PAGE ➔

1.1 Answering the Key Business Question

Business managers typically begin addressing the *Key Business Question* by focusing on a pricing strategy. A strategy is a statement of the key actions that will be taken by the business. Consider the following example of a pricing strategy statement for a fashion product: "It is optimal to price high at the beginning of a fashion season to attract price-insensitive, fashion-forward consumers and then lower prices over time to price discriminate to reach the more price-sensitive customers."[3] This statement clearly articulates how prices will be changed, and to whom the specific prices will be targeted. What is lacking is the specific price point – the number to be charged and changed.

1.1.1 Uniform pricing strategy

A uniform pricing strategy is one that uses a simple, single price for all units sold to all buyers. This is an almost naive pricing strategy because it assumes that all consumers are homogeneous and there is only one product without variations due to quality. Since the price applies to all units sold, no consideration is given to order size (i.e., quantity discounts) or order composition (i.e., bundling). Prices are then linear rather than nonlinear.

1.1.2 Price discrimination strategy

A price discrimination strategy as in the above strategy statement is basically a strategy to sell products (identical or not) to different people at different prices. There are three basic forms, or *degrees*, of price discrimination:

> *First-degree*: Price varies by how much each consumer is willing and able to pay for the product.
>
> *Second-degree*: Price varies with the amount purchased. This is a form of non-linear pricing in which the pricing schedule is a nonlinear function of the quantity. This is typically used by electric utilities, but it also applies to quantity discounts, bundling, and product-line pricing.
>
> *Third-degree*: Price varies by consumer segments. Segments may be defined by age, gender, location, and status (e.g., students vs. non-students, senior citizens vs. non-seniors). This is a common form of price discrimination.

The fashion example is a *third-degree price discrimination* form because consumers are divided into two segments: fashion-forward and non-fashion-forward. The former must have the latest fashion trends as soon as they become available to be ahead of the fashion curve, while the latter wait to see what everyone else is wearing and what is in fashion before they buy. The former would be more inelastic, while the latter would be more elastic.

Therefore, the fashion-forward buyers would be charged a high price, as the strategy statement says, and the non-fashion-forward buyers would be charged a lower price.

Whatever the form, price discrimination as a pricing strategy is widely used. I will return to the theoretical issues of determining the price for a price discrimination strategy in Chapter 3.

1.1.3 Strategy parts

The fashion-forward example illustrates that a pricing strategy has two parts:

> *Price structure*: A high price at the beginning of the fashion season and a low price at the end to target consumers based on their price sensitivities or *elasticities*.
>
> *Price level*: The price that actually will be charged (which is not mentioned in the fashion strategy statement).

A pricing strategy statement simultaneously expresses both structure and level as is clearly done in the fashion example. The *Key Business Question*, however, is about the second part of a strategy statement: the *number* for the level. What is the number and where does it come from?

In this book, I will focus on the *level* because that is the number that goes on the tag that is on the product that must be sold. Chapter 6, however, will deviate from this by discussing ways to gain insight into structure.

1.2 Price effect

Just as important as the price level is the price *effect*. All too often, managers just think of "stimulating demand" by lowering the price, as in the fashion-forward example. But this is too narrow and simplistic a focus. Thinking more broadly, what effect will a different price level have on a key business metric such as:

- Revenue?
- Contribution?
- Customer acquisition?
- Customer retention?
- Market share?

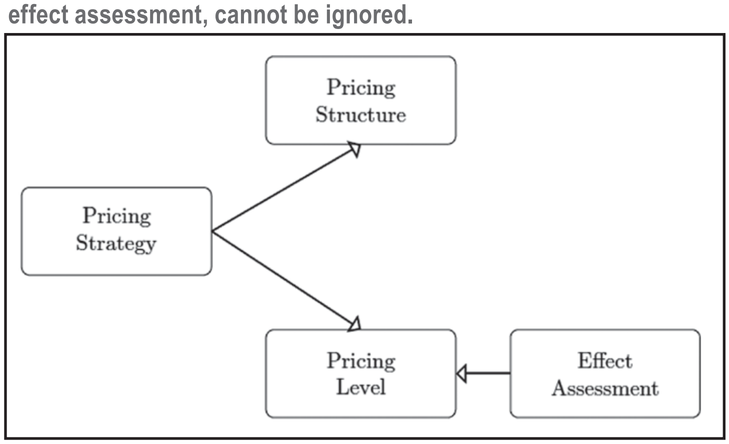
If a price decrease results in a decrease in revenue, then is the new price level the correct one to use? No.

The level part of the answer to the *Key Business Question*, therefore, is subdivided into two parts: a level – or number – answer and an effect assessment answer. This is illustrated in Figure 1. Both level and effect are determined simultaneously by *pricing analytics*.

The remainder of this chapter is divided into six sections. The first section to follow, Section 1.3, discusses two approaches – qualitative and quantitative – a pricing manager could use to set

CONTINUED ON NEXT PAGE →

Figure 1: The basic pricing framework most marketing professionals use has two components: strategy and level. Most of the emphasis by marketers is on the strategy part. But the level part, along with the effect assessment, cannot be ignored.



the price level. This section sets the stage for the next two sections that discuss the qualitative and quantitative approaches in more detail. The remainder of the book, of course, focuses on the quantitative.

Sections 1.4 and 1.5 discuss in more detail the two approaches and argue for the use of a quantitative approach to pricing. Section 1.5 also introduces a framework for doing *pricing analytics*. This framework will be followed throughout the book to answer the *Key Business Question*.

