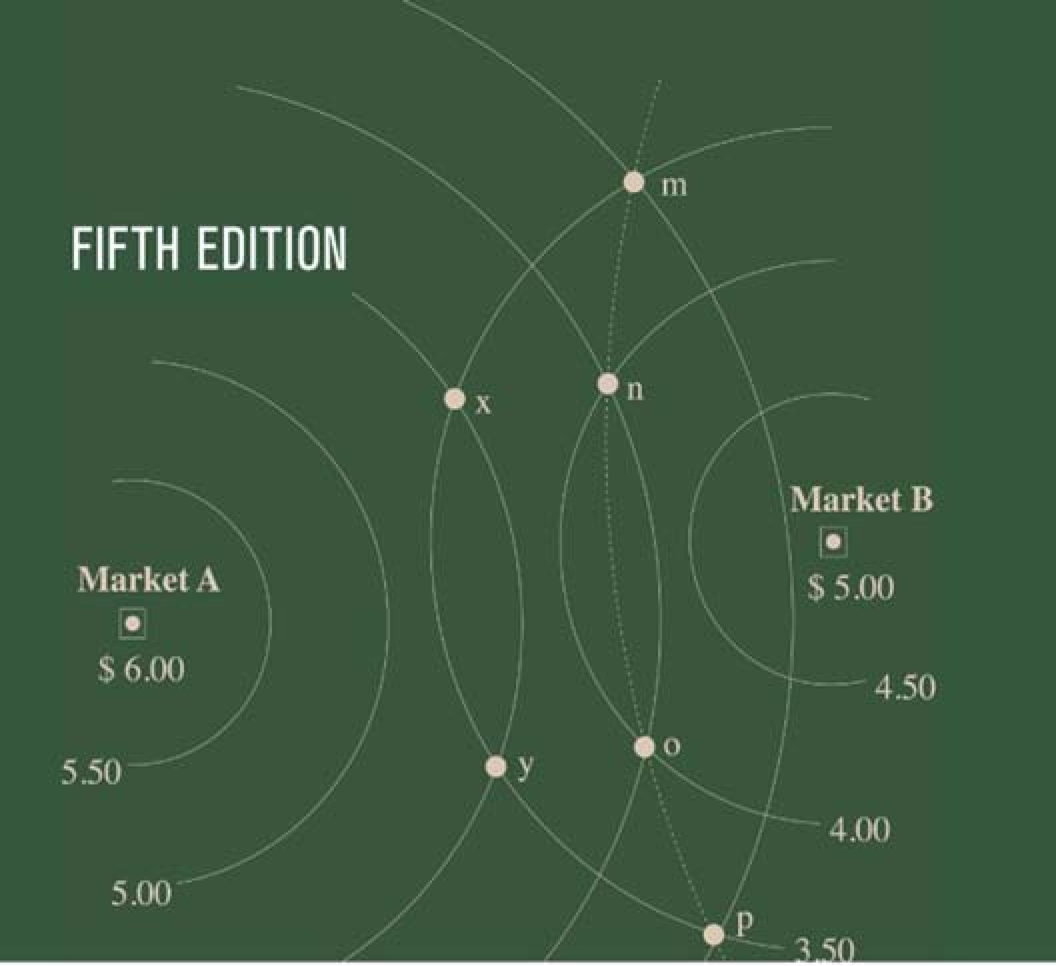
FIFTH EDITION




# AGRICULTURAL PRODUCT PRICES

William G. Tomek
Harry M. Kaiser

# Agricultural Product Prices

# AGRICULTURAL PRODUCT PRICES

FIFTH EDITION

**William G. Tomek**

**Harry M. Kaiser**

**CORNELL UNIVERSITY PRESS**

ITHACA AND LONDON

First published 2014 by Cornell University Press
Printed in the United States of America

*Library of Congress Cataloging-in-Publication Data*

Tomek, William G., 1932– author.
Agricultural product prices / William G. Tomek, Harry M. Kaiser. – Fifth edition.
pages cm
Includes bibliographical references and indexes.
ISBN 978-0-8014-5230-7 (cloth : alk. paper)
1. Agricultural prices – United States. 2. Farm produce – Prices – United States. I. Kaiser, Harry Mason, author. II. Title.

HD9004.T65 2014
338.1′30973 – dc23 2013043251

Cornell University Press strives to use environmentally responsible suppliers and materials to the fullest extent possible in the publishing of its books. Such materials include vegetable-based, low-VOC inks and acid-free papers that are recycled, totally chlorine-free, or partly composed of nonwood fibers. For further information, visit our website at www.cornellpress.cornell.edu.

Cloth printing 10 9 8 7 6 5 4 3 2 1

# Contents

# Kenneth L. Robinson: An Acknowledgment

Kenneth L. Robinson, a Liberty Hyde Bailey Professor of Agricultural Economics at Cornell University, was a co-author of the first three editions of this book. He contributed in important ways to the content and organization of these editions. Although much of the text was rewritten for the fourth edition, he was explicitly recognized as a contributor, as that edition continued to reflect his organizational input and some of his writing. The fifth edition, of course, involves changes from the fourth and has a new co-author, Harry M. Kaiser. Nonetheless, the book still reflects some of Ken's philosophy about the relationship of textbooks to classroom teaching and contains some of his words. It is important that we acknowledge his influence. Robinson was a master teacher of undergraduate and graduate students. He firmly believed that a textbook should provide a foundation of basic principles with applications but that teachers must do much additional work to adapt any text to their specific situation. Ken also wanted the book to have a long shelf life with scholarly content and, hence, be useful for beginning graduate students as well as juniors and seniors. Moreover, he was concerned that the book have a modest price, reflecting his abiding concern for the welfare of students. These views required that the writing be succinct but not obscure and, in 1972, minimize mathematical notation. Because of the reduced cost of using mathematics, the fourth edition and this edition have somewhat more notation than the earlier editions. Otherwise, Robinson's general concept remains: provide a concise, relatively inexpensive book that applies economic principles to the study of agricultural prices. We are pleased to acknowledge his lasting influence on the contents and style of this book.

# Preface to the Fifth Edition

The fifth edition, of course, updates tables and figures from the fourth edition, and as noted below, addresses the effects of changes in markets for agricultural products. The organizational structure of the fifth edition is similar to the fourth, but one organizational change was to move the material on tariffs and prices, previously in the appendix to Chapter 11, to an expanded section on spatial price relationships and trade policy in Chapter 8. A section on domestic agricultural policy is, however, retained in Chapter 11 as part of the discussion of administered prices.

A major challenge for authors of textbooks in applied economics is the growing complexity and diversity of the worldwide economy. The fifth edition discusses the nature of this increased complexity and its effects on price behavior, but without greatly lengthening the book. For example, we discuss the increased diversity of the derived demands for commodities, including both new sources of demand and changes in demands for attributes (e.g., in Chapters 2 and 7). Likewise, Chapters 5, 6, 11, and 12 convey the changing nature of the structure of agricultural markets and pricing institutions. One example is the increased diversity of contracts offered by buyers to farmers. Nonetheless, the fifth edition continues to emphasize the use of basic economic models to provide an understanding of the behavior of agricultural product prices and of the consequences of this behavior.

Thus, this edition is still appropriate for upper-division courses in agricultural markets and prices. Like past editions, the book has sufficient breadth to give teachers flexibility in the choice of topics for their particular course. The references provide more depth and perhaps make the book useful for masters-degree courses. Graduate students have also indicated that the book is useful background for research and for preparing for examinations, and we hope that this remains true. Indeed, if carefully read, the book implies subjects where additional research would be valuable.

As in the past, Cornell University Press sought reviews from knowledgeable, but anonymous, reviewers, and they made many suggestions. Some have been incorporated in the revisions, and we are grateful for them. It was not possible, however, to adopt all the suggestions without massively lengthening the book. These varied suggestions perhaps reflect the increased complexity and diversity of agricultural markets, mentioned above.

We have benefitted from interactions with many outstanding undergraduate and graduate students at Cornell University, too numerous to list, but wish to acknowledge their role in keeping us "sharp." As noted in previous prefaces, Cornell University Press staff provided helpful guidance. We also wish to acknowledge the research support of Nadia Streletskaya, who helped update data and prepare figures and tables, and the assistance of Gretchen Gilbert in converting old Word Perfect files to Word documents.

W. G. T.
H. M. K.

*Ithaca, New York*

# Agricultural Product Prices

CHAPTER 1

# Introduction

The principal objective of this book is to provide students with an understanding of the complex array of forces that influence the level and behavior of agricultural product prices. A secondary objective is to assist students in bridging the gap between theory and empirical analyses of price behavior. Such analyses aid in understanding the performance of the economy, including the consequences of price or policy changes. Models are also used for price forecasting and other areas to assist decision makers, and we will discuss why high-quality forecasts are difficult to attain.

Although the agricultural sector is a declining component of most national economies, agricultural product prices remain important both economically and politically. They strongly influence the level of farm incomes, and in many countries, the levels of food and fiber prices are important determinants of consumer welfare and the amount of export earnings. A decline of only a few cents per pound in the prices of such internationally traded commodities as sugar, coffee, and cocoa can have serious political and economic repercussions in such countries as Mauritius, Colombia, and Ghana. As Deaton (1999) pointed out, inaccurate forecasts of commodity prices led to poor policy prescriptions for African nations.

Concerns about commodity price behavior also exist in the United States. For example, a large drop in the farm price of hogs was reported as likely to drive 24,000 pork producers out of business (*Wall Street Journal* 1998). This, in turn, led to questions about why hog prices had dropped so rapidly and by such a large amount. Large increases in commodity prices and their potential effects on food prices, as in 2006, have also been a cause for concern, and a variety of explanations exist, including blaming "speculators." Or, since ethanol is produced mainly from corn (maize) in the United States, have biofuel policies caused corn and food prices to rise (de Gorter and Just 2010)? This book provides a foundation for addressing such issues.

## Distinguishing Characteristics of Agricultural Prices

Agricultural commodities provide an exceptionally interesting vehicle for the study of price-making forces. The manner in which commodity prices are determined ranges from markets with near monopoly-like institutions, sometimes assisted by governmental regulation and intervention, to approximations of the textbook definition of pricing under competitive conditions. Thus, in examining agricultural prices, one must inevitably study a wide range of models of price determination and pricing institutions.

Agricultural commodity prices are substantially more volatile than are the prices of most nonfarm goods and services. It is not unknown for the prices of commodities to drop by 75 percent or spike upward by 100 or more percent within only a few months. For instance, the monthly price of onions in New York State ranged from $12.70 to $38.60 per hundredweight (cwt.) in the 2006–07 season. Major commodities like sugar, corn, wheat, and cotton also have variable prices. The price of hard red winter wheat in the United States was $6.54 per bushel in December 2011 and $8.21 a year later, a 26 percent increase.

A typical time series of commodity prices exhibits both random variability and systematic behavior.[1] Sometimes, as suggested above, spikes occur; i.e., prices jump rather abruptly to a high level and then fall back to the original level or even lower. These generalizations apply roughly to series of varying frequency, such as daily, monthly, or annual average prices, but it will be convenient in subsequent discussions to distinguish between inter-year and intra-year price behavior.

Much of our discussion will be about the behavior of average prices, say why a series of monthly prices changes with the passage of time. A more complete discussion would include the possibility of changes in other descriptive statistics like the variance. For example, the mean, variance, kurtosis, and skewness of the distributions of monthly corn prices all vary systematically from harvest through the storage period (Peterson and Tomek 2005). We, therefore, provide some discussion of models of the behavior of the variances of prices.

The characteristics of agricultural product price behavior relate in important ways to the biological nature of the production process. Significant time lags

1. One way to describe the systematic behavior of prices through time is by using linear statistical models. Prices "today" may be related to prices "yesterday," and to prices the "day before that," etc. Such prices are said to be autocorrelated, and in a statistical model, current prices can be made a function of lagged prices. This model provides one type of estimate of the systematic behavior. Indeed, we shall show in subsequent chapters of the book that cash (i.e., for immediate delivery) prices of commodities should be autocorrelated. In contrast, the prices of financial assets, such as the price of Coca-Cola Company stock or the price of an industrial bond, are commonly modeled as not autocorrelated.

exist between a decision to produce and the realization of output, and actual production may exceed or fall short of planned production by a considerable margin. At least a year is required for producers to change hog production, 3 years to expand the supply of beef, and 5–10 years for growers to increase the output of tree crops, such as apples. Yields vary from year to year because of variability in weather conditions and the presence or absence of diseases or insect infestations.

Farmers' production decisions are based partly on their expectations about future yields and prices (i.e., expected profitability) of the alternative commodities that they might produce. Since these expectations are not always realized, price and yield risks exist in farming, and the way expectations are formed and acted on by farmers may impart a cyclical component to supply and prices. Also, the nature of resources, like land and equipment, used in farming is such that producers cannot easily make major changes in production plans in response to expected price changes. Thus, given a change in the price that farmers expected to prevail, the change in the quantity supplied, at least over short periods of time, is relatively small (where both changes are measured in percentage); supply is said to be *price inelastic* (see Chapter 4).

The nature of the demand for farm products is also a factor in price instability. Farm commodities are often the raw materials used to manufacture a large variety of end products; they are also exported and stored for future use. Farm-level demand is derived from this broad array of uses, and the magnitude of these uses depends on such factors as population growth and income levels. Like supply, the demand for most commodities is price inelastic (see Chapter 3); changes in prices have relatively small effects on the quantities consumers are willing to buy. Put another way, relatively small changes in production induce a relatively large change in prices. Moreover, the diverse sources of demand mean that many sources of potential "demand shocks" exist. A financial crisis in Southeast Asian countries can, for example, reduce the demand for food imports and thereby reduce the prices of the imported commodities, or a health scare can reduce demand and prices, at least temporarily (e.g., Attavanich, McCarl, and Bessler 2011).

Macroeconomic variables, including policies related to the money supply, governmental budget deficits, exchange rates, trade and foreign aid (which affect the capacity of countries to pay for imports), now play a greater role in determining farm prices than they did four or more decades ago. Many agricultural markets are now international in scope. For example, soybean prices are sensitive to changes in soybean production in Brazil and Argentina as well as in the United States; they are also affected by growing demands for soybeans in countries like China.

Price determination at the first-handler level tends to be more competitive and more decentralized in agriculture than in other industries (Chapter 8). For many commodities, there are a relatively large number of producing units that are geographically dispersed. Although farms are becoming larger, their concentration is still less than in most other industries. The buying side of these markets is becoming more concentrated, often with only one or few buyers in local areas. Nonetheless, when viewed from a broader, national or international perspective, these markets still appear relatively competitive (for additional discussion of market structure, see Chapter 5).

In any case, it is unquestionably true that commodity prices are variable, not only from year to year but from day to day. Commodity markets are sometimes called *flex price* markets, which contrast with *fixed price* markets where prices change slowly and by small amounts. Automobile manufacturers set list prices and might raise prices by 5 percent from one model year to the next; commodity prices can change by more than 5 percent in one day.

Moreover, it has been argued that, in the short run, agricultural commodity prices can overshoot long-run equilibrium levels in response to changes in economic forces, including macroeconomic variables (Saghaian, Reed, and Marchant 2002). Overshooting occurs partly because so many prices in the nonfarm sector change slowly (Frankel 1986).

## Role of Prices

Prices play a central role in economic theory in guiding production and consumption. We are under no illusion, however, that either the production decisions of farmers or the buying decisions of consumers are governed solely by prices. Government programs, including land-retirement or conservation measures, as well as the constraints of climate and soil types, the availability of equipment and production technology, and personal preferences obviously influence farmers' production decisions. Consumers are likewise influenced by advertising, convenience attributes, age and experience, income level, and personal whims and habits as well as by prices.

Notwithstanding the complications introduced by nonprice factors, consumers do respond to changes in relative prices, say a change in the price of chicken relative to beef. Farmers, likewise, have demonstrated repeatedly that they will produce more cabbage, corn, milk, onions, and other products in response to relatively favorable prices. Thus, an understanding of the concepts of economic theory and their application to commodity markets does provide valuable insight into the way prices behave.

A complication in analyzing supply, demand, and price behavior is that neither consumers nor producers respond to price changes in fixed manner with the passage of time. The degree of responsiveness of quantity to a given size price change may itself change with the passage of time. As consumers become more affluent, their purchases of particular food products may become less responsive to changes in prices (and changes in income), and as farmers become more specialized and make large fixed investments, their output also may become less responsive to price changes.[2]

Governments have played varying roles in pricing farm products. From the 1930s through much of the 1980s, price-support programs in the United States influenced the prices of commodities that represented about one-half the value of farm output. Government programs have also influenced prices in many other countries around the world. Since the beginning of the 1990s, there has been a trend toward lower price-support levels and fewer trade restrictions, although government programs and regulations have not been completely abandoned. The changing nature and role of government programs are a potential complication in analyzing the behavior of commodity prices over a period of years.

It is important to note that pricing decisions, whether made on the basis of market forces or political considerations, have important economic consequences. For this reason, tools of analysis that will help one anticipate the economic effects of changes in government programs or of the decisions made by private firms are still important. Farmers, marketing and supply firms, and government officials have to make many decisions that require a knowledge of what will happen if the price of a particular commodity rises or falls. Meat packers, for example, want to know how much the production of hogs will change in response to the price of hogs or, more important, to the price of hogs relative to corn. Government officials need to know how domestic use, exports, and production will be influenced by changes in price-support levels. Firms that store apples would like to anticipate the consequences for late-season prices of putting more apples in controlled-atmosphere storage, which lengthens the storage life of fruit. Students of prices can help answer such questions.

## Plan of Book

The book is divided into four major sections, with two or more chapters per section. The first section is devoted to a review of the economic concepts that underlie price determination, particularly as they apply to agricultural commodities. The second section deals with price variation through time, across space, and by quality attributes as well as with the linkages among prices at the farm, wholesale, and retail levels. The next section is devoted to a description and analysis of alternative pricing arrangements for agricultural commodities, such as auctions and negotiated prices. Particular attention is paid to the role of futures markets in the price discovery process. The final section provides an

2. These statements illustrate possible changes in the degree of responsiveness of producers and consumers to price, and they should be treated as hypotheses subject to empirical verification. Since economic growth is associated with the development of new products and new substitutes and since increases in the number and importance of substitutes imply greater sensitivity of quantity to price, increased affluence could lead to greater rather than less responsiveness by consumers to price changes.

introduction to topics related to the empirical analysis of commodity supply, demand, and prices.

## References

Attavanich, W., B. A. McCarl, and D. Bessler. 2011. "The Effect of H1N1 (Swine Flu) Media Coverage on Agricultural Commodity Markets," *Applied Econ. Persp. & Policy* 33:241–259.

Deaton, A. 1999. "Commodity Prices and Growth in Africa," *J. Econ. Persp.* 13:23–40.

de Gorter, H., and D. R. Just. 2010. "The Social Costs and Benefits of Biofuels: The Intersection of Environmental, Energy and Agricultural Policy," *Applied Econ. Persp. & Policy* 32:4–32.