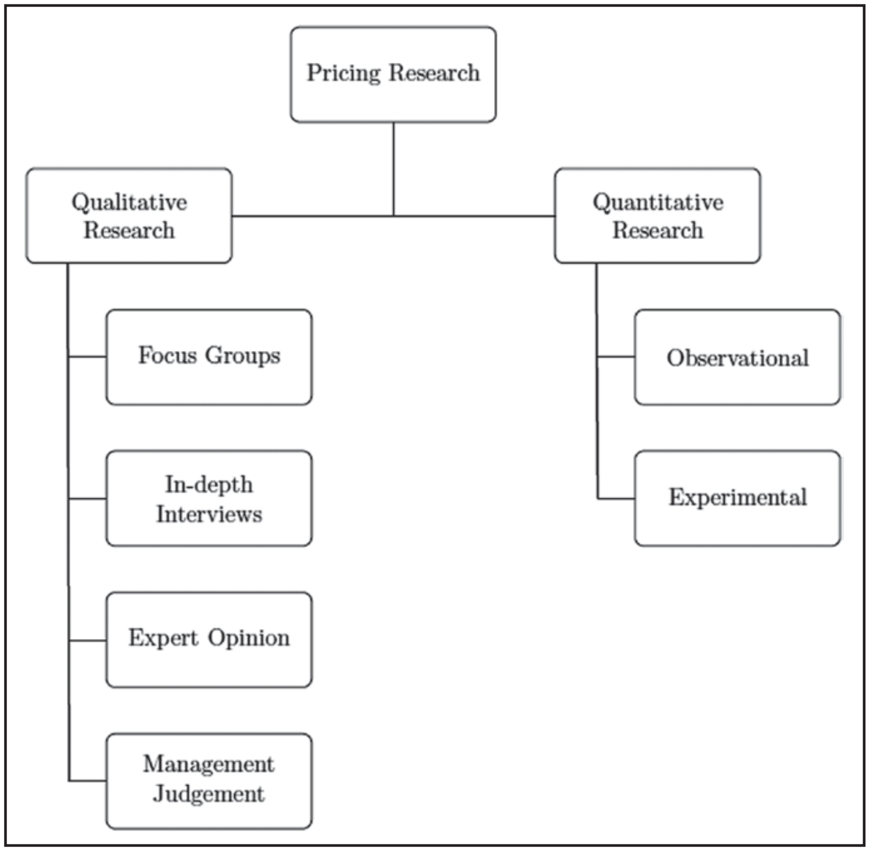
Sections 1.6 and 1.7 introduce simulators as a tool for pricing analytics and the economist's notion of a price elasticity. This is a very brief introduction as befitting an introductory chapter. This important concept is more fully discussed and applied in Chapters 2 and 3. The final section, Section 1.8, is a summary.

1.3 Pricing research approaches

In this section, I will discuss two broad approaches for pricing research: qualitative and quantitative research (see Figure 2).

Qualitative research is largely opinion based and unscientific. Sampling methodologies and statistical tools are not used to determine the price level or even to assess effect – they cannot because qualitative research does not rely on data per se. This is not meant to disparage or condemn qualitative research, because it does have a valid role to play in determining a price level and assessing effect, but that role is not a sole or primary one. Qualitative research is discussed in Section 1.4.

Figure 2: There are two major approaches analysts have used for pricing: qualitative and quantitative. This book is concerned with the quantitative approach using observational and experimental data. We can observe consumers' purchase behaviors that reveal their preferences and build econometric models to estimate elasticities, but you can also ask them to state their preferences using an experimental design and then build econometric models to estimate elasticities.



Quantitative research consists of data analysis and statistical modeling using either observational or experimental data and founded on the basic economic paradigm of consumer demand. The data analysis, aside from what is fundamentally used in model building, includes methods to find data anomalies, correlations, patterns, and trends. Statistical models based on observational data – data measuring actual consumer purchase behavior that reveal preferences (and, hence, are called *revealed preference data*) – are usually regression models of varying complexity.[4] Statistical models based on experimental data – data collected via an experimental design in which consumers state their preferences (and, hence, are called *stated preference data*) – are either conjoint or discrete choice models. Quantitative research is discussed in Section 1.5.

Going forward, observational data will be referred to as "revealed preference data" and experimental data as "stated preference data."

Qualitative and quantitative research can be done to support one another at different stages of the overall research process. Qualitative research done before the quantitative work can provide information and insight into factors other than price that drive or determine the demand for a product, which can then be used in the quantitative research. It can also be done after the quantitative research to give support or credibility to the quantitative results by using, for example, expert assessments. Likewise, quantitative research can be used to give credibility to qualitative research results or to support or bolster expert or managerial opinion or judgement.

Although quantitative pricing research methodologies based on revealed or stated data will be discussed in this book, something should be said about qualitative research. This is done in the next section.

1.4 Qualitative pricing research

Qualitative research consists of focus groups, in-depth interviews, expert opinions, and managerial judgements, to mention a few possibilities.

Focus groups are popular for learning about consumer preferences. Led by a trained and experienced moderator, a well-thought-out and executed series of focus groups coupled with a detailed moderator guide can yield useful information about consumer preferences and buying intentions. This extends to pricing by asking consumers in the groups how much they would be willing to pay. This information, however, can only yield approximate prices, or, better yet, price ranges because the consumers in the focus groups, even though they are chosen based on specific demographic and buying behavior criteria, are not a scientifically selected random sample. Insufficient sample

CONTINUED ON NEXT PAGE →

size is always an issue because focus groups tend to be small. What the focus group participants say cannot be generalized to the market, but can only be used as guides and input into further scientific and quantitative work – the quantitative branch of the pricing tree in Figure 2.

A specialization of the focus group is the in-depth personal interview in which a subject matter expert (SME) or key opinion leader (KOL) is questioned in detail about some issue. The SMEs may talk to the sales force, which has a direct link to retailers and customers, and the SMEs inform the sales force about what the price should be. The SMEs may also talk to consultants who know the market and understand dynamic market operations for the product's category. The KOLs write blogs, lengthy industry reports, and columns in various trade press publications prognosticating on the future of their industry or products. Although the information garnered is useful, it is still biased by the few interviews that can be conducted and who is interviewed.

Management judgement plays a large role in retail pricing.

Notice the focus on guesswork.

> A large department store wants to sell a "one-of-a-kind" designer gown. Although the manager has some idea about the price that the gown can command, there is generally some guesswork associated with the process. How should he choose his initial price?

Management, or the management team, uses information from SMEs and KOLs, plus their own instincts, to determine the best price or how prices should be changed in an attempt to remove the guesswork. They do know something independent of what statistics tells them. This information is called *prior information*, or just a *prior*. The information, however, is personal to them, having been built through years of experience with similar products and situations. They may have developed a "sixth sense" or "gut feel" or a set of rules of thumb (ROTs) that guide their decision. In seasonal retailing, for instance, managers often set prices high at the beginning of the season and then lower them dramatically later in the season just to sell or move merchandise because experience taught them about seasonal effects.

The use of prior information is the basis of Bayesian data analysis. This is a complex, highly technical area of statistics that is now used more frequently because of great strides in software and algorithmic procedures, especially simulations with Markov Chain Monte Carlo (MCMC) methods. Bayesian issues will be discussed in Chapter 5 for the design of discrete choice studies.

All the information gathered from the qualitative side of the pricing tree of Figure 2, whether from the judgement or ROTs of managers, SMEs, KOLs, or consumers in focus groups, has a place in pricing. In the early stage of new product pricing, before any quantitative work is or can be done, a rough approximation of prices may be needed for developing scenarios for a business case to prove or support the continuation of the development of the new product. If the product is in the conceptual development stage, there may not be enough product details to specify a quantitative study, yet the effects of price are needed to decide whether or not it is profitable to continue product development. This is illustrated in the next section.

1.4.1 Pharmaceuticals case study

Pharmaceutical new product development provides an excellent case study of pricing research issues and techniques. As background, new pharmaceutical products must pass a series of rigorous tests or checkpoints that collectively span several years before product launch. The tests are divided into phases[5]:

Phase I: Dose and administration testing; this is preceded by pre-clinical development and testing.

Phase II: Efficacy testing with a test and control group.

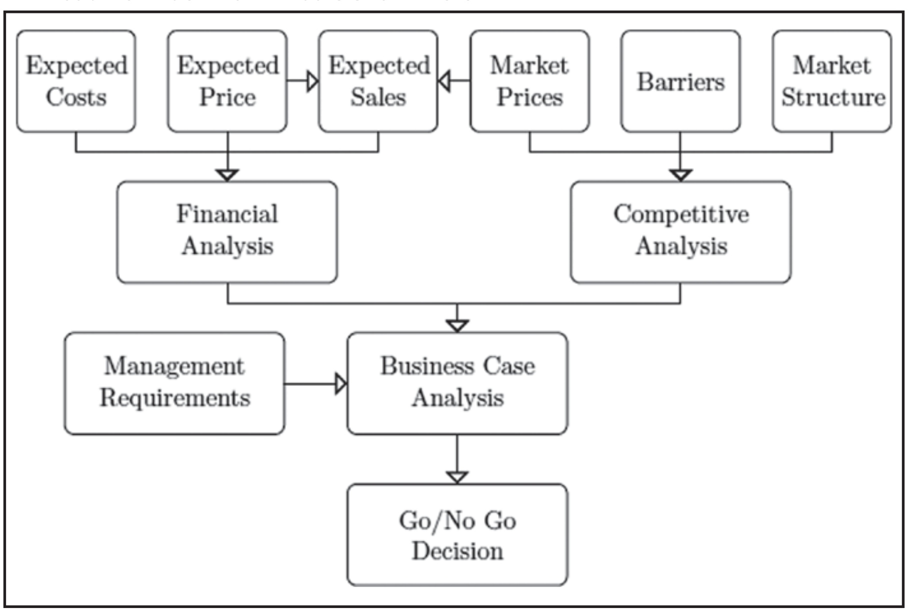
Phase III: Overall effectiveness testing in a large population.

Phase IV (optional): Tracking after launch.

Each phase can last several years. The total time for Phases I–III could be a decade, but the actual time depends on many factors.[6] Given the time span of these phases plus the time required to develop the new drug before the phase testing even begins, it should be clear that new drug development is a costly endeavor.

At each stage of this process, not only is the new drug tested, but its financial viability is also closely monitored through the *business case* process. A business case is a competitive market and financial assessment of the viability of a product. This process is illustrated in Figure 3.

Figure 3: This schematic of the business case process is, of course, highly stylized. The actual process will vary across companies, but the key components and linkages will not differ much from what is shown here.



The competitive assessment part of a business

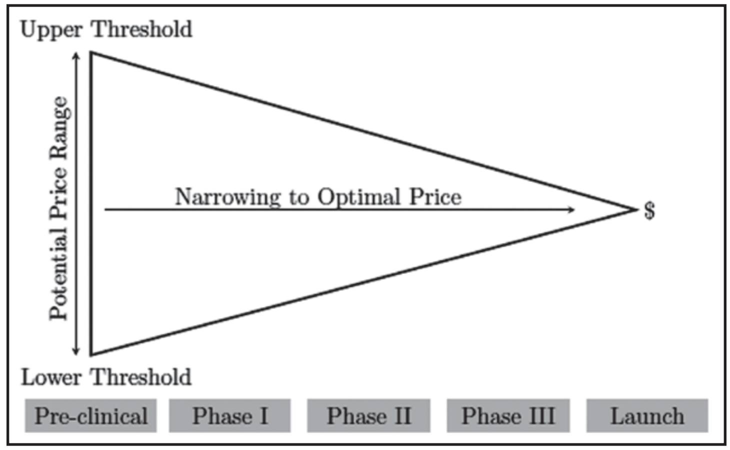
CONTINUED ON NEXT PAGE →

case considers the structure of the market (competitive, fragmented), entry barriers, regulatory restrictions, perceived attribute similarities (from the consumers' viewpoint), likely competitive reactions to marketing programs, and, of course, price structure and levels.

The financial assessment part of a business case considers unit sales forecasts, expected average unit costs, and proposed price points to calculate an expected rate of return for the product. The goal is to determine if the product's estimated rate of return under these conditions would meet or exceed a management-mandated rate of return, or "hurdle rate."

Determining the optimal selling price a decade before the new product is launched, let alone developing it a year before launch, is an impossible task for obvious reasons. Nonetheless, a price point for the business case is needed. In the early business cases, it is best to use a price range of "most likely" prices, with a specific price point developed closer to actual launch. This is where qualitative research can play a big role. KOLs can also be relied on in this phase. This should then be followed by quantitative research as the launch date comes closer, especially stated preference research, with a price range replaced by a price point by the day of launch. This is illustrated in Figure 4.

Figure 4: Due to the long time frame for new product development in the pharmaceutical industry, different levels of pricing are needed and are possible during the development process. Wide ranges are the best that can be developed in early stages. A price point, however, is needed at launch. The quantitative methodologies are different at the different stages. Adapted from [21].