Frankel, J. A. 1986. "Expectations and Commodity Price Dynamics: The Overshooting Model," *Am. J. Ag. Econ.* 68:344–348.

Peterson, H. H., and W. G. Tomek. 2005. "How Much of Commodity Price Behavior Can a Rational Expectations Storage Model Explain?" *Ag. Econ.* 33:289–303.

Saghaian, S. H., M. R. Reed, and M. A. Marchant. 2002. "Monetary Impacts and Overshooting of Agricultural Prices in an Open Economy," *Am. J. Ag. Econ.* 84:90–103.

*Wall Street Journal*. 1998. News item. November 27, A2.

PART I

# PRINCIPLES OF PRICE DETERMINATION

Selected elements of the principles of price determination are reviewed in this part of the book. Principles of demand theory and their application to the demand for agricultural products are discussed in Chapter 2, and elasticity and price flexibility concepts are described in Chapter 3. Principles of supply theory with special reference to the supply of farm products are discussed in Chapter 4. Principles of demand and supply are combined in selected models of price determination in Chapter 5; models of particular interest in agricultural markets are stressed. Indeed, each chapter goes beyond simple economic principles to elaborate on their application to agricultural product markets.

CHAPTER 2

# Demand for Agricultural Products

The objective of this chapter is to review elements of demand theory, relating them to the demand for agricultural commodities. An understanding of demand concepts is important because it helps to explain price behavior. A common approach is to think in terms of retail-level demand by consumers of final products such as the demand for fluid milk at retail, but in studying prices of agricultural commodities, it will also be necessary to consider demand at the farm level, i.e., the derived demand for commodities. Both topics are covered in this chapter.

## Basics of Demand Theory[1]

The basic unit of demand theory is the individual consumer. Each consumer is confronted by a problem of choice. A large number of wants arise from basic needs (e.g., food and shelter), personal characteristics, and the social and physical environment. On the other hand, the consumer has a limited income. Thus, the problem is to choose the specific goods and services that "best" satisfy these wants within the limits imposed by income.

Economists define *best* in terms of the consumer's attempt to maximize utility, which is a measure of well-being that depends on the consumption of goods and services. The utility approach to the theory of demand can be stated mathematically. This involves the maximization of a utility function subject to a budget constraint. The theoretical concept of utility as a function of goods consumed could be given empirical content if we knew the algebraic form and coefficients of the utility function. Then, the classical mathematics of constrained optimization could be used to derive explicit demand relations for the

1. This chapter provides an introduction to demand concepts using intuitive, rather than completely rigorous, arguments. Students should consult an intermediate microeconomics text for a more in-depth discussion of demand theory (e.g., Varian 2010).

consumer. In practice, the utility function is used as a conceptual device to illustrate consumption theory.

On the basis of such theory, we can conclude that a consumer tends to prefer more to less of a good but will buy more only at a lower price. That is, there is an inverse relationship between quantity demanded and price. Also, a number of useful general theorems about relationships among elasticities have been derived from the idea of maximizing a utility function subject to a set of constraints, and empirical analyses of demand often use this framework to obtain internally consistent estimates of elasticities of demand. These topics are discussed in Chapter 3.

## Consumer and Market Demand

*Consumer demand* is defined as the various quantities of a particular good that an individual consumer is willing and able to buy as the price of that good varies, with all other factors that affect demand held constant. The consumer demand relation can be described either as a table of prices and quantities (a demand schedule) or as a graph or algebraic function of prices and quantities (a demand curve). The demand relation simply defines the relationship between price and quantity demanded per unit of time while holding other factors constant.

Price and quantity vary inversely; that is, the demand curve has a negative slope. This inverse relationship is sometimes called the *law of demand,* and it can be explained in terms of the substitution and income effects of a price change.

These effects can be illustrated by constructing an "indifference map," which is a graphical method of describing a consumer's preferences, recalling that these preferences are given or fixed at any point in time. In Figure 2-1, the quantity of X (say, food) is shown on the horizontal axis, and the aggregate quantity of all other goods (Y) is shown on the vertical axis. Each indifference curve or isoquant ($U_1$, $U_2$, or $U_3$) identifies the various combinations of X and Y that will give the consumer equal satisfaction (i.e., utility). For example, the consumer depicted in Figure 2-1 will be equally as well off on $U_1$ with 20 units of X and 65 units of Y as with 55 units of X and 23 units of Y. The higher the indifference curve, the greater the consumer's utility. It also should be noted in passing that the illustration incorporates two other common assumptions: namely that the commodities are continuously divisible and that the consumer finds both commodities desirable. Thus, the indifference curves are continuous and have a negative slope.

Assuming that the consumer has $500 to spend each "month" and that the price of Y is $5 per unit, the maximum amount of Y that can be purchased per month is 100 units. Likewise, if the price of X is $10, the consumer can purchase a maximum of 50 units per month. Based on these price and income assump-

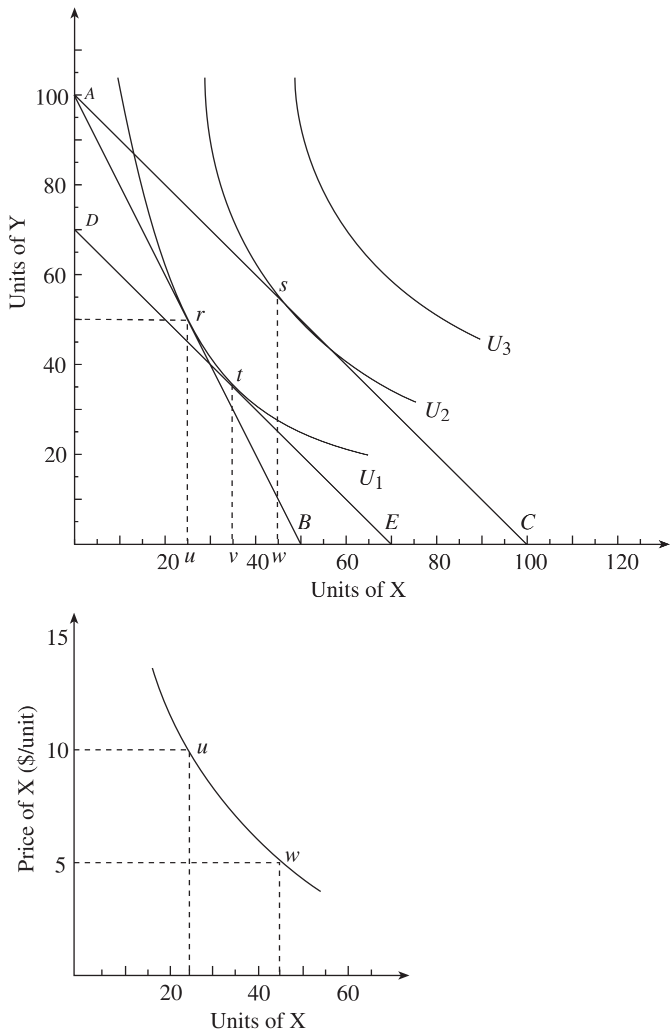



**Figure 2-1.** Relationship of consumer preferences to demand



tions, a line AB can be constructed showing the maximum quantities of X and Y that can be purchased per month with $500. This "budget" or "price-ratio" line establishes the upper boundary for purchases of X and Y. Any combination of X and Y on or below line AB represents a feasible purchase plan; any combination above and to the right is not feasible. The slope of AB is determined by the price of Y relative to X, and the distance of the line from the origin—its level—is dictated by the income constraint.

The consumer's total utility is maximized by selecting the combination of X and Y that enables the individual to reach the highest feasible point on the utility surface. This is the point at which the budget line just touches the highest indifference curve, the point of tangency *r* in Figure 2-1. The precise quantity of X that should be purchased to maximize utility is determined by dropping a line from the point of tangency (*r*) to the horizontal axis. A corresponding line can be drawn parallel to the horizontal axis through *r* to determine the optimum quantity of Y. Thus, in Figure 2-1, the combination that maximizes utility within the income constraint of $500 is 25 units of X and 50 units of Y (25 units of X times $10 per unit plus 50 units of Y times $5 per unit equals $500, the total income available).

The foregoing example is a static snapshot assuming fixed prices. A decrease in the price of X will enable the consumer to purchase more of X with the same income. The change in the price of X will also influence the purchase of Y. The price-ratio line intersects the *X* axis at a greater distance from the origin as the price of X declines, given that the price of Y and income remain constant. For example, if the price of X falls to $5 per unit, 100 units can now be purchased. This is represented by the new line *AC*. Since the prices of Y and X are both $5, the slope of the line is −1.0. The new point of tangency that maximizes utility is at *s* on $U_2$. The new optimum purchase plan consists of 45 units of X and 55 units Y ($5X + 5Y = 5(45) + 5(55) = 500$). With the decrease in the price of X, the consumer maximizes utility by purchasing 20 additional units of X and still has sufficient income left over to purchase an additional 5 units of Y. In other instances, the consumption of Y could decrease as a result of a decline in the price of X.[2] The general point is, however, that a change in the price of one product can have important secondary effects on the consumption of other products.

A decrease in the price of any product is equivalent to an increase in real income; more can be purchased with the same amount of money. Income effects are generally positive; that is, an increase in income leads consumers to purchase more of the good. (Exceptions are noted below.) Assuming the income response is positive, the consumption of X will benefit from a fall in price because of the income effect as well as the substitution effect. If the product in question accounts for a large proportion of total expenditures by the consumer, then clearly the income effect of a change in price of that product will be substantial. It will be less important for a product that accounts for a small percentage of total expenditures. The substitution effect is always inversely related to

2. If a decrease in the price of X results in an increase in the consumption of Y, then the demand for X is price inelastic. The percentage increase in the quantity of X is less than the percentage decrease in the price of X, and as a consequence, the total expenditure on X increases with the decrease in the price of X. This inelasticity is a function of the indifference map. The demand for X could be price elastic, and in this case the decline in the price of X would result in a decrease in the consumption of Y. The concept of elasticity is explored in greater depth in Chapter 3.

price changes; that is, a consumer will tend to buy more as the price decreases, and vice versa.

The income effect can be separated from the substitution effect of a price change using the analytical apparatus of indifference curves. In the example used earlier, a decrease in the price of X from $10 to $5 led to an increase in consumption of X from 25 to 45 units. To separate the income and substitution effects, one can draw a line *DE* parallel to line *AC* that just touches the highest indifference curve previously attainable, $U_1$ (Figure 2-1). This represents the change in income required to offset or compensate for the real income effect of the fall in the price of X. With the lower price of X and lower level of income represented by line *DE*, the consumer is just as well off (on the same indifference curve) as with the higher price and higher money income. The point *t* represents the optimum expenditure pattern with the new prices and the lower level of money income. The move along indifference curve $U_1$ from *r* to *t* is attributable to the substitution effect of lowering the price of X. The change from *t* to *s* is the result of the increase in real income arising from the fall in the price of X. By dropping vertical lines from points *r, s,* and *t* the total increase in consumption (*uw* on the horizontal axis) can be partitioned into the substitution effect (*uv*) and the income effect (*vw*).[3]

In the foregoing illustration, the income effect served to reinforce the substitution effect, but this need not be the case. The substitution effect of a price change is always negative (Okrent and Alston 2011: 5–8); that is, an increase in price invariably results in a decrease in consumption if there is an offsetting change in money income, thus keeping real income constant. The income effect also is generally negative; an increase in price results in a decrease in real income and hence a decrease in demand. For a few goods, the relationship is reversed; real income and demand are inversely related. Such goods are called *inferior goods.* Perhaps the demand for dry beans, at least by high-income consumers in the United States, has a negative income effect. For such goods, the income effect offsets part or all of the substitution effect.

At the extreme, the income effect of an inferior good's price change could outweigh the substitution effect. Then, a price increase would result in an increase in the quantity demanded—a positively sloped demand curve. This situation is called *Giffen*'s *paradox.* It might occur when a staple commodity, such as rice, constitutes a large portion of a consumer's expenditures and has a low price. A price increase would cause a large decline in real income, and if few or no close substitutes existed at prevailing prices, the substitution effect

3. The decomposition discussed here follows J. Hicks; i.e., a price change is accompanied by a compensating income change, which leaves the consumer on the initial indifference curve. A decomposition of the price and income effects proposed by E. Slutsky is somewhat different, since the price change is assumed to be accompanied by an income change that enables the consumer to purchase the original consumption bundle. Johnson et al. (1984: 30f.) provide a graphical and mathematical treatment of the two approaches.

would be overwhelmed by the decline in real income. A good must be an inferior product for Giffen's paradox to occur, but not all inferior goods exhibit Giffen's paradox. Normally, of course, price and the quantity demand have an inverse relationship, and this is the model used throughout this book.

In theory, the magnitude of the change in quantity in response to change in price can be determined from the indifference map and the underlying assumption that consumers maximize utility. In Figure 2-1, the optimum quantity of X that the consumer would purchase at a price of $10 per unit of X was 25 units. When the price declined to $5 per unit, the quantity demanded increased to 45 units (holding the price of Y and income constant). Thus, two points on the demand function for X were derived from the indifference curves; they are plotted in the lower portion of Figure 2-1 as points *u* and *w*. The nature of the demand curve is suggested by the line connecting the two points. Actual intermediate points on the demand function could be determined by rotating the price-ratio line and noting the points of tangency with successive indifference curves. The resulting curve represents the demand for X of an individual consumer in a particular time period; the individual's tastes and preferences, as measured by the indifference curves, are assumed to be constant.

The demand function that we have derived is called the ordinary or Marshallian (after the economist Alfred Marshall) demand curve. For future reference, we note that a compensated or Hicksian (after the economist John Hicks) demand curve also can be derived using the principles discussed above. A consumer's compensated demand function gives the quantities he or she will buy as prices change, but it assumes that the consumer's income is taxed or subsidized so that the consumer's level of utility remains constant. That is, to derive an ordinary demand function, a consumer's utility is maximized subject to a fixed income, while for the compensated demand function, the consumer's expenditures are minimized subject to a fixed level of utility (Okrent and Alston 2011: 6).

*Market demand* is defined as the alternative quantities of a product that all consumers in a particular market are willing to buy as price varies and as all other factors are held constant. A market demand relation can be thought of as a summation of individual demand relations. This includes consumers who enter the market as price declines or who drop out as prices increase. Thus, a change in price influences the number of consumers as well as the quantity each consumes. Of course, since utility functions (indifference maps) of individual consumers are not observable, it is not feasible to build up a market demand curve from aggregating individual utility functions.

Data are often available on total sales and average prices by time periods, such as a month, a quarter, or a year, and such data are often used to estimate market demand functions. Inevitably, with the passage of time, numerous factors that affect demand will change. Thus, strictly speaking, it is impossible to ascertain the true, *ceteris paribus* relationship between quantity demanded and price (for a discussion of the problems, see Davis 1997). Nonetheless, in principle, empirical estimates of demand relations can be useful for policy analysis and forecasting, and the literature contains numerous estimates of retail-level equations.

In deriving theoretical individual consumer demand schedules, prices are treated as given; quantity demanded is thus a function of price. Causation is viewed as running from prices to quantities. However, in practice for market demand relations, prices and quantities may be simultaneously determined, or for agricultural products in particular, causation may run from changes in quantities to changes in prices (see Chapters 5 and 9). In either case, it is valid to specify an inverse demand function, $P = D^{-1}(Q)$, where $P$ equals price and $Q$ equals the quantity demanded. If $P$ and $Q$ are simultaneously determined, it is equally valid to place $P$ or $Q$ to the left of the equal sign. If $Q$ determines $P$, then $P$ should be written on the left-hand side of the function, $P = D^{-1}(Q)$. Inverse demand functions are quite important in agricultural economics. In any case, it is conventional for price to be shown on the vertical axis and for quantity to be shown on the horizontal axis of graphs of demand functions (Figure 2-2).

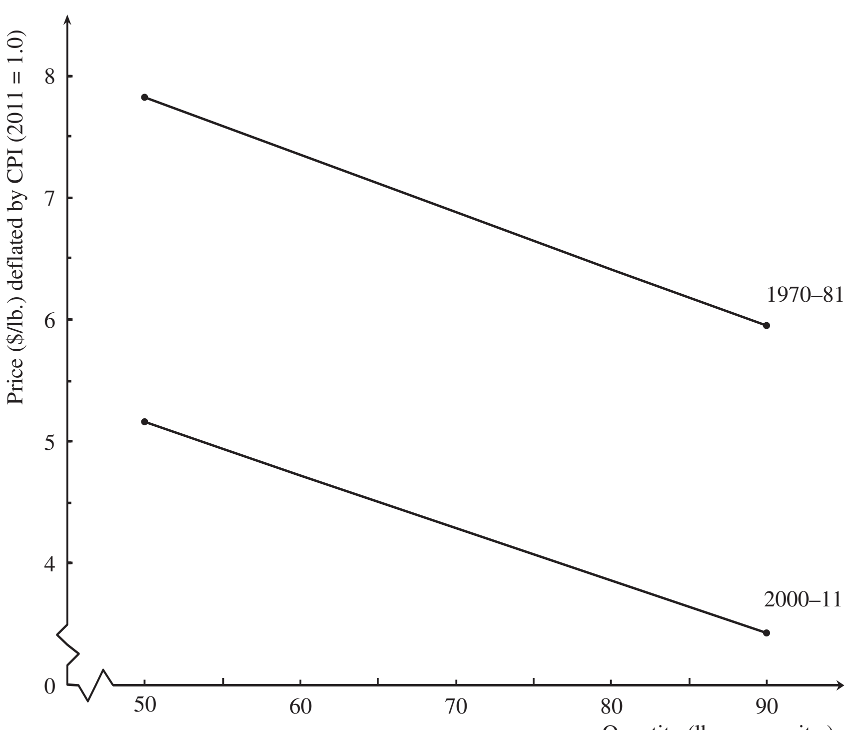



**Figure 2-2.** Estimated retail demand for beef, United States, in two time periods

Estimates of the retail demand for beef in the United States for two time periods are shown in Figure 2-2. The retail price has been deflated by the Consumer Price Index (CPI), and quantity consumed is deflated by the population. The data are grouped by the years 1970–81 and 2000–11 (30 years apart), as shown in the figure, and separate equations were estimated for the two periods.

The results suggest that, for any given level of consumption, the real price of beef was higher in 1970–81 than in the 2000s. That is, the demand for beef decreased over the last 30 years. This decrease in demand for beef is probably due to increasing health concerns about beef's fat content and also to the availability of lower-priced alternatives like chicken. Although the particular numerical results are probably sensitive to the model specification and the sample periods used for the analysis, they illustrate that the level of demand can change with the passage of time. It is also possible that the slope of the relationship can change. This is discussed further below.

## Static and Dynamic Aspects of Demand

The static concept of demand refers to movements along a demand curve; this is called a change in quantity demanded. It is static in the sense that we are looking only at the quantity response to price, and all other factors that may influence demand are assumed constant. With the passage of time, however, other things do not remain constant. Thus, the strictly defined demand curve of economic theory shows how much consumers stand ready to purchase at alternative prices at a particular moment in time. It is also implied that consumers can and will respond instantaneously to a change in price.

The static concept, as just outlined, may seem artificial, but it enables one to think logically about the factors influencing demand and prices. The *ceteris paribus* assumption permits one to ascertain the effect of one variable at a time.

The foregoing discussion implies that the term *dynamic* is used in two ways in demand theory. First, it may refer to changes in demand that are associated with factors that influence the level of demand and that are likely to change with the passage of time. Second, it may refer to lags in adjustment. Quantity adjustments to changes in prices (or price adjustments to changes in quantities, depending on the model) do not take place instantaneously because of the costs of adjustment and the effects of consumers' expectations about the future. The next two subsections elaborate on these dynamic factors.

### Changes in Demand

A change in demand is defined as a shift in the demand curve, as distinguished from movements along a demand curve (i.e., a change in quantity demanded due to a price change). The major factors influencing the level of demand may be grouped as follows:

1. demographic factors, such as population size, its distribution by age, ethnicity, gender, etc.;
2. economic factors, such as income and its distribution, and the prices and availability of other products;
3. consumer tastes and preferences that may be influenced by education levels, life experiences, information and advertising, and the social context in which consumers live (i.e., lifestyle effects).

These factors are sometimes called determinants of demand or demand shifters. Economic theory emphasizes prices of substitutes and complements and income as demand determinants, but for empirical analysis, one may need to consider other factors. As emphasized previously, all these factors are assumed constant for a given level of a demand function.

The effects of changes in the demand determinants may be depicted in two ways, which are best demonstrated by an example. For this purpose, we assume a simple demand function in which quantity ($Q$) is a straight-line function of its price ($P$) and of consumer income ($Y$).

$$Q = \alpha - \beta P + \gamma Y,$$

and $\alpha$, $\beta$, and $\gamma$ are assumed to be fixed parameters that indicate how the variables are related. A graph (demand function) of $Q$ and $P$ can be plotted for a given level of $Y$. If the magnitude of $Y$ changes, then the $P$-$Q$ function shifts to a new level. This illustrates one effect of a change in a demand determinant, a parallel shift in the level of the function. That is, demand increases or decreases as a consequence of a change in the level of the demand determinant, while the parameters $\alpha$, $\beta$, and $\gamma$ remain constant.

The foregoing is a common treatment of the effects of the determinants of demand because, although income may change, consumers' preferences are still assumed not to change. A second possibility is, however, that preferences change in such a way that one or more of the parameters ($\alpha$, $\beta$, $\gamma$), or the functional form, change. This is called a "structural change" in demand. The effect of a price change on quantity may become larger or smaller. Although the slope coefficients depicted in Figure 2-2 are about the same for the two periods, a plausible hypothesis is that consumer preferences for beef have deteriorated in such a way that the effect of a unit change in quantity on price was larger (in absolute value) in the earlier period. That is, it would have not been surprising to find that a given price change had a larger effect on quantity demanded in 2000–11 than in 1970–81.

Conceptually, structural change can be attributed to changes in consumers' preferences. As noted earlier, an individual consumer's demand function is derived from a given set of indifference curves. Thus, as long as these preferences remain unchanged, the relationship among price, income, and quantity also remained unchanged (in the simplified case where these are the only relevant variables). In practice, in attempting to estimate market demand functions, it is difficult to model the combined effects of changes in the many variables that may influence demand, and consequently, it is difficult to determine whether or not structural change has occurred. For instance, as consumers' incomes increase, they may obtain better information about the health effects of alternative diets, their lifestyles may change, etc. Davis (1997) and Tomek and Kaiser (1999), among others, discuss the difficulties of disentangling "parallel" shifts in demand from structural changes in demand.

For the remainder of this chapter, we will refer to changes in demand or shifts in demand. An *increase in demand* means, therefore, that consumers are willing to buy more of the commodity at the same price or that they are willing to buy the same quantity at a higher price. The demand curve has moved to the right. A *decrease in demand* has the opposite effect, a shift to the left.

Increases in demand both for food in the aggregate and for individual food products are closely linked to the rate of population growth. The age distribution of the population also can influence the demand for various products. A teenage population consumes more calories than one made up of people over age 65. Baby food manufacturers gain relative to those selling soft drinks during the early stages of a population boom, but as the population grows older, suppliers of the latter gain relative to the former. In addition, a shift in the demand for commodities like rice, pork, and milk may be related to changes in the ethnic composition of the population.

For most foods, income and demand are positively related; that is, an increase in income shifts demand to the right. For a few commodities, the reverse is likely to be true. Dry beans are perhaps an example of a food for which demand in the United States is likely to decrease as incomes increase. Thus, in principle, the relationship between income and demand can range from positive to zero to negative.

Market demand also depends on the distribution of income among households, and hence a change in income distribution can result in changes in demand. It may be possible, for example, to increase the demand for meats and citrus fruits by transferring income to families near or below the poverty line without changing the total or average level of income. This assumes, as is likely, that very little of the marginal tax dollars collected from upper-income families is used for food, while a substantial portion of the increase in income going to lower-income families (in the form of welfare payments or food stamps) is used to purchase more food or to upgrade diets. Of course, an income redistribution scheme can reduce the demand for certain products, such as high-quality wines.

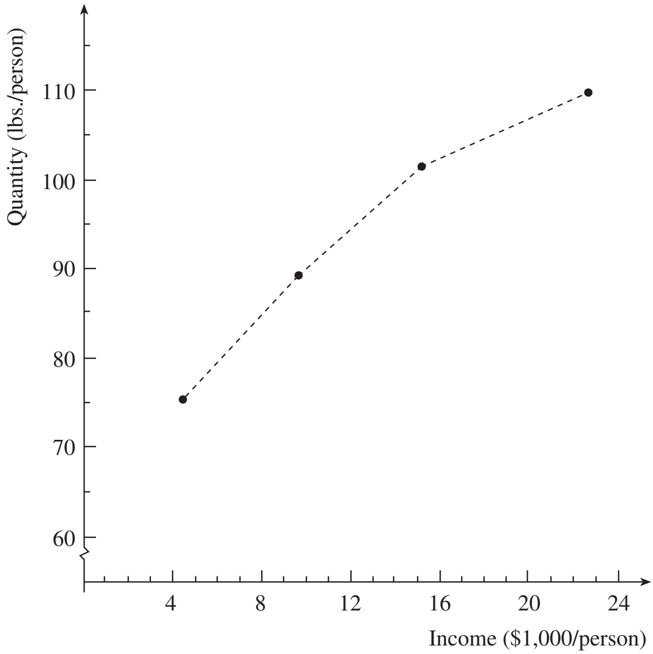



**Figure 2-3.** Average annual household cheese use versus per capita household income, 1987–88 Nationwide Food Consumption Survey, United States. *Source:* Lutz et al. (1992: table 4)

A related idea is that the quantity purchased of a product (other than inferior goods) rises with income but at a decreasing rate; i.e., a curvilinear relationship exists between income and quantity. Total expenditures usually rise even more rapidly because families shift to higher-quality products or buy foods with more built-in services. The relationship between total income and the quantity purchased or the expenditure on a particular food (or commodity group) is sometimes referred to as an Engel curve.[4] Such curves are also called consumption functions; consumption, as measured by quantity or expenditure, is a function of income.

The relationship between per capita household income and per capita consumption of cheese in the United States is shown in Figure 2-3 (Lutz, et al. 1992: table 4). These data were obtained from the 1987–88 Nationwide Food Consumption Survey conducted by the Human Nutrition Information Service,

4. The German statistician Ernst Engel was among the first to undertake empirical studies of the proportion of income spent on food and other items such as clothing and housing as incomes rise. One relationship that he observed has persisted and has become known as Engel's law. Essentially, it states that as consumer incomes increase, the proportion of income spent for food decreases. This implicitly assumes that other things remain constant. With the passage of time, services are added to food, food prices change, etc.

U.S. Department of Agriculture (USDA), and they are not adjusted for other factors that may influence food consumption. Nonetheless, the graph is representative of a relationship between income and the consumption of a "normal" good.

The demand for each commodity is a function not only of its own price but of the prices of other goods. All prices, in theory at least, are linked together in an interdependent system. A change in the price of one product brings about shifts in the demand for other products. The direction of change depends on the direction of change in the price of the related product and whether this product is a substitute or complement. For substitutes, the change in the price of the substitute and the change in demand are usually positively related. (We say *usually* because there are exceptions to the statement, which are discussed in Chapter 3.) If, for instance, the price of beef decreases, then the demand for pork can be expected to decrease. In this case, consumers would partially shift from pork to the relatively cheaper beef, other factors remaining the same. Most agricultural products are substitutes with each other, some much more so than others.

For complements, the change in the price of the related product and the change in demand are usually inversely related. Assuming breakfast cereal and milk are complementary commodities (as they appear to be in the United States), an increase in the price of cereal would decrease the demand for milk (because purchases of cereal decline). The price of cereal and the quantity of milk move in opposite directions. While all prices in an economy are technically interrelated, some products are treated as independent in empirical analyses. Although some people use milk in their coffee, the small quantities of milk used in coffee make it unlikely that a significant relationship can be found between the price of coffee and the quantity of milk.

The development and introduction of new products can be an important shifter of demand, as well as creating structural changes in demand, for agricultural products. Artificial fibers became important substitutes for natural fibers such as cotton and wool. The demand for sugar has been significantly impacted by the introduction on high fructose corn syrup. The introduction of margarine as a substitute for butter, starting in the 1930s, is perhaps the classic example of the effect of the introduction of a new product on the demand for an existing product. The new substitute can not only reduce the demand for the existing product but also create structural change. The introduction of margarine quite probably made the consumption of butter more sensitive to changes in the price of butter than had been the case prior to the existence of margarine; if the price of butter rises, consumers have a viable alternative available.

Changes in preferences obviously contribute to shifts in the demand for agricultural products. As already noted, consumers' preferences are probably influenced by advertising, experience, and education. For example, educational programs about health and nutrition can influence the types of foods purchased.

Also, demand has been shown to be impacted by generic advertising and promotional programs for a host of commodities, such as milk, cheese, fruits and vegetables, and meat products (Kaiser et al., 2005). Since the 1970s, improved knowledge about the health effects of diet has probably reduced the demand for red meats while increasing the demand for poultry, fish, breads, fruits, and vegetables. But, as also suggested earlier, it is difficult in empirical work to isolate the separate effects of variables that are highly correlated with each other. The declining demand for beef, for example, has been variously attributed to (1) the growing concern about the health effects of animal fat; (2) changes in the nature of competition from substitutes, such as chicken; and (3) shifts in the ethnic, age, and income distributions of the U.S. population.

Long-run trends in per capita *consumption* are sometimes used as indicators of changes in preferences. However, they are not necessarily synonymous. For example, an upward trend in chicken consumption is not necessarily dependent on a shift in preferences toward chicken. Growers produced more chicken, in part, because of technological changes, which lowered the costs of chicken production. Thus, the supply function for chicken shifted to the right (supply is discussed in the next chapter), and the consequence has been an upward trend in the availability (hence, the consumption of) chicken. Thus, the increased consumption of chicken may be a consequence of both supply and demand factors and is not necessarily dependent on a shift in preferences.

In sum, the demand for food and fiber products depends on a host of socioeconomic variables. The growing complexity and proliferation of food products and the diversity of changes in socioeconomic variables appear to have made demand analysis even more complex than it used to be. Thus, although much research has been done on the demands for foods and fibers, much remains to be learned about the effects of individual variables on the demands for specific foods (e.g., Tomek and Kaiser 1999).

## Lengths of Run in Demand Analysis

We turn now to a second aspect of the dynamics of demand—the concept of length of run. Simple static theory assumes instantaneous adjustments to price changes. In the real world, however, there are a number of reasons why instantaneous adjustments do not occur; the quantity demanded is likely to change gradually in response to a price change. There are two broad reasons for gradual change: costs of adjustment and the effect of consumers' expectations. The cost of adjustment model assumes that consumers are responding to market prices but that the long-run adjustment to these prices is delayed by the costs of adjustment and, hence, that the long-run quantity is not observable. Costs of adjustment include search costs; consumers are unlikely to learn of price changes immediately. A related factor is the possible habitual patterns of behavior of consumers. It may be impractical (too costly) for a consumer to remake all consumption decisions every time a price changes. Thus, the least-cost solution for some individual consumers may be to delay responses to prices changes, thereby resulting in aggregate adjustments being distributed through time.

Costs can also include technological and institutional barriers to change. For example, if the price of diesel fuel declines relative to the price of gasoline, consumers cannot instantly shift to diesel because they own a stock of gasoline-burning vehicles. Consumers tend to wear out durable goods before replacing them. It is also possible that some consumers' incomes are largely committed at the time the price change occurs. They have installment payments, rent, insurance premiums, and electric and other utility bills to pay. Thus, their discretionary income (i.e., the income remaining after deducting all cash commitments, including debt repayment) may be small, and consequently consumers cannot take advantage of a price change, at least not immediately.

In the expectations model, it is assumed that the current quantity demanded is a function of expected prices, where expectations are dependent on current and past information. Consumers may be uncertain about future price changes, and a change in the current price may influence consumers' expectations about future price changes. Purchases may be postponed or accelerated, depending on consumers' expectations about the future. If a decrease in price is taken as a signal of possible further declines in price, then the immediate effect may be to defer purchases in anticipation of the still lower price. Conversely, if a price increase is assumed to signal further increases in price, purchases may be speeded up. This implies asymmetric responses to price changes (which are more difficult to model).

Expected price is not observable, and a variety of assumptions about expectations formation exist in the literature. These include naive, adaptive, quasi-rational, and rational expectations models. We will have occasion to refer to these concepts later in the book. For now, the point is that consumer (or producer) expectations are not observable and must be modeled.

In demand theory, the concept of *permanent income* is somewhat analogous to that of expected price. The hypothesis is that observable income consists of two components: permanent and transitory income. The transitory component is like a random error term, and consumers' consumption behavior is assumed to depend on the permanent component of their income. Of course, the permanent component of income is unobservable, and like expected price, models must make some assumption about how consumers value their permanent income. A common approach in empirical demand analysis is to introduce lagged dependent variables (e.g., lagged quantity demanded) into the model.

Based on the foregoing types of arguments, a distinction is sometimes made between *short-run* and *long-run* demand. The long run is usually defined as the time required for a complete or total quantity adjustment to occur in response to a permanent price change. The long-run time period corresponds to the adjustment period for each product, but since the time required for adjustment is likely to vary among products, the long-run time period will necessarily differ for different items. The time required to complete the adjustment process is surely greater for durable goods, which are purchased infrequently, than for food products, which are purchased weekly. Indeed, it seems likely that many of the reasons outlined here for lagged adjustments are less applicable to food products than to major purchases of durable products.

The short-run demand schedule is simply the initial effect of the price change. For example, using months as the unit of observation, the short-run effect would be the quantity response this month contemporaneous with the price change this month. Adjustments in subsequent months represent intermediate effects, and as noted, the long-run effect is based on the total quantity adjustment. Logically, the short-run effect of a price change is smaller than the ultimate total (long-run) adjustment to a given price change. (In terms of elasticity coefficients, which will be discussed in Chapter 3, demand is expected to be relatively more own-price elastic in the long run than in the short run.)

The estimation of long-run demand relationships from market data is difficult because prices and other factors that affect demand do not remain constant for a long enough period of time for the full effects of a given combination of variables to work themselves out. Further changes in prices or other variables are likely to occur before the adjustment to the initial price change is completed. Long-run effects are typically modeled via distributed lag models, which are discussed briefly below.

Another complication is that if a price change persists at a relatively high or low level over a long period of time, it may induce structural change. A persistently high price provides an incentive for new substitutes to be developed. Moreover, consumers learn new ways to economize in the use of high-priced products, and even if prices subsequently decline, these economizing techniques are retained. A persistently low price provides an incentive for new uses of the product to arise and/or habits to form that are costly to change.

The effects of changes in energy prices help illustrate the foregoing points. Relatively low gasoline prices provided incentives for consumers to purchase large vehicles with poor gasoline mileage and otherwise encouraged consumption. On the other hand, higher prices provide incentives to conserve more and to buy more fuel-efficient cars. These adjustments would occur gradually, but when adopted, they would likely persist, at least to some degree. Fuel-efficient cars that enter the stock of cars are not discarded when prices decline. Or, if high costs of heating homes induced more families to insulate their homes, this insulation would remain even if prices decreased at some point in the future.

For commodities that can be stored, it may also be important to distinguish between very short-run (daily or weekly) effects and those that occur over longer time periods. We discuss this point in the later section on speculative demand. Before turning to that topic, however, we address the concept of a distributed lag, which is related to the analysis of long-run demand.

### Concept of a Distributed Lag

The idea of a delayed adjustment to a price change leads rather naturally to the concept of a distributed lag model. The lapse of time between a cause and its effect is called a lag, and the effect is likely to be spread through time rather than occurring at a point in time. Hence, the term *distributed lag* arises from a delayed response that is spread over time. In explaining consumer demand, the price change is typically specified as the cause and the quantity change as the effect, but in some cases, interest may center on lagged responses of prices to changes in the availability of the commodity. In evaluating the effects of an advertising program on purchases, it is very likely that the effects will be spread through time.

Given a change in the causal variable (say, price), there are many alternative paths of adjustment that the other variable (quantity demanded) might follow through time. For example, the adjustment in quantity might follow a path where the largest effect occurs initially and then decays, say at a geometric rate. Alternatively, there may be a small initial effect, followed by a larger rate of adjustment, and then a smaller rate of adjustment. One of the problems of empirical analysis is that many alternative paths of adjustment are, in principle, possible. Model specification requires assumptions about the length of the adjustment period and the nature of the path of adjustment. Examples of such models are provided in any modern textbook on econometrics. Although obtaining high-quality estimates of distributed lag models is fraught with difficulties, these models are important because of the likely lags in adjustments that exist in time-series observations (Tomek and Kaiser 1999). This fact is also true for models of supply response, which will be discussed in Chapter 4.

## Speculative Demand

The reader perhaps has thought of demand concepts only in terms of demand by consumers for current use. Speculative demand represents a type of demand related to anticipated use and prices relative to current use and prices. Since numerous agricultural commodities are produced seasonally but are consumed throughout the year, the concept of storage or speculative demand is of particular interest to agricultural economists. Inventory holders, for example, provide the service of carrying stocks from harvest through the marketing year. They expect that price will rise from harvest through the rest of the crop-year by enough to provide them with a profit. That is, the expected price minus the current price is sufficiently large to cover all the costs of storage.[5]

5. Inventories are carried for purposes other than speculation. Processors and other users of commodities hold "working stocks." A stock outage could be very costly for a processor, and clearly livestock producers must have inventories on hand to feed their animals every day. Some minimal level of inventories is needed to avoid the costs of stopping and starting production. Likewise, households carry stocks of food; they do not shop prior to every meal. Stocks provide benefits. A related concept, the convenience yield of inventories, is discussed in Chapter 12.