1.4.2 Cost-plus pricing

I would be remiss if I did not say something about cost-plus pricing. Cost-plus pricing is probably the most common pricing approach, typically among those companies that are "less sophisticated." Cost-plus pricing

> … is used primarily because it is easy to calculate and requires little information. There are several varieties, but the common thread in all of them is that one first calculates the cost of the product, then includes an additional amount to represent profit. It is a way for companies to calculate how much profit they will make. Cost-plus pricing is often used on government contracts, and has been criticized as promoting wasteful expenditures.
>
> The method determines the price of a product or service that uses direct costs, indirect costs, and fixed costs whether related to the production and sale of the product or service or not. These costs are converted to per unit costs for the product and then a predetermined percentage of these costs is added to provide a profit margin.[7]

I advocate for the use of a strong quantitative approach focused on what the customer is willing and able to pay for a product or service.

In essence, this approach to pricing involves finding the average cost of producing the product and then adding a markup over that cost. Cost-plus pricing has the appearance of a quantitative approach to pricing since costs, a quantitative concept, has a markup, another quantitative concept, applied to it. Adding a markup to costs is a financial issue, which is certainly quantitative. My position is that the markup itself, although a number, is actually based on management judgement about what is acceptable or required for the business to return to its owners as a profit and so falls into the realm of the qualitative approaches. The result of this qualitative action is quantitative, not the input. As such, it is outside the scope of this book.

Cost-plus pricing has many other issues. It has been pointed out that cost-plus pricing can even lead to perverse pricing in that a company could be led to increase prices in a weak market and have prices that are too low in a strong market.

The central issues with cost-plus pricing is that the customers are not considered; only management's judgement about the size of the markup is considered. I advocate for the use of a strong quantitative approach focused on what the customer is willing and able to pay for a product or service. Estimates of demand elasticities are at the heart of this approach. Elasticities and their uses are discussed in detail in Chapters 2 and 3, respectively.

Although my position is that cost-plus pricing per se is strictly qualitative, you can, nonetheless, determine the percentage markup over average incremental cost using elasticities. The elasticities are quantitatively based so you can calculate the percentage markup. This is different, however, from *specifying* the markup, which is what cost-plus pricing per se usually entails.

1.4.3 Importance of qualitative information

Although qualitative pricing research is not emphasized in this book, it has an important role in pricing because it establishes prior information – information prior to the quantitative research – that helps guide the quantitative research. Qualitative research, however, does not and cannot lead to useful numbers. The best qualitative techniques can do is provide opinions.

CONTINUED ON NEXT PAGE ➔

As you will see later, discrete choice approaches to modeling can be greatly enhanced by using prior information (simply called *priors*). A basic recommendation is to use priors if they are available. I will discuss how later in this book.

1.5 Quantitative pricing research

The quantitative approach is more complicated than the qualitative because it involves three interconnected parts:

> *Theory foundation*: Establishes the underlying principles for the research.
>
> *Data analysis*: Supports a statistical model.
>
> *Statistical model*: Estimates elasticities for determining price level and effect.

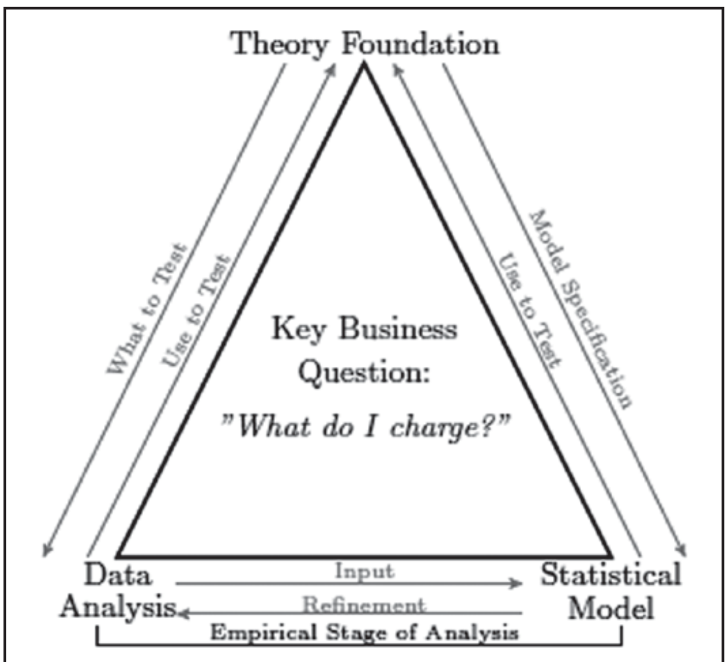
These three parts address the *Key Business Question* ("What do I charge?") and form the basis for estimating price elasticities. We can represent these three parts as a triangle as shown in Figure 5.

The *theory foundation* highlights key variables or drivers of data collection and empirical model specification. For pricing, this is the economic theory of consumer demand. *Data analysis* supports theory development and statistical model specification, with graphs and transformations (e.g., logs) playing major roles.

The *statistical model* estimates parameters consistent with the theory and directs data collection. The *data analysis* and the *statistical model* together form the *empirical stage of analysis*.

At the center of the triangle is the *Key Business Question*, the driving motivation for the quantitative research study. The three vertices of the triangle revolve around, address, and support this *Key Business Question*. The triangle could be rotated in any direction and it would still stand on its own with no part dominating. They all work together to address the *Key Business Question*.

Figure 5: The pricing analytics framework is a stylized depiction of the research process, whether for pricing or any type of quantitative research. Here, the Key Business Question is the focus of the process.



1.5.1 The role of theory

Theory plays a major role in any research because it is an organizing framework. Theories are designed to promote "systematic and organized methods of reasoning"[8] by allowing you to create artificial worlds to test ideas. The role of theory does not stop here. Theory helps you to avoid confusing causation and correlation, which people often think are one and the same – that they can be used interchangeably. Causation is distinct from correlation. Correlation means association, not causation.

Theory also helps you avoid spurious correlations in which a third, unexamined variable may be latent; it may be there and exert an influence, but you may not be able to identify it (yet) or measure it (yet). You may find, for example, a high positive correlation between the amount of ice cream sold at the Jersey shore in mid-July and the number of drownings at the Jersey shore also in mid-July. This is spurious because of the existence of another factor: the hot weather. This third unaccounted variable, the hot weather, is the reason for the supposed high correlation.

Let me pause to handle an issue about "theory" vs. hypotheses. Here is a popular quote from the British economist, John Maynard Keynes (emphasis added):

> The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed, the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual influence, are usually the slaves of some defunct economist.