An analogous idea can be applied to households purchasing storable foods. If a special low price for a product is being advertised, consumers are willing to purchase the product not only for current consumption but also to store it for future use. Consumers do this when the sale price is sufficiently below the expected future prices to justify buying and storing the product for future use. Thus, the quantity responses to sale specials can be large, and a plausible hypothesis is that very short-run responses to price changes can be large relative to the short run. In this case, it is assumed that consumers are aware of the sale price, that they expect the price to return to a higher level after the sale, and that they respond immediately to this price (in contrast to the arguments made in the previous section about costs of search and adjustment). Put another way, the very short-run demand schedule may be defined to include the demand for both current use and for inventories.

Alternatively, one could specify a demand for current consumption and a demand for storage. In this case, the total available supply is allocated between consumption and storage; consumption and storage cannot exceed total supply. Complete models of prices (that include the supply side) will be discussed in subsequent chapters; the point here is that the introduction of the possibility of storage helps link future prices with current prices. On the demand side, additional variables contribute to changes in prices. For instance, the prospect of a small crop for next year would increase speculative demand for current inventories that can be carried into the new crop-year.

The current price of a commodity may, therefore, be strongly influenced by expected future events as well as current economic conditions. Changes in factors that influence expectations tend to occur randomly and hence are unpredictable. For instance, a natural disaster may damage the wheat crop in Australia, and hence the demand for U.S. wheat to be exported in the forthcoming year would increase. The disaster causes an unanticipated rise in expected *and* current prices. These adjustments occur in a matter of minutes, hours, or, at most, days; they are not deferred until the actual exports occur or orders are placed. This occurs because prices are adjusting to discourage current consumption and allocate stocks for future use, in light of the information that production in the next crop year is expected to decline.

Speculation is sometimes viewed only as increasing the amplitude and frequency of price changes. It is allegedly "bad." However, speculation that correctly anticipates future events tends to have a moderating effect on price fluctuations. The purchase of stocks by merchants at harvest time increases prices over the level that otherwise would prevailed. Likewise, the sale of stocks during the year keeps prices below levels that would have prevailed with smaller inventories. In an analogous way, if speculators correctly anticipate a relatively small crop in the next crop year, carrying additional inventory into the new year helps ameliorate the price effects of the small crop. This is sometimes called inter-temporal arbitrage.

Of course, if expectations are not realized, the initial price change was unneeded, and prices return to their original level. But, this "unnecessary" fluctuation was a consequence of the arrival of new information. As noted earlier, changes in information that influence expectations occur randomly; they are not predictable. This is not the fault of the merchant-speculator. Indeed, one of the costs of carrying inventories is the risk of unfavorable price changes. Thus, it is important in appraising price behavior to discriminate between useful speculative carrying of inventories and possible illegal or anti-competitive behavior by merchants.

In sum, total demand is influenced by speculative demand, and speculative demand introduces the probability of current prices being influenced by expected future events. Current prices can change because of changes in expectations.

## Derived Demand

The ultimate consumer determines the shape and position of the demand function for a final product. For this reason, consumer demand relations are often referred to as *primary demand*. In empirical analyses, primary demand functions are estimated from retail price and quantity data.

The term *derived demand* is used to denote demand schedules for inputs that are used to produce the final products. This is an especially important concept for agricultural commodities because their demands are derived from many uses.

For example, corn (maize) produced in the United States is used for animal feed; manufacturing sweeteners, alcohol and other products; exports; and inventories. Hence, the farm-level demand for corn depends on a host of factors. Variables influencing the demand for exports include the incomes and population levels in other countries, as well as the availability of substitutes and the level of exchange rates. The farm price of corn can be influenced by the rate of economic growth (or the lack thereof) in importing countries, such as China, as well as by their domestic production. Another use of corn is to produce ethanol, which is blended with gasoline. Thus, the derived demand for corn also depends on the price of petroleum, which in turn depends on supply and demand conditions in the petroleum market.

Wheat, like corn, has a variety of end uses. Many countries are important suppliers of wheat of varying types. As we will discuss in Chapter 7, the different varieties of wheat have different characteristics (e.g., varying levels of protein). Hard red winter and spring wheats produced in North America are processed into flour that, in turn, is used to make a variety of breads. The general point is that the derived demand for a particular lot of a product will also depend on its attributes.

In sum, the demand for commodities at the farm level is derived from the demand for the many different uses of the commodity; they are inputs in the production of many intermediate and final products. Moreover, these uses have become more varied and complex as the economy has grown. Analyzing the derived demands for agricultural commodities is, therefore, increasingly difficult.

The concept of derived demand has broad applications. As we will see in Chapter 4, the demands for inputs used in farming, such as seed, fertilizer, and land, are derived from the demand for the commodities which farmers produce. The demand for a retail product, like that for a farm commodity, may be viewed as derived from the demand for its characteristics (Lancaster 1971). For example, a feed manufacturer wants to know the least-cost set of ingredients to make, say, a dairy feed with a particular protein, energy, and fiber content. Thus, the manufacturer is interested in the protein content of the alternative commodities that might be used to make the feed. Commodities can also have undesirable attributes, such as excess moisture or pesticide residues, that detract from their demand and lower their price. A genetically modified (GM) commodity is considered to be undesirable by some, but not all, buyers.

The idea of households demanding attributes rather than products may be less apparent, but Lancaster (1971) and others point out that households should be treated both as producing and consuming units. The consumer is buying foods to produce meals that are nutritious, have variety, and are tasty. Some consumers are demanding products that are organically or "naturally" grown and/or are not produced from GM commodities, while other consumers are not concerned about these kinds of attributes. Hence, the demand for food products may be viewed as derived from the demand for a large variety of characteristics, which are of varying importance to different consumers. A literature has developed to estimate the value of such characteristics, often using experimental economics approaches. This topic is discussed in Chapter 7.

In empirical modeling, there are two basic approaches to obtaining derived demand functions. One of these is discussed in Chapter 4, based on traditional microeconomic theory, which assumes that a firm's output is a continuous function of the various inputs. Then, assuming competitive, profit-maximizing firms, it can be shown that the derived demand for an input (say, wheat) is a function of the price of the input (wheat) relative to the price of the output (flour and bread).

In the second approach, the derived demand for the farm commodity differs from primary demand for the final product by the amount of the per unit marketing and processing costs. The demand for meat animals at the farm, for example, is based on the retail-level demand function for meat, minus marketing costs such as slaughtering, processing, transporting, and retailing meat. This approach, as discussed in Chapter 6, makes the implicit assumption that the final product is produced in a fixed proportion to the quantity of commodity input. For example, it is assumed that 1.7 pounds of live hog is required to obtain 1 pound of pork at retail. The transformation coefficient is exactly 0.588. With this assumption and others to be discussed, the derived demand function depends on the same shift variables as the primary function, but the derived demand equation also should logically contain a variable (or variables) to account for marketing costs. That is, derived demand can change because the primary function changed and because marketing costs changed.

In discussing agricultural product prices, we are concerned with both retail-level and farm-level demand functions. It is unquestionably the case that modeling these functions is becoming increasingly difficult.

## References

Davis, L. G. 1997. "The Logic of Testing Structural Change in Meat Demand: A Methodological Analysis and Appraisal," *Am. J. Ag. Econ.* 79:1186–1192.

Johnson, S. R., Z. A. Hassan, and R. D. Green. 1984. *Demand Systems Estimation*. Ames, Iowa: Iowa State Univ. Press. Chapter 2.

Kaiser, H. M., J. M. Alston, J. M. Crespi, and R. J. Sexton, Editors. 2005. *The Economics of Commodity Promotion Programs: Lessons from California*. New York: Peter Lang Publishing.

Lancaster, K. 1971. *Consumer Demand: A New Approach*. New York: Columbia Univ. Press.

Lutz, S. M., D. M. Smallwood, J. R. Blaylock, and M. Y. Hama. 1992. *Changes in Food Consumption and Expenditures in American Households During the 1980's*. Econ. Res. Ser. USDA Stat. Bul. 849.

Okrent, A.M., and J. M. Alston. 2011. *Demand for Food in the United States*. Univ. of California Giannini Foundation Monograph 48.

Tomek, W. G., and H. M. Kaiser. 1999. "On Improving Econometric Analyses of Generic Advertising Impacts," *Agribusiness* 15:485–500.

Varian, H. R. 2010. *Intermediate Microeconomics, a Modern Approach*, 8th ed. New York: W.W. Norton & Co.

CHAPTER 3

# Demand Elasticities and Related Coefficients

This chapter reviews the concepts of own-price, cross-price, and income elasticities of demand. Interrelationships among these coefficients, as suggested by the underlying demand theory, are described. The concepts of *total elasticity* and *flexibility* coefficients are also introduced.

## Price Elasticity

### Definition

The concept of a demand function or a demand schedule was discussed in Chapter 2. It provides a description of the relationship between price and the quantity buyers are willing and able to buy, other factors held constant. Price theory suggests an inverse relationship between price and quantity, but the inverse relationship by itself says nothing about the responsiveness of quantity demanded to a price change. This responsiveness is likely to vary from product to product. It is logical to think that the quantity of salt purchased is not very responsive to changes in the price of salt, while on the other hand, the quantity of grapes purchased is relatively more responsive to a change in the price of grapes. From the perspective of consumers, grapes have more substitutes than salt.

An explicit demand curve can be defined by an algebraic equation, which in practice must be estimated from available data. The quantity variable is normally expressed in physical units, while price is expressed in monetary terms per physical unit. But since different units of measurement are often employed (bushels, pounds, kilograms), it is difficult to make direct comparisons from the equations of the impact a given change in price will have on different products. To facilitate comparisons, economists frequently make use of percentage relationships, which are independent of the size of units used to measure price and quantity. The most common of these percentage relationships is the own-price elasticity of demand, which is a ratio that expresses the percentage change in quantity demanded associated with a given percentage change in price.

Price elasticity is defined for a point on the demand curve, and hence for most demand functions, the magnitude of the elasticity coefficient varies along the function. Let *d* equal a very small change, then a mathematical definition of elasticity for product *i* is

$$E_{ii} = (dQ_i / Q_i)/(dP_i / P_i).$$

This equation is merely a way to state the percentage change in $Q$ relative to the percentage change in $P$, evaluated at a point $P_i$,$Q_i$. Note, the elasticity definition can be rewritten as $(dQ_i / dP_i)(P_i / Q_i)$, where $dQ_i/dP_i$ is defined as the first derivative of the demand function $Q_i = D_i(P_i)$. Thus, in principle, if one knows $dQ_i/dP_i$, then the elasticity can be evaluated at particular values of $P_i$ and $Q_i$ that are on the function. In empirical analyses, a common point to use is the arithmetic means of $P_i$ and $Q_i$.

An alternative equation for defining price elasticity is the arc or average formula

$$\begin{aligned} E &= [(Q_0 - Q_1)/(Q_0 + Q_1)]/[(P_0 - P_1)/(P_0 + P_1]\\ &= [(Q_0 - Q_1)/(Q_0 + Q_1)][(P_0 + P_1)/(P_0 - P_1)]. \end{aligned}$$

The subscripts now represent two different points on a demand curve. The arc equation is mainly a device for computing an elasticity at an average between two points—not the average of the elasticities on the arc between the points. The smaller the arc or segment, the more nearly the elasticities computed from the arc and point formulas approach each other. Remember, elasticity is strictly defined only with respect to a particular point and, hence, will change as the point is varied.

## Interpretation

The own-price elasticity-of-demand coefficient for any product can be interpreted as the percentage change in quantity demanded given a very small percentage change in the price of that product, other factors held constant. A convenient way to think of a price elasticity is as the percentage change in quantity corresponding to a 1 percent change in price. Since the slopes of demand curves are negative, price elasticity-of-demand coefficients have a negative sign.

The range of the price-elasticity coefficient is from zero to minus infinity. This range is divided into three parts. (1) If the absolute value (i.e., neglecting the sign) of the coefficient is larger than 1, demand is said to be *elastic*. The percentage change in quantity demanded is greater than the corresponding percentage change in price. The extreme case is a horizontal demand curve—

demand is perfectly elastic (the coefficient is infinite). (2) If the absolute value of the coefficient is less than 1, demand is *inelastic*. The percentage change in quantity is less than the corresponding percentage change in price. Quantity demanded is relatively unresponsive to price changes. The extreme case is an elasticity of zero—demand is perfectly inelastic. (3) A coefficient of exactly −1 represents the case of *unitary elasticity*. The percentage change in quantity equals the percentage change in price.

As already noted, the elasticity coefficient varies along the demand curve for most functional forms. If a straight-line demand function is extended to the two axes of the graph, the elasticity varies from infinity on the price axis through various (negative) values to zero at the point on the quantity axis, with unity elasticity at the midpoint. This may be verified by referring to the definition of price elasticity (point equation) and noting that price is zero when the demand function intersects the quantity axis and that quantity is zero when the function intersects the price axis. A few special cases exist where the elasticity is a constant over the range of the curve. These cases include a straight horizontal line, a straight vertical line, a power function, and a rectangular hyperbola.[1]

Since, in general, the elasticity coefficient varies in magnitude along the demand curve, it is not technically correct to say that the demand for a good is elastic or inelastic. Demand is elastic (or inelastic) only within some range of prices. However, it is convenient to categorize and speak of the demand for a product as being either elastic or inelastic. This shorthand reference should be interpreted as referring to the elasticity within the usual range of prices. Relatedly, since elasticity is defined for a very small change in price, it should not be used to predict changes in quantity demanded for a large change in price since the elasticity, itself, would likely change over a large price change.

## Price Elasticity and Total Revenue

Total revenue equals price multiplied by quantity and consequently has two components, price and quantity. Assuming a stable demand curve, price and quantity vary inversely as we move along the function, and it is not obvious how changes in price will influence total revenue. For example, the question whether a given percentage increase in price will increase or decrease total revenue depends on the magnitude of the corresponding percentage change in quantity. The question is answered by the magnitude of the price elasticity-of-demand coefficient.

If demand is elastic in the relevant range of prices, then price and total revenue vary inversely. A price increase will decrease total revenue, and a price decrease will increase total revenue. This follows from the definition of an elastic demand, which is that the percentage change in quantity demanded is

1. Letting $Q$ = quantity demanded and $P$ = price, the power function is $Q = aP^b$, where $a$ and $b$ are parameters of the equation. A rectangular or equilateral hyperbola is $QP = b$, or $Q = b(1/P)$; in this case $E = -1$.

greater than the percentage change in price. A decrease in price, for example, results in a more than offsetting percentage increase in quantity demanded. However, it does not follow that total revenue will increase indefinitely as price decreases. At some point, price will presumably move into an inelastic range of the demand relation.

If demand is inelastic in the relevant range of prices, then price and total revenue vary directly. A price increase will increase total revenue and vice versa. The reader should reason out why this is true based on the definition of an inelastic demand.[2]

The derived demands for most, if not all, farm commodities are thought to be price inelastic. Thus, price and total revenue are expected to be directly related or, put another way, an increase in output, with other factors held constant, will reduce prices and the total revenue obtained by farmers. This was recognized at least as early as 1915 when Henry A. Wallace wrote, "The Demand Laws . . . indicate to me that the farming class as a whole is penalized by over-production and rewarded for under-production" (as cited in Stigler 1962: 17). Governments have sometimes undertaken farm policy measures to limit supply, implicitly assuming that the demand for the commodity is inelastic. Otherwise, reducing output would lower the total revenue received by farmers. (Reducing production also reduces total costs.)

The price-elasticity concept measures responsiveness *along* a demand curve. This, of course, is implied in the discussion of elasticities and total revenue. Shifts in demand (relative to supply) also influence revenue. If, for example, demand increases, then for a given price total revenue and quantity increase even though demand is inelastic at that price. This is a consequence of the shift in demand and not of the elasticity evaluated on one function. This point is illustrated in Figure 3-1 for the apple crop in the United States. The observations for the period 2006–11 are up and to the right of the observations for the years 2000–05, but within each time period, total revenue (deflated by a measure of the general price level) is inversely related with production. It appears that the aggregate demand for apples has increased, but that the demand for apples at the farm level is price inelastic.[3]

A related point is that it is difficult to compare elasticities in two different time periods. Sometimes researchers state that demand is more (or less) elastic than it use to be. But, with the passage of time, the demand function likely has

2. Another way of discussing this point is in terms of the marginal revenue function. Total revenue, $R$, equals price multiplied by quantity, $PQ$. Recalling the demand function $Q = D(P)$, $R = QP = D(P)P$. Marginal revenue, then, is the first derivative $dR/dP$. When price is in the inelastic range of the demand equation, marginal revenue is negative, and when price is in the elastic range, marginal revenue is positive.

3. It perhaps should be emphasized that the figure relates to the total revenue from all sales of apples. Apples have many uses: for fresh fruit, for canning and freezing, and for juice. The demand functions for the individual uses differ from the aggregate function, and thus the elasticities for individual uses can differ from the elasticity related to aggregate demand.

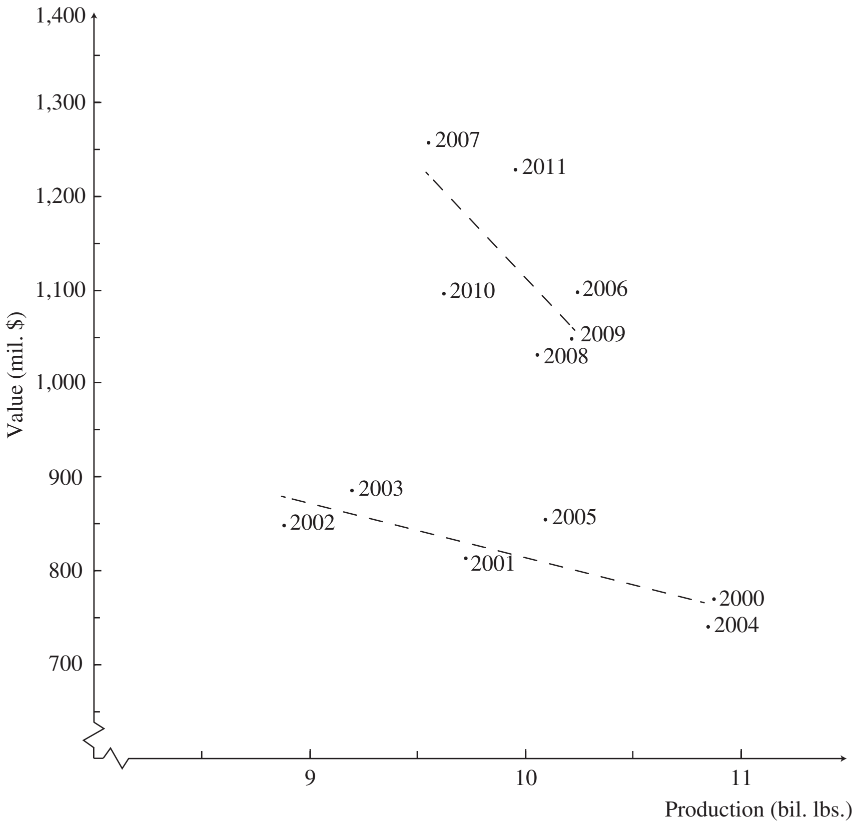



**Figure 3-1.** Apples: production and value of production, United States, 2000–11 crop years

shifted levels. Thus, the comparisons involve elasticities obtained from two different functions, i.e., estimated for two different pairs of prices and quantities. Researchers need to be precise when discussing the comparisons of coefficients estimated for different time periods. Is the change in elasticity due to a shift in demand, a change in the slope of the demand function, a shift in supply along a demand function, or some combination of factors?

## Income Elasticity

Income elasticity of demand is a measure of the responsiveness of quantity to changes in income, other factors held constant. The income-quantity relationship, of course, can be expressed algebraically. As indicated in Chapter 2, this relation is sometimes called a consumption, or Engel, function. The income elasticity is defined as a point on the function and typically varies along the range of the curve.

Let $Y$ equal income and $d$ equal a small change; then the definition of income elasticity at a point is

$$E_{iy} = (dQ_i/Q_i)/(dY/Y) = (dQ_i/dY)(Y/Q_i).$$

Also, if the consumption function is $Q_i = f(Y)$, $dQ_i/dY$ is the first derivative of $Q_i$ with respect to $Y$. The income elasticity may be interpreted as the percentage change in quantity corresponding to a 1 percent change in income, other factors held constant.

In most cases, the elasticity coefficient is positive. As discussed in Chapter 2, as income increases, consumers are expected to buy more of most products, and when income decreases, the opposite occurs. The income elasticity for food in the aggregate is perhaps 0.2. A few foods probably have negative income elasticities at the average level of incomes in the United States and other high-income countries, but whether a particular commodity has a positive or negative income elasticity is a question that must be answered empirically.

For individual foods, at least those with relatively few built-in services, income elasticities are thought to decline as incomes increase. Households with high incomes will generally have smaller income elasticities for foods than households with low incomes. That is, as incomes increase, it seems logical that the marginal effect of $Y$ on $Q_i$ ($dQ_i/dY$) will decrease; the function $Q_i = f(Y)$ is probably non-linear. Because income elasticities are used in making demand projections (i.e., to estimate the effect of increasing income on the demands for specific products) and because the elasticity itself can change as incomes increase, the researcher must exercise great caution in making projections from a single coefficient. Also, since $Q_i = f(Y)$ is non-linear, the choice of a functional form in empirical analyses will influence forecasts, especially if $Q_i$ is estimated from a level of $Y$ that is outside the range of the historical experience.

In empirical analyses, "income elasticities" are sometimes estimated from observations on expenditures rather than from observations on physical quantities and incomes. Expenditures on a particular product may be a function of total expenditures. When data are obtained from a survey of households, the reported household income may contain errors. The respondent may overlook income earned by another member of the household or forget non-wage income. In time-series analysis, it is relatively easy to obtain estimates of total expenditures on food compared to obtaining the total physical quantity of food purchased.

When expenditures are used, the elasticity represents the percentage response of expenditure on the individual product to a 1 percent change in total expenditures. Coefficients that measure the responsiveness of expenditures are sometimes called *expenditure elasticities.* These elasticities are generally larger than those based on physical quantities; i.e., expenditures are usually more responsive than quantities to a given change in income. Consumers with high incomes

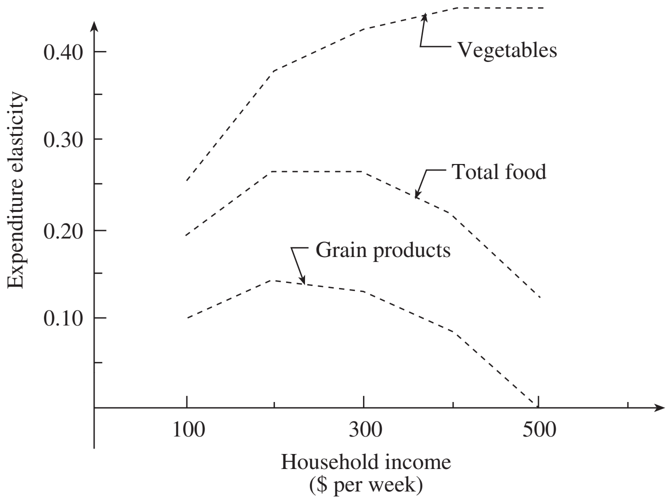



**Figure 3-2.** Estimated expenditure elasticities by household income level, United States, 1965. Elasticities are computed from coefficients presented by Salathe and Buse (1979: table 3)

are likely to buy products of higher quality and with more services attached than are consumers with low incomes. These higher-quality products, with more services, usually cost more. Thus, the expenditure change in response to an income change includes a quality-service effect as well as a quantity effect.

Estimated expenditure elasticities for two commodity groups plus total food are plotted against household income in Figure 3-2. Although based on older data, this figure illustrates how elasticities may vary with income levels; that is, the expenditure elasticities for food tend to rise and then decline as income grows. The particular shape of the relationships, however, is related to the quadratic functional form used by the analysts (Salathe and Buse 1979).

## Cross-Price Elasticity

Cross-price elasticities of demand are measures of how the quantity purchased of one product responds to changes in the price of another product. More precisely, the cross elasticity of product *i* with respect to product *j* is defined as

$$E_{ij} = (dQ_i / Q_i)/(dP_j / P_j) = (dQ_i / dP_j)(P_j / Q_i).$$

This may be interpreted as the percentage change in the quantity of *i* given a 1 percent change in the price of *j*, other factors held constant.

In practice, three types of relationships can be identified: the products are substitutes, complements, or independent. The definition of the three categories is based on the *substitution effect* of the price change of *j*. The substitution effect is positive for substitute commodities; the price of *j* and the quantity of *i* move in the same direction. If the price of *j* increases, then consumers tend to substitute *i* for *j*. If the price of *j* decreases, then consumers tend to substitute the relatively cheaper *j* for *i*.

The substitution effect is negative for complementary commodities such as milk and breakfast cereal. In this case, the price of *j* and the quantity of *i* move in opposite directions. An increase in the price of *j* (cereal) means that the quantity demand of *j* decreases and hence the quantity of the complementary commodity *i* (milk) also decreases. The symmetrical argument can be made for increases in the price of *j*.

The substitution effect is zero for independent commodities. Independence means that no substitution or complementary relationship exists between the two commodities, but as we will see, the price of *j* and the quantity *i* may not be truly independent because of the *income effect* of a change in the price of *j*.

On the basis of the foregoing reasoning, economists generally say that substitute commodities have positive cross elasticities, complementary commodities have negative cross elasticities, and independent commodities have zero cross elasticities. However, because of the potential income effect of a change in the price of *j,* these generalizations need not be true. The income effect on the demand for *i* is generally, but not always, negative for cross elasticities. A decrease in price, other factors held constant, increases real income and hence will increase quantities purchased (unless the product is an inferior good). An increase in price decreases real income and has the effect of decreasing quantity.

If the income effect outweighs the substitution effect, there will be a net reduction in the demand for commodity *i* when the price of *j* increases. Consumers will normally substitute *i* for *j* when the price of *j* increases; but an increase in the price of *j* is equivalent to a reduction in real income. This will adversely affect consumption of both *i* and *j*. Thus, the real income effect on consumption of *i* will be negative, while the substitution effect will be positive. If the former exceeds the latter, the net effect may be negative even though the two products are substitutes.

The income effect usually reinforces the substitution effect for complementary goods. However, two commodities could be independent from the viewpoint of substitution and still have a negative cross elasticity from the income effect of the price change. The interpretation of cross elasticities is further complicated by the fact that the income effect is not always inversely related to price. There are inferior commodities, which means some commodities have negative income elasticities. This implies a positive income effect that would reinforce or add to the substitution effect for some commodities.

The importance of the income effect depends on the size of expenditure on the commodity relative to total expenditures. Typically, the expenditure on one product is a small fraction of total expenditures, and hence the income effect does not outweigh the substitution effect. Thus, the generalization that economists make about the signs of cross elasticities usually hold—a positive cross elasticity implies that the products are substitutes, while a negative cross elasticity implies the commodities are complements.[4] We have discussed the topic in some depth, however, because in low-income countries where one food, like rice, makes up a large proportion of total expenditures, the income effect of a change in the price of that commodity can be large.

Based on logical reasoning, butter and margarine are examples of substitutes, and hot dogs (wieners) and hot-dog buns (rolls) are complementary. In practice, economists have found the measurement of cross elasticities to often be difficult.

Reversing the products in the cross-elasticity equation does not necessarily give the same coefficient; indeed, most of the time it will not. The cross elasticity of sugar with respect to coffee is likely not the same as the cross elasticity of coffee with respect to sugar. A change in the price of coffee will have some effect on the consumption of coffee and hence a small effect on the demand for sugar. But a change in the price of sugar probably will have no measurable influence on the demand for coffee.

The explicit relationship between any pair of cross elasticities can be defined mathematically. The substitution effect (whether for substitutes or complements) is symmetrical, but the income effect is not. The exact relationship is given in the next section.

## Relationships among Elasticities

### Consumer Demand

There is a lot of research devoted to the theoretical relationships among elasticities and to the applications of these concepts to the estimation of elasticities for food products. Brandow (1961) and George and King (1971) are early classic examples of empirical applications, and a huge literature has developed out of these early beginnings. Okrent and Alston (2011) provide a review of this literature and present new estimates of demand elasticities for food products.

The important relationships among demand elasticities include the homogeneity condition, the symmetry condition, and the Cournot and Engel aggregation conditions; each of them is discussed in this section. The theory from which these relationships are derived makes certain assumptions about consumer behavior. Thus, the elasticity conditions hold for an individual consumer with a given utility function, which satisfies these assumptions, including the assumption that the individual's tastes and preferences are "reasonable." A detailed

4. Estimated elasticities are subject to sampling error and perhaps to biases from errors in specifying the statistical model (see final section of this chapter). Thus, users of empirical results should be cautious in interpreting negative cross-price elasticities as truly representing complementary products. Most products are substitutes.

discussion of the underlying axioms, which define "reasonableness," is beyond the scope of this book. Roughly, the consumer is assumed to be able to rank commodity bundles in order of preference, and the ranking must avoid ambiguity and be consistent (transitivity). Moreover, the consumer is assumed to not be satiated, preferring more to less. A further assumption is required about the nature of the utility function to assure that the usual conditions for a constrained maximum (from the calculus) are necessary and sufficient. More rigorous discussions are available in a variety of sources (e.g., Deaton and Muellbauer 1980; Capps and Havlicek 1987).

*Homogeneity Condition.* In discussing the restrictions on a system of elasticities, it is convenient to think of a matrix of elasticities that has $i = 1, 2, \ldots, n$ rows and $j = 1, 2, \ldots, n + 1$ columns, where the last column will represent income, $y$. In other words, the economy has a total of $n$ goods and services. Then, Equation (1) is written for the $i$th row and states that the sum of the own- and cross-price elasticities and the income elasticity for a particular ($i$th) good is zero.

$$(1) \qquad E_{i1} + E_{i2} + \cdots + E_{ii} + \cdots + E_{in} + E_{iy} = 0,$$

where $E_{ii}$ is the own-price elasticity for the $i$th good, $E_{ij}$'s are the cross-price elasticities, and $E_{iy}$ is the income elasticity. This constraint holds for all $n$ rows (goods).

The homogeneity condition means that the substitution effect and the income effect of an own-price change must be consistent with the cross and income elasticities for the product. A large income elasticity tends to imply a large (in absolute value) own-price elasticity. A large number of substitutes and/or a strong substitute also suggests a relatively large own-price elasticity for the product. If in Equation (1) the cross elasticities are collectively large and positive and given that the income elasticity is positive, then the own-price elasticity must necessarily be a large (in absolute value) negative number. The reverse reasoning also holds. Salt has commonly been used as an example of a product with a price-inelastic demand. The intuitive argument is that it has few substitutes and has a small income elasticity. If the cross elasticities and income elasticities are near zero, then the own-price elasticity will be near zero, although negative (i.e., inelastic).

The following illustration provides estimated demand elasticities for beef ($i$ = beef). The estimates are based on annual, retail-level data in the United States (Huang 1993).

| | |
|---|---|
| beef own-price elasticity | –0.62 |
| cross-price with pork | 0.11 |
| cross-price with chicken | 0.02 |

| | |
|---|---|
| all other cross elasticities | 0.10 |
| income (expenditure) elasticity | 0.39 |
| sum | 0 |

The homogeneity condition combined with two reasonable assumptions suggests that the absolute value of the own-price elasticity is likely to be greater than the values of the cross-price elasticities. The assumptions are (1) that the income elasticity is positive and (2) that most of the cross relationships are substitutes and hence the cross elasticities are mostly positive numbers. Consequently, for the sum of the elasticities to be zero, the commodities own-price elasticity must be a large (in absolute value) negative number.

The homogeneity condition also can be used to set a lower limit on the price elasticity and an upper limit on the income elasticity. We again use the reasonable assumption that the sum of the cross elasticities is positive and that the lower limit of this sum is zero. If the sum of the cross elasticities were about zero, then $|E_{ii}|$ approximately equals $|E_{iy}|$. If the sum of the cross elasticities were positive, then the price elasticity would be larger (in absolute value) than the income elasticity. Thus, the magnitude of the income elasticity tends to set the lower limit of the own-price elasticity. In a similar way, we can argue that the price elasticity, if known, sets an upper limit for the product's income elasticity.

For example, suppose previous research indicates that the income elasticity for food grains in India is 0.7; then the price elasticity is implied to be –0.7 or smaller and, because of substitutes, is probably less inelastic (more elastic). Thus, the estimate of one elasticity can provide a guide to another elasticity. The general point is that elasticities should have a logical relationship to one another.

*Symmetry Condition.* As noted in the previous section, cross-elasticity coefficients are not symmetric in the sense that $E_{ij}$ equals $E_{ji}$. Rather, the symmetry between a pair of cross elasticities is appropriately defined by Equation (2).

$$(2) \qquad E_{ij} = (R_j/R_i)E_{ji} + R_j(E_{jy} - E_{iy}),$$

where $R_i$ is the expenditure on $i$ as a proportion of total expenditures; $R_j$ is the expenditure on $j$ as a proportion of total expenditures; $E_{ij}$, $E_{ji}$ are the cross-price elasticities; and $E_{iy}$, $E_{jy}$ are the income elasticities.

Note, the last term in Equation (2) will be approximately zero if the two income elasticities are about equal and/or the expenditure share for product $j$ is small. Thus, the following equality may hold approximately:

$$(2') \qquad E_{ij} = (R_j/R_i)E_{ji}.$$

This equation has been called the Hotelling-Jureen relation; it is as an approximation of the symmetry relation.

The substitution effect of a price change is in a sense symmetric, but the income effect is not.[5] The income effect of a price change will be larger for the product that takes a larger proportion of total expenditures. We know, for example, that in the United States a much smaller proportion of the average consumer's income is spent on lamb than beef. Thus, a 1 percent change in the price of lamb has a smaller effect on the consumption of beef (smaller cross elasticity) than the reverse. To illustrate this point, we assume that the average consumer spends 2 percent of total expenditures on beef, b, and 0.1 percent on lamb, a. That is, $R_{\mathrm{a}} = 0.001$, and assuming that the beef and lamb have roughly similar income elasticities, the last term in Equation (2) will be about zero. Now, suppose the cross elasticity of lamb with respect to beef ($E_{\mathrm{ab}}$) is 0.6; then, from Equation (2′)

$$E_{\mathrm{ba}} = (0.001/0.02)(0.6) = 0.03.$$

Given the assumed conditions, a 1 percent change in the price of lamb will result in only a 0.03 percent change in the demand for beef, and this is true even though a 1 percent change in the price of beef will result in a 0.6 percent change in the demand for lamb.

The symmetry condition, combined with other restrictions on the admissible values of elasticities, is used to help estimate cross elasticities in demand systems. The restrictions help make the empirical estimates consistent with economic principles. Moreover, a demand system has a very large number of coefficients to estimate, and restrictions on the model help reduce the degrees-of-freedom problem. That is, in an unrestricted system, the number of parameters to be estimated is large relative to the number of observations (sample size); the restrictions reduce the number of coefficients to be estimated.

*Engel and Cournot Aggregation Conditions.* The Engel aggregation condition is defined by Equation (3), which means the weighted sum of the income elasticities for all *n* items in a consumer's budget is 1, where the weights are the expenditure shares, the expenditure on the *i*th product relative to total expenditures ($R_i$).

5. The absolute (not percentage) response of the quantity of *i* to a price change in *j* is defined by the partial derivative of $Q_i$ with respect to $P_j$. This partial derivative can be divided into the substitution effect and the income effect. The partial derivative of $Q_j$ with respect to $P_j$ can be partitioned in an analogous way. The symmetry is between the substitution-effect components of the two partial derivatives; using the letter *K* to represent the substitution effects, $K_{ij} = K_{ji}$. But, as discussed in Chapter 2, a change in price also has an income effect. There, we also made the distinction between uncompensated and compensated elasticities, following the approach of Hicks. A similar approach is to consider the partial derivative of $Q_i$ with respect to $P_i$ and partition it into a price and income effect; the combined effects is the uncompensated change in $Q_i$. If one adjusts for the income effect, i.e., consider only the substitution effect $K_{ii}$, this gives the compensated effect of the own-price change. This definition implicitly follows the approach of E. Slutsky, rather than Hicks, but in the limit, the two approaches give the same result. Uncompensated and compensated price elasticities have analogous definitions.

$$(3) \qquad R_1E_{1y} + R_2E_{2y} + \cdots + R_nE_{ny} = 1.$$

Equation (3) should not be interpreted as meaning that all income elasticities must be small. The weights are proportions, which must be less than 1, and indeed may be tiny. A hypothetical example for $n = 3$ with assumed income elasticities of 5, 1, and 0.2 illustrates the point: (0.1)(5) + (0.4)(1) + (0.5)(0.2) = 1. Note, the expenditure shares sum to 1.

The Cournot aggregation condition is defined by Equation (4) and applies to weighted price elasticities in each column, $j$.

$$(4) \qquad R_1E_{1j} + R_2E_{2j} + \cdots + R_nE_{nj} = -R_j.$$

That is, for the $j$th product (column), the price elasticities weighted by their respective expenditure shares must equal the negative of the expenditure share for that ($j$th) product. The practical import of this result is in the estimation of demand systems, to which we now turn.

*Applications.* Estimates of market demand elasticities are important for policy analyses. Procedures for obtaining such estimates fall into two broad categories. The first is the specification and statistical estimation of demand equations for individual commodities without the imposition of the constraints, just discussed. In theory, the demand for a particular good depends on the prices of all of the goods in the economy, but in practice many prices are ignored. This approach is ad hoc in the sense that judgment is required to specify the prices that will be excluded from the demand equation of interest. Often, such demand equations are estimated in the context of modeling markets, i.e., including supply equations, inventory equations, etc. The emphasis in such research is on a particular market and perhaps on the relationship of that market to other closely allied markets.