The important part of this quote is the emphasized sentence on "Practical men." Economists develop theories about how the world works and, one way or another, usually through the infamous Economics 101 college course, those theories are learned, although often not well. Nonetheless, they are learned and applied, whether we know it or not. They influence us as part of our stock of knowledge and for empirical work they are our ultimate guide as to what we should do, whether we acknowledge it or not. So Keynes' economists are always there.

Despite these economic theories still having an influence, some people prefer not to acknowledge them and instead talk about "hypotheses." You could substitute the word "hypothesis" for "theory" and nothing would change. In some sense, all analytical work is hypothesis driven. Analytical work involves the use of data, data often too voluminous to wade through effectively without a guide. The guide for going through the data is the set of hypotheses.

Referring to the research triangle, theories or hypotheses, whichever word is preferred, guide what data you need to collect and tell you how you should analyze them. Here are some example pricing hypotheses:

CONTINUED ON NEXT PAGE →

- Setting our price at $4.50 for each unit will yield sales of 1000 units.
- Increasing price 1% will increase our profit margin.
- Male customers will be willing to pay $100 more for a convertible option.
- Customers over 45 years old are more price sensitive than younger customers.
- There are four price segments in our market, each with a different price elasticity.
- Our competitor's pricing strategy has little effect on our market share.

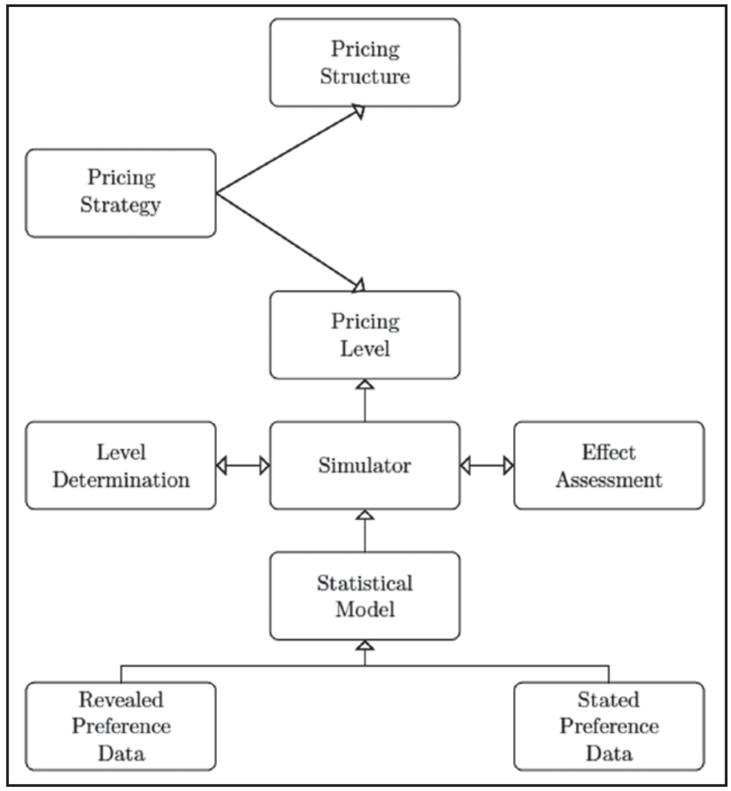
The statistical model should be founded on economic theory. After all, you are concerned with how people make decisions and economic theory is about economic decision making. Since you are trying to say something about people's behavior regarding prices – a definite economic concept – you should use economic theory (in particular, consumer demand theory) as the framework for the statistical model. Economic theory is the guide. If economic theory is not used to guide the development of a statistical model, then you would just have parameter estimates that are uninterpretable. I provide some background on consumer demand theory in Chapter 2.

1.5.2 The role of data and data analysis

Once the statistical model based on economic theory is specified, the model's parameters can be estimated. This means that data are needed. There are two types of data: observational and experimental. Once the model's parameters are estimated, elasticities can be calculated. Sometimes a simulator is built using the elasticities to estimate demand and other key business metrics (e.g., revenue, contribution) under different conditions or *scenarios*. The optimal price is determined by studying these scenarios to test the hypotheses.

Figure 6 is an expanded version of Figure 1 showing how the quantitative pieces of data, statistical model, and simulator are used.

Figure 6: There are a number of steps in pricing research that feed into the price level part of a price strategy. The key component feeding into the level determination is the simulator, which in turn takes input from a statistical model. The types of models are discussed throughout this book.



Revealed preference data

Revealed preference data are data on consumer transactions that are usually maintained in a large database. The database typically has details on the items purchased, prices paid, characteristics of the consumers (e.g., demographics, shopping history), and sales and advertising history. For extremely large operations, the data may be stored in Hadoop clusters with very sophisticated software managing, querying, and analyzing the data. Some might say that these large databases approach "Big Data" status, especially if the company also stores web searches for products, customer online comments or reviews, emails, customer service rep interactions, and so forth. Internal customer databases are prime sources for much of the modeling done in pricing. For example, telecommunications companies gather detailed information on each subscriber regarding the number of minutes of a call, the type of call, the time of day and day of week of a call, where the call was placed to, and the consumers' calling plans. In addition, details on consumers are either maintained (e.g., address) or easily obtained either by sampling consumers in the database or by purchasing demographic data that map very closely to customer characteristics. Pricing analysts typically build models, usually regression models, from internal databases such as a data warehouse or "data mart" organized into tables of customer purchases.

A data mart is a specialized subset of a data warehouse for a business department or function. For example, the marketing department of a national seller of exercise equipment maintains a data mart of all the equipment sold to private gyms, schools, institutions, and households. The sales data are in tables called *Orders*, *Products*, and *Customers*. The Orders table contains data about each order with fields for an order number, order date, ship date, order quantity, product identification number (PID), and customer identification number (CID). The Products table has fields for linking PIDs, product description, current price, and maybe price change history. The Customers table has a linking CID, customer name, address, and maybe some demographics. By linking, filtering, and aggregating the tables, perhaps with Structured Query Language (SQL) commands, you can build a table of quantity ordered of a product, the relevant prices, and pertinent aggregate customer information. You could sample the records using an appropriate sampling technique such as simple random sampling or stratified random sampling and dividing the data into training, validation, and test tables. In the end, a pricing data mart (PDM) – a collection of data tables ready for modeling – will be produced. Figure 7 illustrates this process. Issues associated with working with Big Data plus more background on PDMs are discussed in Chapter 10.