The second category of analysis is the estimation of a system of demand equations, where the restrictions derived from demand theory are formally imposed on the system. In this case, the research emphasis is on the interrelationships among demands for the $n$ products (or product groups) included in the system. The theoretical constraints provide a formal structure for the interdependencies among the products.[6] The resulting literature involves the statistical estimation of a system of demand equations, still requiring considerable judgment in model

6. In an economy with $n$ goods, there will be $n$ own- and cross-price elasticities plus an income elasticity for each good. For example, if $n = 50$, then 2,550 coefficients must be estimated [$n(n + 1) = 50 \times 51$]. The restrictions help reduce the number of parameters to be estimated, but a degrees-of-freedom problem will often still exist. That is, the number of coefficients to be estimated in the statistical model is large relative to the number of observations. A common approach to making estimation manageable is to work with sub-systems, e.g., a sub-system of meat demand equations, and the selection of sub-systems is based on separability conditions. A discussion of these conditions is, however, beyond the scope of this book (e.g., see Capps and Havlicek 1987).

specification and estimation. Consequently, a large literature exists, including numerous applications for foods (e.g., Okrent and Alston 2011).

In empirical applications, the data pertain to markets, while the constraints are derived from assumptions about the behavior of individual consumers. Because market data represent averages or aggregates based on many individual consumer decisions, certain aggregation conditions should be fulfilled for the constrained demand system to be a valid representation of the market data (e.g., see Deaton and Muellbauer 1980: Chapter 6). Further, model specification should accommodate distributed lag effects and other variables (in addition to prices and income) that determine the level of market demand (see Chapter 2).

Finally, the typical demand system is estimated under the implicit assumption of one-way causation from prices to quantities, but this assumption is the opposite of a common view about the way that prices for agricultural commodities are thought to be determined. That is, as will be discussed in Chapters 4 and 5, current production of agricultural commodities is based on decisions made by farmers many months, even years, in the past. Thus, current supply may be influenced very little by current price. If so, then causality runs essentially from quantity to price; quantity is predetermined by prior events. As a consequence, a literature has also developed for complete inverse demand systems (e.g., Huang 1991). A subsequent section discusses flexibility coefficients, which are related to inverse, i.e, price-dependent, demand equations.

To summarize, the foregoing issues are about the question, what model "best" describes the changes in prices and quantities through time? Are the elasticities estimated from market data, imposing the theoretical constraints, "good" approximations of the true but unknown elasticities? The answers to such questions are a matter of debate and judgment, and it is perhaps fair to say that no single correct answer exists. With our current state of knowledge, empirical results must be judged relative to their intended use and cannot be judged relative to some absolute standard of correctness.[7]

## Total Elasticity

The own-price elasticity of demand provides a measure of the percentage change in quantity demanded in response to a 1 percent change in price, *other factors held constant*. As such, it is a partial measure of the change in quantity demanded, since other factors are not allowed to change. However, if the price of one good changes, the prices of other goods, especially substitutes, will change as well, at least in competitive markets. The prices of substitutes tend to move in the same direction, but by varying amounts. A change in the price of one product sets in motion changes in other prices and quantities that ulti-

7. King (1979) provides one of the early useful critiques of the empirical implementation of demand systems. Davis (1997) discusses the large chasm between theory and empirical practice. A basic question is, can stable (robust), useful estimates can be obtained? Many of the empirical results in the literature appear to be unstable.

mately result in a new structure of prices. For example, if the price of beef declines, the demand for its substitutes (e.g., pork) also declines. Given a fixed supply of pork, the price of pork decreases, and the change in the price of pork will, in turn, influence the demand for (and hence the price of) beef. Consequently, one needs to consider more than the partial own-price elasticity of demand for beef to estimate the net final effect on the consumption of beef of a change in the price of beef.

The foregoing discussion assumes a very short-run situation in which supplies are constant. The relationship becomes even more complex if one takes account of the possible effects of changes in the relative prices of competing commodities on quantities supplied. To predict the total effect of one initial price change, a complete knowledge of all own-price and cross-price elasticities of demand and supply is required. Such knowledge could come from what are called, in econometrics, the structural supply and demand equations. In this section, we limit the discussion to the demand side of the market.

The interrelationship among prices leads to the idea of a total demand response curve and total elasticities. The total demand response curve is defined as the price-quantity relationship that results when all other important demand variables are "allowed to act and interact as the market structure requires to reach a new equilibrium level" (Buse 1958: 882). The elasticity of total demand response, or simply total elasticity, is defined as the *net* percentage change in quantity resulting from a 1 percent change in own price, taking account of the interactions of related variables.

Assuming the principal relationships are substitute relations, then Figure 3-3 depicts the idea of a logical response relation and hence illustrates the total elasticity concept. The prices of substitutes are likely to decline as the commodity's own price declines. This would appear as a decrease in demand for the commodity under consideration. The total response curve is less elastic or more inelastic than the *ceteris paribus* demand curve. Thus, if the main relationships are among substitutes, the total elasticity is smaller in absolute value than the corresponding own-price elasticity.

Conceptually, the total elasticity coefficient can be viewed as the sum of two terms: the own-price elasticity and the cross-price elasticities multiplied by the elasticities of "price transmission," e.g., the elasticity of the price of $j$ with respect to a change in the price of $i$. The own-price elasticity is adjusted for the cross effects. Assuming just one substitute ($j$) for the commodity $i$, the total elasticity for $i$ ($T_i$) is

$$T_i = E_{ii} + E_{ij} S_{ji},$$

where $E_{ii}$ is the own-price elasticity, $E_{ij}$ is the cross-price elasticity, and $S_{ji}$ = the percentage change in the price of $j$ given a 1 percent change in the price of $i$. $E_{ij}$ and $S_{ji}$ are usually positive and less than 1 for substitutes (Buse 1958: 889).

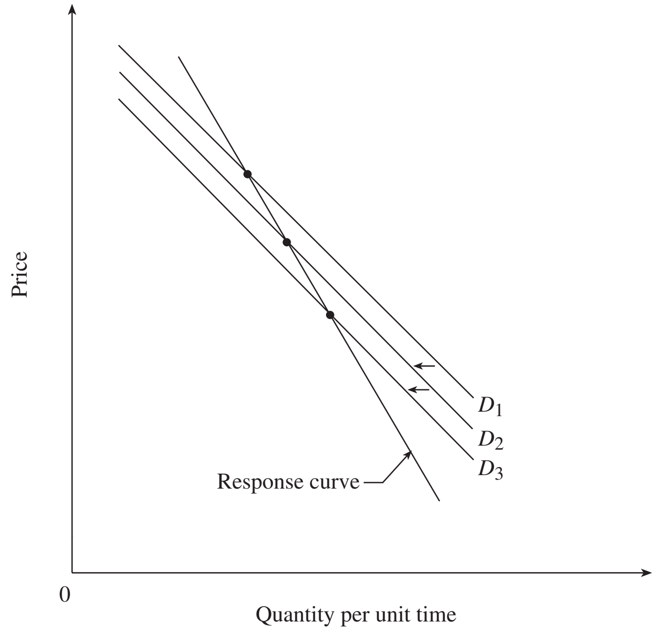



**Figure 3-3.** Hypothetical total demand response curve

$E_{ii}$ is, of course, negative and larger in absolute value than the cross elasticities. Thus, $T_i$ is expected to be negative but smaller in absolute value than $E_{ii}$. For example, if beef is treated as the sole substitute for pork, then an estimate of the total elasticity for pork (if $i$ = pork and $j$ = beef) is $T_i = -0.94 + (0.72)(0.29) = -0.73$ (Buse 1958: 886f).[8] Although these are hypothetical numbers, they illustrate that net percentage change in quantity of pork is less than that implied by the own-price elasticity.

It should be noted that the foregoing discussion implicitly assumed a static model that does not differentiate between short- and long-run own-price elasticities (which are defined in Chapter 2). In practice, if time is allowed for adjustments, the structural model may also need to include the distributed lag effects of long-run demand. The long-run own-price elasticity presumably is larger in absolute value than the short-run elasticity, but the principles related to total

8. Although the Buse paper is over 50 years old, it represents the original application in agricultural economics. His analysis uses a static four-equation supply-demand model for beef and pork. Readers familiar with econometric models will note that Buse derives the total elasticities by solving the structural equations for the reduced-form equations. A general issue in such models is, what variables are endogenous (determined by the model) and what variables are exogenous (givens, determined outside of the model)? Also, what are the important cross effects to be modeled? In other words, the Buse approach can be modified in terms of the assumptions made about the process generating the time-series data, hence the implicit assumptions being made about the net cross effects among variables. Also, as Buse notes, the ideas trace back to work by J. R. Hicks in 1939.

elasticity should still be applicable. Namely, the *net* total long-run effect is likely to be smaller than the *ceteris paribus* total long-run effect. The details are largely beyond the scope of this book (but see the section on equilibrium displacement models in Chapter 16).

## Price Flexibility Coefficients

The *own-price flexibility* is defined as the percentage change in price associated with a 1 percent change in quantity, other factors remaining constant. The price flexibility concept is potentially important for agricultural commodities because it is possible that causality runs from quantity to price. Briefly, the idea is that the biological nature of the production process means that important time lags exist between the decision to produce and the actual arrival of the commodity on the market. Thus, current production cannot be influenced by current prices. For example, the domestic production of apples this year is determined by past decisions and cannot be changed by current prices; this is sometimes referred as a "predetermined" supply. Prices must adjust to clear the market of the available supply (Chapters 4 and 5).

The situation is one of a fixed or predetermined supply and a given level of demand for a specific time period (like a quarter or a year). Within the time period, the level of current production cannot be changed. Thus, it may be logical to think in terms of an inverse demand function, $P = D^{-1}(Q)$, and to estimate the responsiveness of prices to changes in quantities.[9]

The own-price flexibility for the $i$th product is defined as

$$F_{ii} = (dP_i/P_i)/(dQ_i/Q_i) = (dP_i/dQ_i)(Q_i/P_i).$$

Like the price elasticity of demand, the own-price flexibility is negative. A flexibility of –3.0 means that a 1 percent increase in quantity is associated with a 3 percent decrease in price, *ceteris paribus.*

It appears on first examination that the flexibility coefficient can be treated as the reciprocal of the corresponding price elasticity. Thus, if demand is inelastic, the absolute value of the price flexibility coefficient is bigger than 1; prices are flexible. If demand is elastic, then prices are relatively inflexible; i.e., the flexibility coefficient is less than 1 in absolute value.

To be precise, however, one must take account of the other variables in the demand model. It is common in inverse demand functions for agricultural products to make price a function of quantities of the substitutes as well as of the

9. Although domestic production can adjust to price changes only after some (potentially long) time lag, the level of imports and exports may be able to adjust faster. Also, inventory adjustments can occur. Thus, it becomes a logical and empirical issue as to whether or not causality runs from quantity to price. In any case, as noted in Chapter 2, it is appropriate to write demand equations in inverse form even if price and quantity are simultaneously determined.

product itself. This, of course, is in contrast to the demand function, which makes quantity a function of prices. Since different variables are being held constant in the two functions, the reciprocal of the flexibility need not equal the elasticity.[10] Rather, it can be shown that the reciprocal of the flexibility sets the lower limit of the corresponding elasticity:

$$|E_{ii}| > |1/F_{ii}|.$$

In the special case when the cross effects are zero, an equality would hold. Otherwise, the own-price elasticity will be larger than the reciprocal of the flexibility.

Flexibility coefficients that are analogous to cross-price and income elasticities may also be defined. The cross flexibility of $i$ with respect to $j$ is the percentage change in the price of commodity $i$ in response to a 1 percent change in the quantity of $j$, other factors held constant. An algebraic definition is

$$F_{ij} = (dP_i/dQ_j)(Q_j/P_i).$$

The cross flexibility based on the quantity of a substitute is expected to be negative, and this is often a source of confusion because cross-price elasticities for substitutes are positive. The negative cross flexibility is perhaps best understood from the viewpoint of the mathematics of a system of demand equations (see Chapter 16). Intuitively, the reason for negative cross flexibilities for substitutes relates to the logic of the model, which is implying causality running from quantities to prices. Thus, given that the quantity of a substitute product $j$ increases, its price decreases. A decrease in the price of the substitute means that the demand for commodity $i$ decreases, and given a fixed supply of $i$, its price decreases. In sum, an increase in the quantity of *j, ceteris paribus,* results in a decrease in the price of $i$.

The flexibility of price with respect to income changes is a measure of the percentage change in price in response to a 1 percent change in income. For normal goods, the income flexibility is positive. Namely, prices move directly with shifts in demand. A larger income implies a higher price, the quantity variables fixed.

10. The derivative $dP/dQ$ from $P = D^{-1}(Q)$ is the reciprocal of $dQ/dP$ from $Q = D(P)$. However, in practice, we must compare partial derivatives from more complex models, say from

$$P_i = D^{-1}(Q_i, Q_j, Y) \text{ versus } Q_i = D(P_i, P_j, Y).$$

Since different variables are being held constant in the two equations, the partial derivative of $P_i$ with respect to $Q_i$ is not the reciprocal of $Q_i$ with respect to $P_i$ (for more discussion, see Houck 1965). A further complication arises from the statistical estimation of the equations. Namely, if the price-dependent equation is estimated by the least squares method and then solved algebraically for quantity, the resulting coefficients will differ from those obtained by estimating the quantity-dependent equation directly by least squares. This is a property of the statistical estimator.

Relationships among flexibilities analogous to those for elasticities, such as symmetry, have been derived. Thus, it is possible to fit a system of price-dependent demand functions subject to the types of restrictions used in traditional demand systems. Huang (1991) provides a brief discussion of the conceptual framework as well as references to the background literature.

## Empirical Elasticities

Economists have estimated a variety of demand elasticities using various data sources and models. The results must be treated as approximations of true, but unknown, parameters, and users of estimated elasticities (or flexibilities) should attempt to understand the limitations of the results they are using. Typically, estimated elasticities are obtained for a specific purpose, and they may be difficult to generalize to other applications.

Any statistical estimate is just that, an estimate, which at a minimum is subject to sampling error. In other words, the estimate is almost certainly not equal to the true parameter; if it were possible to obtain another sample of data from the same population, the estimates from the different samples would vary. In addition, it is possible that the data and model have problems. For example, reported prices may not be a good measure of the prices to which consumers are responding; errors in variables may exist. Likewise, the model may contain omitted variables or other inappropriate restrictions. Economists are aware of these issues, and as noted previously, a considerable literature exists about the problems of empirical demand analysis (Alston and Chalfant 1991; Davis 1997; Tomek 2000).

Further, in interpreting elasticities, one should be careful in speaking about "the" price elasticity of demand. Elasticity estimates depend on the functional form, typically varying as price and quantity change along the function. Recall also, there is a distinction between short- and long-run elasticities. Moreover, the estimates should be treated as specific to the time period, location, and product definition used in the analysis. (The estimation and interpretation of elasticity coefficients are discussed further in Chapter 15.)

As illustrated in Table 3-1, price and income (expenditure) elasticities for broad, generic commodity categories, like beef or apples, tend to be inelastic. This is an accepted generalization in agricultural economics. But, if we were measuring the responsiveness of the purchases of T-bone beef steak to variations in its price in a particular supermarket chain in Ithaca, NY, the resulting estimate would probably be price elastic. In this case, the elasticity pertains to a specific product, T-bone steak, in a particular location and not to the aggregate of all beef in the United States. The specific product has far more substitutes: other cuts of meats, including other beef products, and meats available in other stores in the same area. Similarly, the price elasticity for Golden Delicious apples sold as fresh fruit is probably elastic even though the demand for all apple varieties for the aggregate of all uses is price inelastic.

**Table 3-1.** Estimated elasticities of demand, selected foods and beverages, United States

| Product | Own-price elasticity | Income (expenditure) elasticity |
|---|---|---|
| Beef | –0.76 | 0.99 |
| Pork | –0.64 | 1.16 |
| Poultry | –0.38 | 0.64 |
| Apples | –0.57 | 0.20 |
| Bananas | –0.80 | 0.32 |
| Eggs | –0.73 | –0.71 |
| Potatoes | –0.45 | 0.17 |
| Lettuce | –0.77 | 0.28 |
| Milk | –0.15 | 0.61 |
| Juice | –0.27 | 0.66 |
| Coffee/Tea | –0.46 | 3.14 |

*Sources:* Beef, pork, and poultry, Beach et al. (2007: table ES-2); eggs, Okrent and Alston (2011: table A-6); apples, bananas, potatoes, and lettuce, Okrent and Alston (2011: table 30); milk, juice, and coffee/tea, Zheng and Kaiser (2008: table 3).

Thus, for the reasons indicated, no set of estimates is perfect. It seems doubtful, for example, that the expenditure elasticity for eggs is negative or that the expenditure elasticity for coffee and tea is so large (Table 3-1). The foregoing demonstrates the importance of being careful in the interpretation and use of estimated elasticities.

## References

Alston, J. M., and J. A. Chalfant. 1991. "Can We Take the Con out of Meat Demand Studies?" *West. J. Ag. Econ.* 16:36–45.

Beach, R. H., C. Zhen, N. E. Piggott, M. K. Wohlgenant, C. L. Viator, and S. C. Cates. 2007. *An Economic Analysis of the Effectiveness of the Pork Checkoff Program*. RTI Project Number 0210314.000, Research Triangle, Inc.

Brandow, G. E. 1961. *Interrelations among Demands for Farm Products and Implications for Control of Market Supply*. Penn. State Univ. Ag. Exp. Sta. Bul. 680.

Buse, R. C. 1958. "Total Elasticities—A Predictive Device," *J. Farm Econ.* 40:881–891.

Capps, O., Jr., and J. Havlicek, Jr. 1987. "Concepts of Consumer Demand Theory," in *Food Demand Analysis*, R. Raunikar and C-L. Huang, eds. Ames, Iowa: Iowa State Univ. Press. Chapter 1.

Davis, L. G. 1997. "The Logic of Testing Structural Change in Meat Demand: A Methodological Analysis and Appraisal," *Am. J. Ag. Econ.* 79:1186–1192.

Deaton, A., and J. Muellbauer. 1980. *Economics and Consumer Behavior*. Cambridge, UK: Cambridge Univ. Press. Chapters 1–6.

George, P. S., and G. A. King. 1971. *Consumer Demand for Food Commodities in the United States with Projections to 1980*. Giannini Foundation Monograph 26, Univ. of Calif. Div. of Ag. Sciences.

Houck, J. P. 1965. "The Relationship of Direct Price Flexibilities and Elasticities," *J. Farm Econ.* 47:789–792.

Huang, K. S. 1991. *U.S. Demand for Food: A Complete System of Quantity Effects on Prices*. USDA Tech. Bul. 1795.

———. 1993. *A Complete System of U.S. Demand for Food*. USDA Tech. Bul. 1821.

King, R. A. 1979. "Choices and Consequences," *Am. J. Ag. Econ.* 61:839–848.