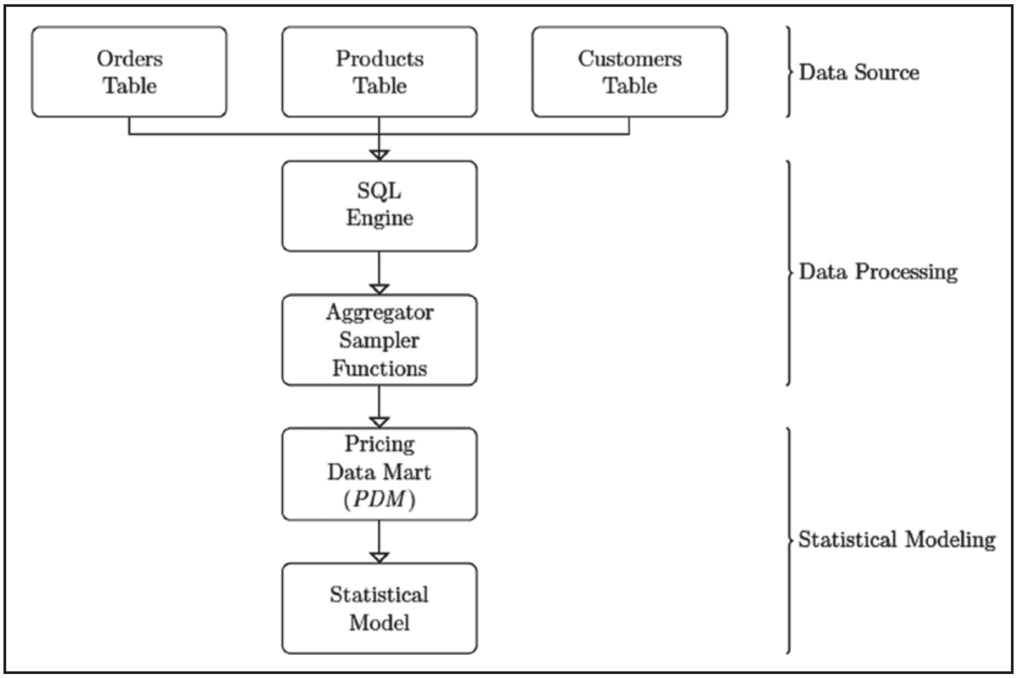
Models based on revealed preference data in the PDM run straight into five difficult and challenging data issues:

CONTINUED ON NEXT PAGE ➔

1. Multicollinearity;
2. Lack of variation;
3. Timeliness;
4. Existence;
5. Self-selection.

Multicollinearity
Many key drivers of purchases are highly correlated because of how markets work. This leads to the statistical problem known as *multicollinearity*. Simply put, some variables in the PDM are highly correlated, making them redundant. Software cannot identify which variables are effective in a regression estimation – the higher the correlation, the more the redundancy and the less able the software is to distinguish one variable from another.

Figure 7: A typical pricing modeling framework using a data warehouse of revealed preference data. The pricing data mart (PDM) must be built before any statistical model can be estimated.



Lack of variation
Some key drivers, such as price, may not vary in the market, especially for the period under study. Variation is important because statistical models work only if there is variation in the data – without variation there is nothing to estimate.

Timeliness
Observational data may be "old." If the data are, for example, monthly for five years, many of them may not reflect current market conditions, especially in rapidly changing markets such as computers and cell phones.

Existence
Since observational data are what actually exist (i.e., observed), the price levels product managers may want to test may not be reflected in the data. How can anyone say how customers would respond to prices they have never seen?

Self-selection
This is a subtle problem. The people in the data mart self-selected to be there. Self-selection means that people chose to be in the data mart because of their purchases. If they purchased elsewhere or chose not to purchase at all because the price was too high or the product quality too low, they would not be in the company's data mart as a buyer; there would be no record of them. Yet they are there because they liked the product features (including price) better than a competitor's product.

For those in the data mart, all you know is what they bought and the price they paid. You do not know what they compared the product to in order to make their final purchase decision (basically, you do not know the competitive price and features they saw). For all those in the data mart who made a purchase decision, there are many others not in the data mart either because they chose not to buy anything or because they chose to buy the competitive product at a better price; you just do not know which one. In short, you do not know the true, total population of buyers.

You could argue that at least the competitive prices are known because a competitive assessment group in marketing tracks these data using input from the sales force, advertisements, and surveys, so this is not an issue. Unfortunately, just because you know the prices and features does not mean the customers (and non-customers) knew those same prices when they made their purchase decisions. Information about the market is said to be asymmetric. You also cannot say with certainty what prices people saw because you simply do not know.

Since people self-select to buy and thus appear in the data mart, any regression estimation (e.g., ordinary least squares; OLS) would be impacted. Basically, because of self-selection, the sample is not a random sample – it is biased. A random sample would provide a description of the entire purchasing population, not just the one in the company's data mart. OLS estimates will be inconsistent because the models fail to account for this self-selection; they are misspecified. To put it briefly, inconsistency means that if you could allow the sample size to become larger, the OLS estimates would not converge to the true parameter value, in our case the elasticities; they would be biased. Intuitively, you want the estimates to be consistent so that in very large samples you get the correct answer. With a non-random sample, no matter how large the sample, it will still not reflect the true population of buyers. If the estimated elasticities based on the non-random, self-selected sample are biased, then how can you make the correct pricing decision? The whole purpose of modeling is to take the raw and chaotic data and extract the information needed for pricing to "beat the competition" in a competitive market. But if the estimated elasticities are wrong, then how can you beat anyone? In fact, the competition will win if they use correctly (or better) specified models based on a random sample.

The model misspecification is due to two decisions. The first is a vendor choice decision (who they will buy from) and the second

CONTINUED ON NEXT PAGE ➔

is a quantity decision (how much they will buy). The typical model focuses only on the second of the two decisions, the first being ignored. But the first cannot be ignored since the two decisions are not independent. They are functions, typically, of the same key drivers such as price and product attributes. There are two models! The vendor decision model can be summarized in a key ratio[9] that is incorporated into the quantity decision model. To ignore this new variable is to omit a relevant variable and thus to misspecify the model, and misspecified models lead to biased results.

This is a complicated problem, but there are two recommendations. First, recognize that a problem exists because of self-selection. This is not an insurmountable problem or a killer of data warehouses or data marts. Failure to recognize it or, worse, ignoring it is to help the competition win. Second, do not rely solely on data warehouses and data marts as *the* sources of all data. Aside from self-selection, the four other problems mentioned above still hold and have to be addressed. These are characteristics of observational data. Consider using discrete choice experiments in which these problems, including self-selection, are avoided. The self-selection problem, for example, is a non-issue because the sample used in an experiment is a random sample of all potential buyers. Discrete choice experiments are discussed in Chapter 5.

Stated preference data

Stated preference data are data on key drivers of purchases, especially price, which are collected and manipulated under controlled conditions determined by an experimental design. Conjoint and discrete choice studies are the main approaches used to collect and analyze data. Statistical models are estimated for stated preference data just as they are used for revealed preference data, but unlike revealed preference data, stated preference data do not have the five flaws. The stated preference designs ensure that key factors are not correlated and have sufficient variation. Also, by the nature of experimentation, hypothetical situations can be posed to customers to get their reactions, so timeliness and existence are not issues. Finally, both buyers and non-buyers can be included in the study, so self-selection is not an issue. In fact, people will actually state that they will not buy as part of the study. For many pricing studies, stated preference data approaches are far superior and highly recommended. The gold standard for pricing analytics should be stated preference analytics.

1.5.3 The role of statistical models

A statistical model – the third part of the research triangle – is necessary because the raw, almost chaotic data have to be organized and the important elements (i.e., information) extracted. That can only be done by estimating the key (unknown) parameters of a model.

The model is a statement of relationships you expect to learn about from the data. Without a model, you would just be searching (mining?) for relationships that seem plausible. Such searching is called *unsupervised learning*; using a model is called *supervised learning*. This is not to say that unsupervised learning is not important – it is, in some cases. But this is to say that a structured approach to learning is better, especially in this case since you are looking for a particular numeric quantity – an elasticity – to help price a product. Fishing in a pool of data will not yield the number; it has to be calculated, and a model tells you how.

You should distinguish between a *mechanistic model* and a *statistical model* because there are similarities that may obfuscate the critical differences.

Mechanistic model

A mechanistic model is one for which you know exactly the value for a variable Y when you know a value for one or more other variables. The relationship is exact. You could refer to a mechanistic model as *deterministic* because of this exactness. A classic example from physics is Einstein's famous equation, $E = MC^2$. If you know the mass of an object, M, and the speed of light, C, then you know the amount of energy, E, and you know it exactly. Energy, E, is determined mechanically from this equation.

You can write any mechanistic model as

$$Y = f(X) \tag{1}$$

where Y is a variable to be explained, sometimes called the *dependent variable* or the *response variable*, and X is a vector of variables that explain or drive or determine Y. These variables are sometimes called *independent variables*, *key driver variables*, or *predictors*. They are non-random (C, the speed of light in Einstein's equation, is, after all, fixed by nature) and can be measured perfectly. The notation $f(\cdot)$ represents some function of the non-random X. For Einstein's equation, $f(\cdot)$ is a multiplicative function of two variables, M and C^2.

A common functional form is a linear model such as

$$Y = \beta_0 + \beta_1 X \tag{2}$$

where β_0 and β_1 are two (usually unknown) parameters. If you have two independent variables, you write

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2. \tag{3}$$

In general, for p independent variables, you write

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \tag{4}$$

$$= \sum_{i=0}^{p} \beta_i X_i \tag{5}$$

where $X_0 = 1$.

Einstein's equation is called *inherently linear* because it can be linearized by taking the log of both sides[10]:

$$\ln(E) = \ln(M) + 2 \times \ln(C) \tag{6}$$

which has the same form as (2). That is, you can write

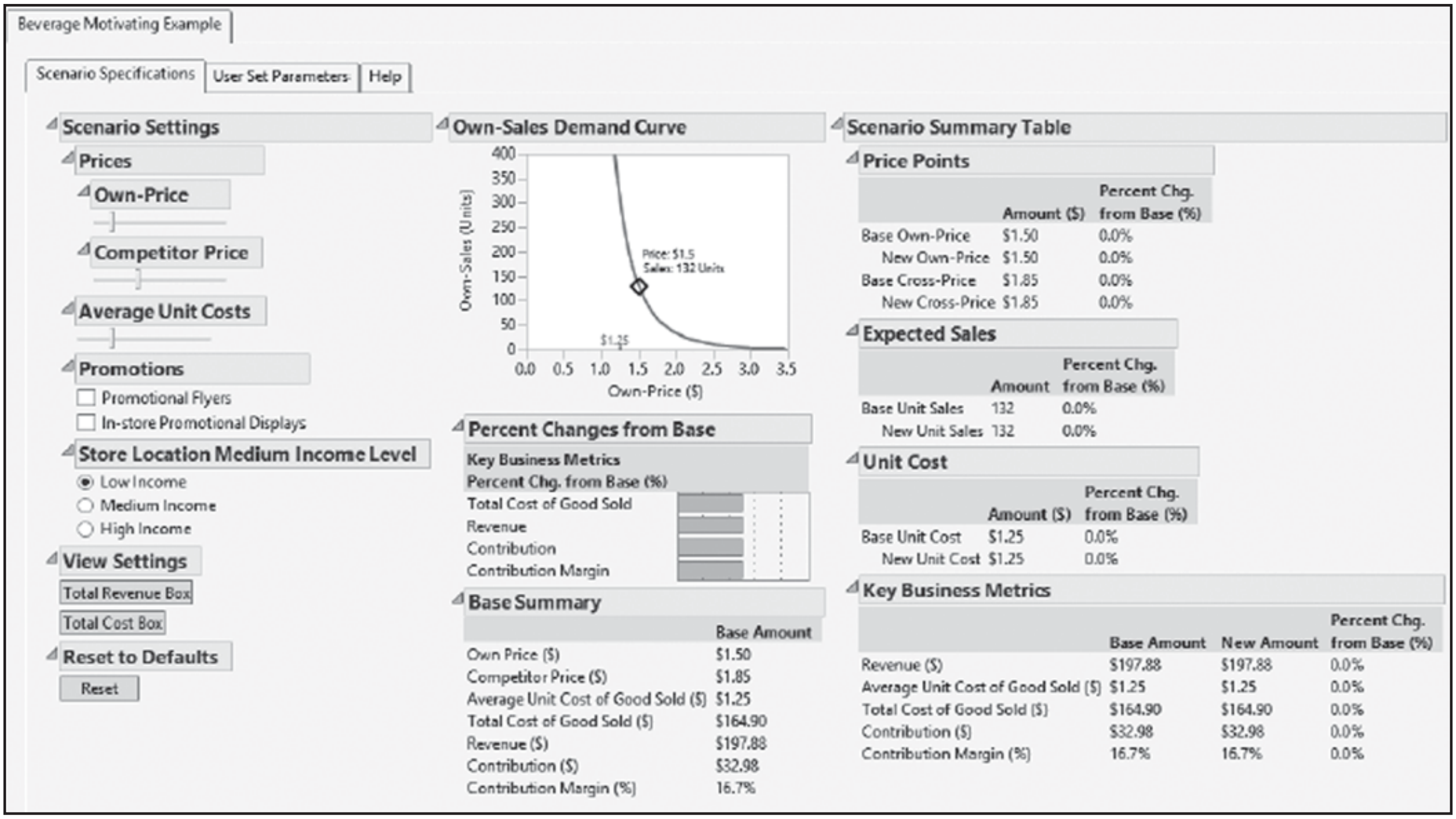
$$\ln(E) = \beta_0 + \beta_1 \times \ln(M) + \beta_2 \times \ln(C) \tag{7}$$