Okrent, A. M., and J. M. Alston. 2011. *Demand for Food in the United States*. Giannini Foundation Monograph 48, Univ. of Calif. Div. of Ag. Sciences.

Salathe, L. E., and R. C. Buse. 1979. *Household Food Consumption Patterns in the United States*. USDA Tech. Bul. 1587.

Stigler, G. J. 1962. "Henry L. Moore and Statistical Economics," *Econometrica* 30:1–21.

Tomek, W. G. 2000. "Commodity Prices Revisited," *Ag. & Res. Econ. Rev.* 29:125–137.

Zheng, Y., and H. M. Kaiser. 2008. "Advertising and U.S. Nonalcoholic Beverage Demand," *Ag. & Res. Econ. Rev.* 37:147–159.

CHAPTER **4**

# Supply Relationships in Agriculture

In this chapter, we explain supply concepts, with special emphasis on the supply of farm commodities. The production of farm commodities has unique features that require the underlying economic principles be adapted to these features. The basic economic principles of supply analysis can also be applied to marketing and processing firms that supply foods to final consumers. Thus, attention is focused first on the theoretical basis of supply functions and related elasticity concepts. Then, we discuss the factors that shift the level of the supply function, with an emphasis on the supply of farm commodities. Finally, we cover some special topics in supply analysis related to the adoption of new technology and aggregate output from the farm sector.

## Theoretical Basis of Supply Functions

### Firm Level

The theory of the firm under certainty commonly assumes that the firm's objective is to maximize profits.[1] For a farmer producing one product ($a$) using two inputs ($x_1$ and $x_2$), the short-run profit is defined simply as $P_A a - P_1 x_1 - P_2 x_2$ where the $P$'s represent prices per unit of output and inputs, respectively. We also use the notation that $a$ is the firm's output and A is the aggregate output of all the suppliers; all firms are assumed to face the same prices in this competitive model. A firm's output is related to the inputs through a production function: $a = f(x_1, x_2)$. For example, the output could be corn, and the inputs could be

1. Economists in defining *profits* take account of returns to (costs of) all factors of production, including the capital and labor provided by the owners of the firm. Profit is the return to the owner's management after adjusting for the costs of all of the inputs. Moreover, as will be discussed below, a distinction can be made between the short and the long run. Firms may accept small or negative profits in the short run in the expectation of long-run profitability.

nitrogen fertilizer and seed (for simplicity, we assume only two inputs). The relevant prices are for corn, fertilizer, and seed.

Note, the profit function can be rewritten $P_A[f(x_1, x_2)] - P_1x_1 - P_2x_2$ and that this equation can be used as a basis for deriving the producer's demand for inputs (the $x$'s) and the producer's supply of $a$. The precise algebraic expression for the supply function depends on the form of the production function, hence on the form of the cost functions. We do not present the mathematics, which use the first-order conditions (first derivatives of the profit equation) for profit maximization, but it can be shown that a firm maximizes profits by producing at the level at which marginal costs equal marginal revenue. *Marginal costs* are defined as the increment in total costs associated with producing one more unit of output. *Marginal revenue* is the increment in total revenue associated with producing and selling an additional unit of output.

In this analysis, prices are assumed to be exogenously given to the producer; the optimal allocation of inputs by the firm is in response to the market-determined prices. With this assumption, which is true in a competitive industry, the firm's marginal revenue is the prevailing market price. As the market price changes, so does the marginal revenue confronting the firm. With an upward-sloping marginal cost curve, the profit-maximizing output rises as the market price increases.

Hypothetical average and marginal cost curves for a representative firm, producing a single product, are shown in Figure 4-1. These curves are based not only on fixed input prices but also on the assumption of a given technology—a given production function—for a given (optimal) plant size. In other words, these variables are held constant in deriving a supply schedule. As previously suggested, the rates of change of the marginal and average costs functions are related to the form of the underlying production function (input-output relationship). The upward slope of the marginal cost curve assumes that diminishing returns occur as additional units of the input are used. For example, increasing the level of fertilizer increases yield, but probably at a diminishing rate, and ultimately yields can decline if "too much" fertilizer is used.

It should be noted, as an aside, that the behavior of marginal costs can vary with the type of output under analysis. For example, if the output is the service of storing a commodity, marginal costs might plausibly be a fixed, constant amount over a considerable range of storage; as long as storage capacity exists, the addition of another unit of commodity to the inventory may not result in a larger marginal cost. Diminishing returns are probably not perceptible over a broad range of the storage function.

For our purposes, however, the point to emphasize is that given the common assumption of rising marginal costs, the optimal output varies as the price of the product rises or falls (Figure 4-1). The price of A equals marginal revenue, assuming firms in a competitive industry (see Chapter 5), and the optimal output of the firm depends on the intersection of this price with marginal costs. If, for

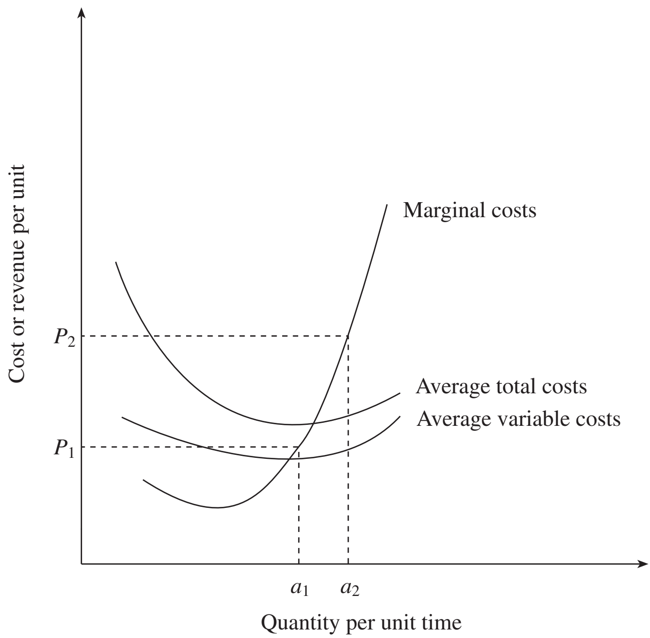



**Figure 4-1.** Cost curves and optimum output at alternative prices

instance, the output amount is at a marginal cost that is less than marginal revenue, then an increase in output would result in an incremental increase in revenue that is larger than the incremental increase in cost.

The average total and average variable cost curves are also important in understanding supply response. Variable costs are those that change with the quantity produced and can be avoided by ceasing production. Total costs combine variable and fixed costs (those costs that do not vary with output). Thus, the average total cost function lies above the average variable cost function (Figure 4-1). If the product's price does not cover the average total costs, the firm will incur losses; however, losses will be minimized (and the returns over variable costs maximized) by equating the marginal cost and price. As long as price is above average variable costs, the firm can produce and make a partial contribution to fixed costs. If price drops below average variable costs, then the optimal decision is to cease production; the costs that can be avoided exceed the returns from producing and selling the product.

The distinction between fixed and variable costs depends on the length of run (the time period) allowed for adjustment. In the short run, inputs such as fertilizer, seed, and hired labor can be varied; a firm's equipment and land resources may be fixed. Intuitively, as longer time periods are allowed for adjustment to price changes, more resources become variable. In the long run, from the firm's perspective, all resources can be changed; e.g., more land can be purchased or rented. Consequently, all costs become variable. Based on the foregoing discussion, one can conclude that the short-run supply function is upward sloping as long as prices are above average variable costs and that, in the long run, the firm's supply becomes more responsive to a given price change. Indeed, under some circumstances, the supply function may be horizontal; this occurs if the long-run average cost and long-run marginal cost curves are constant over the range of output (e.g., Varian 2010: Chapter 22).

These concepts assume that the prices of inputs and output are certain. In fact, for farmers (and sometimes for processors as well), the price of the output is often uncertain. Lags exist between the decision to produce (planting and breeding decisions) and the realization and sale of the output. In the corn belt of the United States, corn is planted in April and harvested in October and November. The price of most inputs may be known in April, but the price of corn in October or thereafter is unknown in April. Agricultural output, such as corn production, is also subject to the vagaries of the weather, pests, and diseases over the growing season. Consequently, production decisions must be based on *expected* prices and *expected* output; expectations may be based on forecasts from statistical models; futures markets quotes; government, university, and private predictions; and past prices. To the extent that expectations are not met, price and yield *risk* exists.

It is, therefore, appropriate to think of a firm's planned output as being a function of expected price. A notation for expected price is $P^*_{t-1} = E[P_t|I_{t-1}]$, where $E$ is the expectation operator, $I$ is the information, and the subscripts represent time periods. In words, the price expected to occur at time $t$ is determined at a prior time, $t - 1$, based on the information available at time $t - 1$. The notation states that the expected price is conditional on the information available to the decision maker at time $t - 1$. With the passage of time, new information will arrive, which is likely to make the planting-time expectation wrong. Assuming that expectations are not realized, then the difference $P_t - P^*_{t-1}$ represents the error in expectations, and price risk is presumably some type of function of these errors. Likewise, a distinction can be made between planned output and realized output, with a corresponding measure of yield risk. So, the concepts discussed before might better be modified, for farmers, to be in terms of planned supply as a function of expected price.

A common assumption is that producers are risk averse; i.e., producers are willing to pay to shift risk to someone else (say, by buying crop insurance or forward contracting) and hence risk is equivalent to a cost of production. For this reason, the objective of the firm may not be simple profit maximization. Rather, firms may be viewed as maximizing expected utility, where utility is a function of expected returns and the riskiness of these returns. Thus, price and yield risk can be treated as factors that shift the supply function and are held constant in defining a *ceteris paribus* supply function. Alternatively, one could assume that producers are risk neutral; i.e., their production decisions are not influenced by the level of risk.

In either of the foregoing cases, the prices of inputs, such as fertilizer, are held constant in defining the cost curves. Thus, changes in these prices (and for risk-averse producer, changes in levels of risk) will shift marginal costs and hence the supply function.

To this point, our definition of *costs* has been limited to the costs of inputs used in the production of a single commodity. We have ignored the effects of potential alternative uses of resources, i.e., opportunity costs. Many farms are multi-product enterprises. If the relative prices of two commodities change, it may become profitable to shift the use of resources from the production of one commodity to another. Indeed, from the perspective of a firm, the decision could be made to shift entirely from producing one commodity to another; i.e., the firm's supply of *a* may decline to zero even though its price is above its average total costs of production (as previously defined). Put another way, *opportunity costs* should be incorporated into the marginal and average cost curves, where the opportunity cost of producing a particular commodity is the value of the same resources used to produce the next best alternative.

The opportunity cost of some inputs may be difficult to estimate. If a firm has an outmoded set of equipment, the best alternative use may be its salvage value. On the other hand, if the cost of operating a tractor is identical for producing both corn and soybeans, then the opportunity cost depends on the relative prices of the two commodities.

Clearly, opportunity costs can vary from farm to farm because the quality of inputs can vary. Differences in soil quality may make it possible for some farmers to produce a wide range of crops, while others may have few alternative uses for their land. Or a farmer may have inferior management skills, and the most profitable decision may be to use one's labor in an off-farm job and rent the land to another farmer. That is, determining the appropriate measure of opportunity costs depends on defining the appropriate alternative use of resources. What indeed is the next-best alternative use? Notwithstanding the potential difficulty of answering this question, the general conclusions are that a firm's production decisions are a function of expected prices of the alternative outputs that can be produced with the available inputs, the prices of these inputs, and the associated risks.

In sum, from a general model perspective, we can think in terms of a system of equations for a multi-product firm. This system consists of the supply equations for the products produced by the firm and of the demand equations for inputs used in producing these products. The quantities supplied and demanded are a function of the output and input prices, and these prices may be certain or uncertain. Thus, like consumer demand, an internally consistent model of input demand and product supply can be specified (e.g., Ball 1988), but this topic is beyond the scope of this book.

## Market Supply Curves

To understand price behavior in markets, we need to aggregate over the behavior of individual firms and discuss market supply. If a firm's optimal output is $a$ and if for convenience we assume $m$ identical firms, then the aggregate output is $A = ma$. In practice, firms are not identical, but market supply curves are based on the combined responses of all firms to a change in price. Thus, the quantity supplied of a product is generally positively related to changes in its price, other factors held constant, or more precisely in agriculture, the quantity that farmers plan to supply is positively related to expected prices. But the additional output that producers are prepared to offer at higher prices depends on the time allowed for adjustments to take place. Thus, supply needs to be discussed in terms of length of run.

In the very short run, by definition, the supply function is a vertical line; insufficient time exists for quantity supplied to adjust to price changes; supply is fixed. The length of the very short-run time period differs for different commodities. For a crop with an annual harvest, production cannot be changed until the next crop year. (Note, it may be important to make a distinction between production and total supply; production may, depending on the product, be augmented by inventories or imports.)

In the short run, some resources are variable and others are fixed. The number of acres planted to an individual crop can be altered, fertilization rates can be varied, etc. It is even possible that some portion of a crop will not be harvested if prices do not cover harvesting costs. The general point is that the short-run supply function can now be viewed as the aggregation of short-run marginal cost curves. Hence, the short-run supply function depicts the quantities producers are willing to produce and sell in the short run as prices vary, all other factors held constant. Given logical input-output relationships, this supply function will have a positive slope; a larger price results in a larger quantity supplied, and vice versa.

The tendency for supply curves to become more responsive (flatter) as more time is allowed for adjustments is illustrated schematically in Figure 4-2. The figure is drawn so that the long-run supply function intersects the price axis ($P_2$) above the intersection of the short-run function ($P_1$). This illustrates the principle that prices must cover all costs in the long run. In contrast, some costs can be treated as fixed in the short run, and production can continues even though all fixed costs are not fully covered. Of course, in reality, the quantity supplied is unlikely to drop to zero, but in the long run, large adjustments in resource use are possible.[2]

2. In 1960, New York State was estimated to have 88,000 farms and 14.3 million acres of land in farms. By 2010, New York State had approximately 36,300 farms and 7.0 million acres in farms (NASS 2012: table 76). Total land used for farming declined by about 50 percent. Clearly resource changes can be large in a 50-year period.

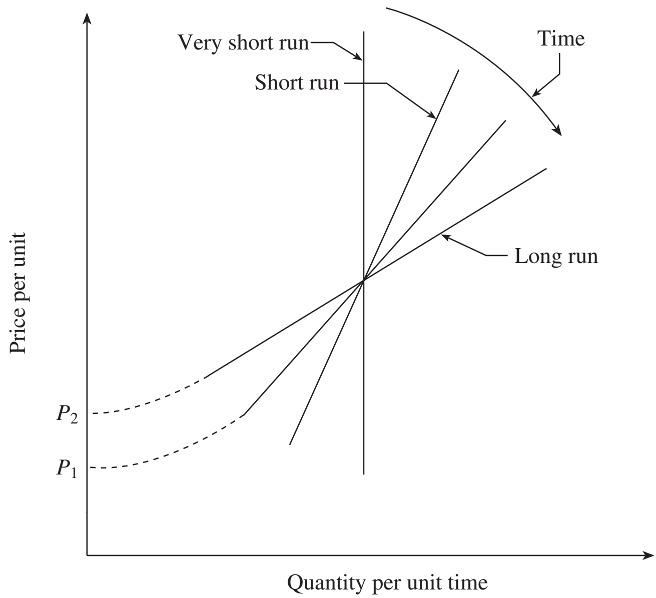



**Figure 4-2.** Changing supply-price relationships through time

In empirical analyses, the short run is the time required for an initial supply response to a price change. For an annually produced crop, this would be 1 year, and annual observations would be appropriate for the analysis. For a poultry enterprise, the response time is quicker than a year, and quarterly observations would be appropriate for analysis. Estimates of intermediate-run and long-run effects are based on distributed lag models, which were introduced in Chapter 2. As mentioned there, it is difficult to obtain unambiguous estimates of the long-run effect of an own-price change, partly because the analyst must specify the length of the long run and the form of quantity adjustments to the price changes. In addition, the supplies of agricultural products are influenced by a large number of factors, which can change with the passage of time, and it is difficult to disentangle the separate effects of the different variables.

Improvements in technology have been especially important influences on supply for most agricultural commodities. Although the usual theoretical definitions of short- and long-run supply assume that technology (the production function) is a given, price changes may provide incentives or disincentives for the generation and adoption of new technology. If technological change causes average costs to drop faster than demand increases, prices will tend to decline (Antle 1999). This has the potential appearance (or equivalence) of a decreasing-cost industry, with a downward-sloping long-run supply function. In this situation, however, it is not external economies or declining factor prices that are causing costs to decline; rather, it is improvements in the production technology for firms within the industry. Empirical models often try to separate the long-run response to price changes from trends that are, perhaps, related to improved technology.

## Price Elasticity of Supply

The price elasticity of supply is defined in a manner analogous to the price elasticity of demand. It expresses the percentage change in quantity supplied in response to a 1 percent change in price, other factors held constant. Using the notation $A$ is the quantity produced and supplied, and $P_A$ = price of $A$, a market supply function can be written $A = S(P_A)$. Then, letting $d$ represent a small change, the price elasticity of supply may be written

$$E_{AA} = (dA/A)/(dP_A/P_A) = (dA/dP_A)(P_A/A).$$

Here, $dA/dP_A$ also may be viewed as a first derivative.

Based on arguments made in the previous section, the price elasticity of supply is non-negative. In the very short run, supply is fixed and the elasticity is zero. The very short-run supply function is said to be perfectly inelastic. An *inelastic* supply refers to the range of elasticities between 0 and 1. The quantity supplied is relatively unresponsive to price changes. An *elastic* supply refers to coefficients larger than 1; the percentage change in quantity supplied exceeds the percentage change in price. In principle, supply could also be perfectly elastic—a horizontal line.

As is the case with demand functions, the elasticity coefficient typically varies in magnitude at different points on the supply function. It is convenient to speak of "the" price elasticity of supply, but such a coefficient is usually measured at the arithmetic mean of the data. Thus, an inelastic supply should be interpreted as applying to the average of historical experience. In empirical research, supply equations are sometimes specified as straight lines, but as Houck (1967) pointed out, this specification places special restrictions on the magnitude of the supply elasticities. As the quantity supplied is increased along the function (i.e., as quantity approaches infinity), the price elasticity of supply approaches 1. If the quantity-dependent function intersects the price axis first (horizontal intercept negative), the elasticity is always greater than 1 but approaches 1 as price and quantity increase. If the function intersects the quantity axis first (horizontal intercept is positive), the elasticity is always less than 1 but, again, approaches 1 as quantity grows. If the linear function with a positive slope intersects the origin (intercept 0), then the elasticity is a constant equaling 1.