with $\beta_0 = 0$, $\beta_1 = 1$, and $\beta_2 = 2$.

Mechanistic models have no role in pricing because they lack a

CONTINUED ON NEXT PAGE →

Figure 8: A pricing simulator allows the you to test different pricing scenarios for developing the optimal price point. This screenshot shows the interface for a simulator that allows you to test different prices in order to determine the effect on key business metrics such as revenue and contribution.



stochastic element. Randomness is a part of the real economic world and that randomness must be accounted for. This is where a statistical model enters the picture.

Statistical model

A statistical model is similar to a mechanistic model in mathematical form, but it differs in that it describes how one or more random variables are related to one or more other random variables. The model is statistical since the dependent and independent variables are not deterministically but *stochastically* related.[11] A simple statistical linear model is written as

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where is a stochastic or random variable that represents the factors that also drive Y but that are unknown. Since is a random variable, Y is also a random variable. Simple statistical modeling typically treats X as non-stochastic, although advanced econometrics recognizes the stochastic nature of X.

The random variable, called a *disturbance term*, has special significance in this model because it represents all the other factors that could drive Y but that are not our central focus or are unknown or unknowable. The X is our central focus and represents our *testable hypothesis* of how Y is determined. It is our hypothesis about what causes Y and it is testable because you can use data with statistical methods. The X is based on a theory, the top part of the research triangle, or solid intuition or experience of what drives or determines Y.

A theory such as, for example, consumer demand theory in economics provides the foundation for identifying just those few special factors from the plethora of factors that determine quantity sold, the Y. These may be stated explicitly in the theory, with their relationships to Y well stated, or they may be derived by implication, as is typically the case. In either case, these are the ones you want to focus on and the theory tells you why. All other factors are subsumed in the ε. As a random variable, we usually assume that $\varepsilon \sim N(0, \sigma^2)$, although other specifications are possible, as you will see with discrete choice models in Chapter 5.

1.6 Simulators

The simulator in Figure 6 is a very important part of the pricing analytics process. This is where the results of the statistical model, in the form of the estimated parameters and elasticities, are pulled together to determine the optimal price level and the effects on the key business metrics. The simulator allows you to test different pricing scenarios and, depending on the quality of the simulator, also allows upper management to do the same if they wish.

1. The simulator should allow for the ability to flexibly change:
2. The business's own product price;
3. Any key competitor's price;
4. Key product attributes if the issue is new product pricing;
5. Average incremental cost;

CONTINUED ON NEXT PAGE ➔

6. Key promotional features.

The simulator should show the key business metrics such as revenue, total costs, contribution, and contribution margin. Figure 8 shows a screenshot of a simulator for an example developed in Chapter 8.

Price elasticities

The economic concept of a price elasticity, also sometimes referred to as *price sensitivity* or *price responsiveness*, is the key result of pricing analytics. Using the elasticities, you can propose a base or reference price (perhaps the existing market price) and then determine how sales would change as price deviates from this base case, but without actually changing the price in the market. The simulator, of course, is the tool to use for this analysis. Since revenue is merely price times quantity, the impact on revenue, or any other key business metric, can be quickly assessed.

Unfortunately, even in books on pricing, the concept of an elasticity is often given the back seat to discussions about strategies and tactics. Being customer focused means understanding how the customer will react to changes in his/her environment, with pricing being just one aspect of that environment.

Elasticities summarize the customer's reaction to changes, whether the changes are the firm's actions or a competitor's actions. I devote Chapter 2 to a detailed examination of elasticities because of their importance, while in Chapter 3 I discuss how elasticities are used in pricing. The remainder of the current book focuses on estimating elasticities.

Notes

1. [49]. McCarthy introduced this framework with four *P*s in 1960. Four is most common, although the list has expanded over the years. I prefer the five listed in the text. See the Wikipedia article on marketing mix for some discussion.

2. There may be minor implementation costs in deploying a new price. After all, computers have to be reprogrammed for the new price. But these costs are small compared to the price of an ad campaign.

3. [4]. Also see [43].

4. [61] originally proposed a theory of revealed preference as an alternative to utility theory. This has since generated much criticism, controversy, and a wide literature. See, for example, [71]. I am not concerned with the pure theory of revealed preference for this book.

5. A new phase, Phase 0, has been added. See the Wikipedia article "Clinical trial" at en.wikipedia.org/wiki/Clinical_trial.

6. See www.cancerresearchuk.org/cancer-help/about-cancer/cancer-questions/how-long- does-it-take-for-a-new-drug-to-go-through-clinical-trials.

7. See en.wikipedia.org/wiki/Cost-plus_pricing.

8. Cited in [16].

9. This is called the inverse Mills ratio and is used in a two-stage regression estimation to handle selection bias. See [25] and [24].Also see [81], which develops a brand and quantity choice model. This is briefly described in Chapter 12.

10. Natural logs are typically used because they have useful mathematical properties, as you will see later.

11. See Wikipedia article "Statistical model."