A linear function may be an acceptable approximation when prices and quantities vary over a limited range. It is worth noting, however, that logical reasoning suggests that the supply function is probably not linear if the output price varies over a large range. As noted in a previous section, if prices drop below

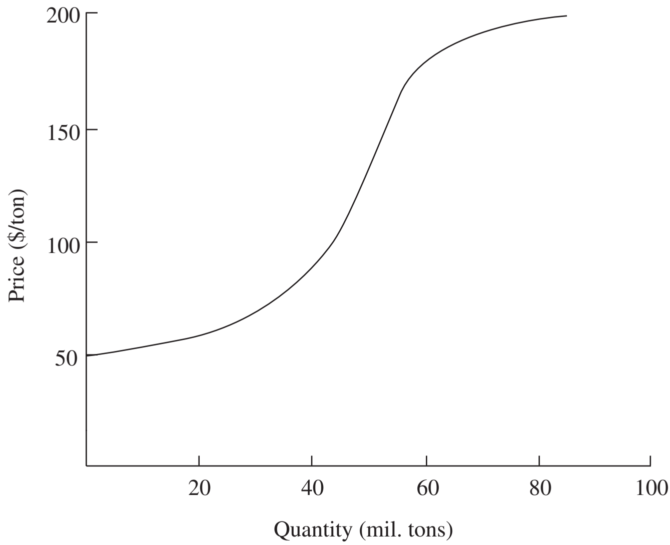



**Figure 4-3.** Hypothetical supply curve for commodity A

a firm's average variable costs, the firm will cease production. In practice, firms do not have identical cost functions, and the aggregate output is unlikely to drop to zero. Nevertheless, it is possible to imagine prices reaching a sufficiently low level for a large shift of resources to alternative uses. Likewise, as prices increase, a point may be reached where resources are switched from other enterprises to the output of the commodity. That is, in a middle range, price variability may have little effect relative to the alternative uses of resources, but upper and lower break points might be reached that have a large effect on producers' choices.

A hypothetical supply relation for commodity A, which illustrates the foregoing point, is given in Figure 4-3. Supply functions may not be linear, and elasticities are not constants. Of course, in practice, it is difficult to specify an appropriate functional form for supply equations, but we should not be surprised if linear equations provide poor forecasts when forecasts are made outside the range of the sample data used to fit the equation.

The technical problems encountered in attempting to obtain reliable estimates of supply elasticities are similar to those discussed in connection with estimating demand elasticities. Indeed, the problems of isolating the effect of price from other variables is likely to be greater in supply analysis.[3] Also, given the arguments made in a previous section, it is logical that elasticity estimates can vary by region, where the number of alternatives differ, and differ depending on

3. Just (2000) provides a discussion of issues in empirical supply analysis.

**Table 4-1.** Price elasticities of supply for selected commodities and regions

| Product | Elasticity |
|---|---|
| Corn | |
| Southeast U.S. | 0.670 |
| North-central U.S. | 0.170 |
| U.S. | 0.293 |
| Soybeans | |
| Southeast U.S. | 0.195 |
| North-central U.S. | 0.300 |
| U.S. | 0.296 |
| Wheat | |
| North-central U.S. | 0.250 |
| U.S. | 0.340 |
| Cotton | |
| Southeast U.S. | 0.506 |
| U.S. | 0.466 |

*Sources:* Southeastern United States, Liang et al. (2011); North-central United States, Lin and Dismukes (2007); total United States, Lin et al. (2000).

whether short-run or long-run adjustments are being estimated (Liang et al. 2011). For livestock, elasticities also may vary by stage of production. Chavas and Johnson (1982) found that in broiler production, the response to a given price change is greater for the breeding flock that supplies chicks for the broiler industry than it is in the growing stage of broiler production.

Estimated magnitudes of price elasticities for selected crops are shown in Table 4-1. These estimates are for the short run; long-run elasticities can be much larger. Empirical analyses for crop supply, such as those in Table 4-1, often use hectares per acres planted as the measure of quantity. Variables like planted-area represent producers' intended supply, while final output can be influenced by the vagaries of the weather, pests, etc. Likewise, for hogs, quantity supplied may be measured by "sows farrowing," i.e., the number of female animals giving birth. Although elasticity estimates vary across studies and their precise values are uncertain, the conventional wisdom of agricultural economics is that the supply of major commodities is price inelastic, especially in the short run.

## Changes in Supply

Changes in a product's price typically (but not always) explain a relatively small proportion of the total variation in output of farm commodities. Short-run changes in output are often influenced by weather conditions, diseases, and pests, while long-run trends in supply are attributable to such factors as improvements in technology that result in higher yields. These and other factors that lead farmers to produce more or less at the same price are referred to as *determinants of supply* or *supply shifters.* For forecasting or policy analysis, it is important to know whether changes in output occur as a result of movements along a static supply curve (changes in quantity supplied) or because of shifts in the supply curve (changes in supply).

A shift in the supply curve to the right (an increase in supply) means that a larger quantity will be offered at a given price; a shift to the left has the opposite effect. Previous sections have alluded to the variables that are held constant in defining the static supply schedule. They can be categorized as follows:

1. changes in input (factor) prices;
2. changes in prices of commodities competing for the same resources or factors of production;
3. changes in the prices of joint products (e.g., soybean oil and meal for soybean production);
4. changes in the level of price and yield risks faced by producers;
5. changes in technology that influence efficiency and the costs of production; and
6. changes in institutional factors, like government programs.

The supply of agricultural commodities is also influenced by random states of nature, such as a drought. These too can be treated as shifters of the supply function, and sometimes particular events are modeled.[4] They also may treated as a part of the random error term in statistical models, i.e., as an unexplainable residual.

As with demand relations, it is possible to distinguish between parallel shifts in the function, which result from changes in the variables, and structural change, implying that the unknown parameters or the form of the function has changed. To illustrate, a hypothetical supply equation may be written as follows:

$$A = \beta_0 + \beta_1 P_A - \beta_2 X,$$

where $A$ is the production of the commodity A, $P_A$ is the commodity's price, and $X$ is the price of an input used in producing A. The $\beta_i$'s are the unknown parameters. In a static supply model, $X$ is assumed to be constant, but when $X$ changes, the level of supply will shift. In this case, an increase in the price of the input will reduce the level of supply.

The foregoing algebraic expression specifies a linear relationship among the variables with constant parameters (the $\beta_i$'s). A change in the functional form or in the parameters is defined as a structural change. Changes in technology,

4. Some studies have explicitly modeled weather as well as economic and technological factors as an explanatory variable in statistical crop yield models (e.g., Li et al. 2011).

such as the development of a new, higher-yielding variety, can result in structural change. The slope as well as the position of the supply function also may change because the size distribution of farms changes, new areas are brought into production, or the government alters the provisions of farm programs. Shift variables (like $X$) and structural changes may be related and can occur simultaneously. Those modeling supply responses try to include a comprehensive set of supply shifters and, within this context, assume the structure is constant. It is possible to specify models, however, that let the parameters change with different program regimes (e.g., Liu et al. 1991).

## Input Prices

Recall that static cost curves are defined assuming input (factor) prices are constants. An increase in factor prices, other variables held constant, shifts the supply curve to the left; for a given product price, supply is reduced. A decrease in input prices has the opposite effect. Thus, the effect on supply of a change in the product price can be reinforced or neutralized by a change in input prices.

A given percentage decrease in the price of all factors of production accompanied by an equal percentage decrease in the price of the product results in the same use of quantities of inputs. This can be viewed as a downward movement along the static supply curve in response to a reduction in the price of the product and a shift to the right in the supply curve as a result of a corresponding decline in the price of factors. This situation is illustrated schematically in Figure 4-4. Production will remain at $A_1$ despite the decline in the product price from $P_1$ to $P_2$ if the factor prices decline by an amount sufficient to shift the supply curve from $S_1$ to $S_2$.

Product-factor price ratios (say, $P_A/P_X$) are often important in determining the output of farm commodities, especially livestock products such a beef, pork, chicken, milk, and eggs. For these commodities, feed costs are usually the most important component of total costs. An understanding of how prices influence input use is also essential if one wants to predict the demand for farm inputs, and the concepts for optimum factor use are an extension of the concepts introduced earlier.

A producer maximizes profits by using each factor up to the point where the last unit just pays for itself or, in economic terminology, where the marginal value product (addition to total revenue) just equals the marginal (incremental input) cost. To determine optimal use, the production function must be known. For example, we previously denoted a firm's production function as $a = f(x_1, x_2)$. For simplicity, we now consider one input, writing $a = f(x)$; a hypothetical function is illustrated in the upper part of Figure 4-5. In this illustration, some output ($a$) is produced without using any of input $x$, say fertilizer. This is shown as a positive intercept on the vertical axis. Production rises as more of the input is added, but output increases at a diminishing rate. The rate of increase in output ($da/dx$) is called the marginal physical product of $x$ ($\mathrm{MP}_x$); this relationship is

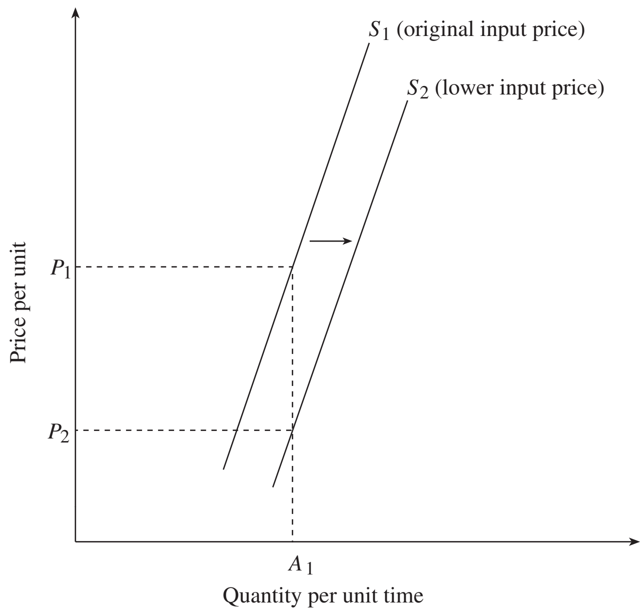



**Figure 4-4.** Changes in supply associated with a decrease in input prices and a corresponding decline in the price of the product

shown in the lower part of Figure 4-5. Following the rule of profit maximization, the optimum level of factor use is at the point where the marginal product of *x* multiplied by the price of *A* ($P_A$) just equals the price or marginal cost of *X* ($P_X$). (The notation is intended to convey that a firm is using the amount *x* to produce the quantity *a,* but the firm is paying market prices for *X* and receiving the market price for *A*.)

In mathematical notation, this rule can be written as follows:

$$(da/dx)P_A = P_X.$$

By dividing both sides by $P_A$, we obtain $da/dx = P_X/P_A$. This equation is simply another way of stating the conditions that must prevail for optimum factor use. Profits will be maximized when the marginal product of each factor ($MP_x$) is equal to the appropriate factor-product price ratio ($P_X/P_A$).

An increase in the price of an input will lead to reduced use of that input, other factors remaining constant. A higher marginal product is necessary to offset the higher price of *X*; the marginal product can be increased only by using less (for given production technology). A reduction in the price of the output will have the same effect as an increase in the price of the input. Either a decrease in $P_X$ or an increase in $P_A$ will reduce the factor-product price ratio and lead to an increase in the optimum level of input use. At the new optimum level,

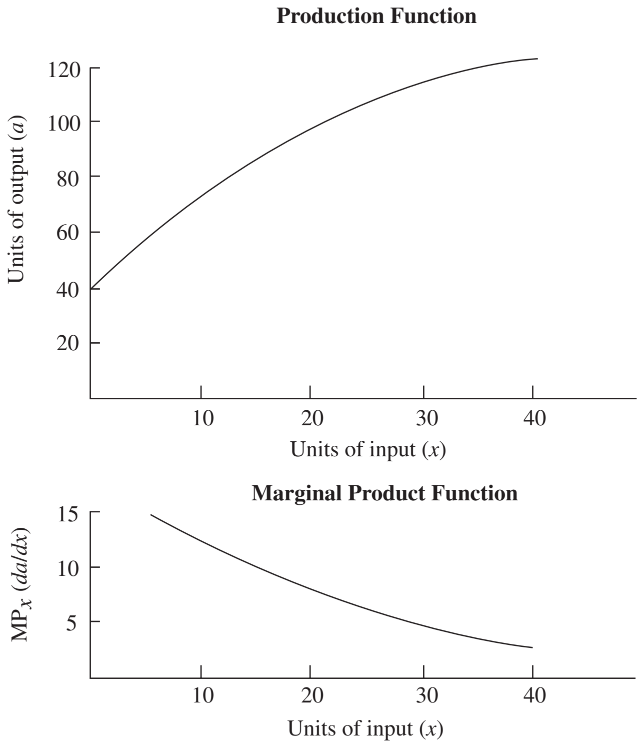



**Figure 4-5.** Hypothetical total and marginal product functions for input *x*

the lower marginal product, associated with the increased factor use, is equal to the smaller factor-product price ratio.

Optimum factor use will change only with an increase or decrease in *relative* prices. It will not change if both product and factor prices rise or fall by an equal percentage. Under some circumstances, the prices of inputs and the end product tend to move together. For example, a decrease in livestock product prices reduces the demand for feed grains. This decline in the demand for the input, *ceteris paribus,* will result in a fall in the price of feed. (The effect of changes in product prices on input use are discussed further in an appendix to this chapter.) The general point is that considering the static effect of a product price change without considering the secondary effects on factor prices (and perhaps on other product prices) is potentially misleading about the net supply response.

Market supply relations are an aggregation of the firm-level relationships. Theory of the firm implies that relative prices are important in analyzing supply, and as a consequence of the importance of relative price changes, product-factor (and sometimes product-product) price ratios are commonly used in supply

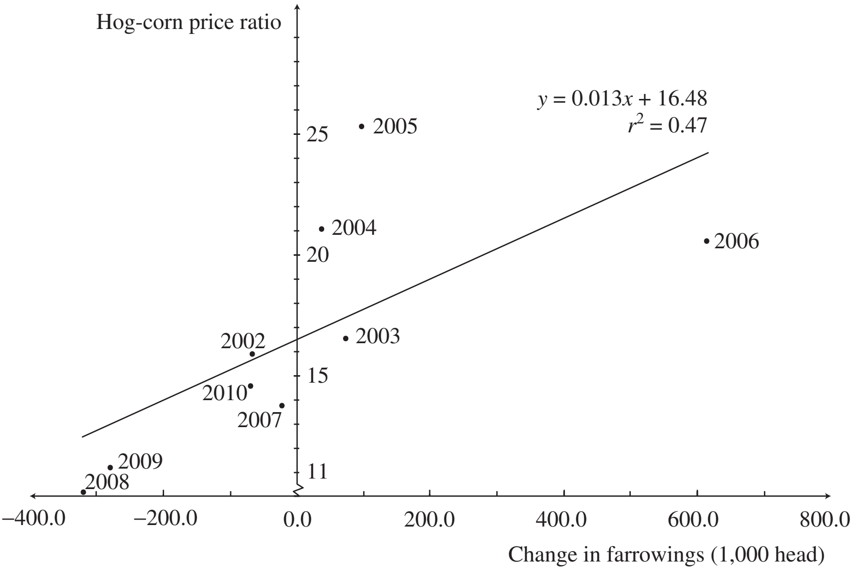



**Figure 4-6.** Relationship between hog to corn price ratio and the change in numbers of sows farrowing in the following year, United States, 2002–10

analysis. One of the best known of these is the hog-corn price ratio, which is the price of hogs divided by the price of corn. For example, when the price of hogs (specifically barrows and gilts) is \$90 per cwt. and the price of corn is \$6 per bushel, the ratio is \$90/\$6 = 15. A high hog-corn price ratio is associated with a subsequent increase in pork output, while a low ratio leads to a decrease in production.

The initial effect of a high hog-corn price ratio is to increase the number of sows bred and, hence, an increase in farrowing (the number of sows producing a litter of pigs). Farrowings are an indicator of future hog supplies. Figure 4-6 illustrates the relationship that prevailed from 2002 through 2010 between the hog-corn price ratio (shown on the vertical axis) and subsequent changes in the number of sows farrowing (shown on the horizontal axis). These data suggest that ratios below 16.5 usually resulted in a decrease in farrowings, while ratios above 16.5 led to an increase in farrowings. Of course, other factors, besides corn prices, influence pork output, and consequently the empirical relationship between farrowings and the price ratio is far from exact; i.e., the simple scatter diagram in Figure 4-6 does not take account of other factors that influence supply.[5]

5. In Figure 4-6, the lagged ratio of hog to corn prices is a proxy for producers' expectations about returns to hog production. It is an imperfect measure, and alternative measures of farmers' expected returns are discussed in Chapter 9. Also, the two-variable relationship depicted in the figure does not account for other variables influencing farrowings. Thus, the scatter of observations reflects the effects of omitted variables and errors in measuring producers' price expectations.

## Competing Commodities

The supply curve for a given commodity will shift to the left if alternative commodities become more profitable to produce; it will shift to the right if other commodities become less profitable. Competing commodities are ones that can be produced with the same resources. A competing commodity, B, can become more profitable because the price of B rises relative to the first commodity, A, or because costs of producing B decline relative to A. The latter situation can arise if a new technology increases the yield of B relative to A. Thus, relative changes in product prices, yields, or efficiency can change the relative profitability of different crops.

Corn and soybeans are important alternative crops over much of the central and southern parts of the United States. When, in the past, the yield of corn grew relative to the yield of soybeans, this made corn production more profitable at least temporarily, but with the passage of time, an increase in the output of corn has the effect of changing relative prices. That is, the initial yield advantage may be offset by subsequent changes in relative prices. This, again, emphasizes the dynamic behavior of price and quantity adjustments.

Gardner (1976) makes the planted acres of soybeans a function of the planting-time prices of soybeans and corn, as well as other variables. In one specification, the prices are planting-time observations for post-harvest futures contracts; these prices are often interpreted as unbiased estimates of the post-harvest prices, i.e., a measure of the rationally expected prices based on the information available at planting. The coefficient for the corn price variable is –10.2, which suggests that, for a $1 per bushel increase in the price of corn, soybean acres planted would decrease 10.2 million acres, other factors held constant. Of course, a $1 per bushel change in the price of corn is a large change, but the point is that changes in the price of corn do indeed influence soybean acreage. Although the magnitudes of the coefficients may vary, this general result is confirmed in other studies (e.g., Lin et al. 2000).

## Joint Product Relationships

The supply of some agricultural commodities is determined in part by joint relationships. Joint products are produced in more or less fixed proportions from the underlying product. In many cases, one of the joint products is minor, such as the value of the hides of beef animals, and is not considered separately from the value of the major component. In a few cases, however, both products have major economic value and should be considered in supply relationships. The size of the sheep breeding flock, for example, is influenced by both the prices of lamb and of wool. Thus, it is also true that the supply of wool is influenced by the price of lamb, and vice versa. Both prices have a positive coefficient in the supply equation. The short-run price elasticity for the sheep breeding flock with respect to the price of lamb is estimated to be 0.87, and the elasticity with respect to the price of wool is 0.32 (Whipple and Menkhaus 1989).

## Price and Yield Risks

Production decisions are based on expected prices and yields. These expectations may not be realized, and consequently price and yield risks exist. Price risk is a function of differences between expected prices and the ultimately realized prices. A variety of empirical measures of risk can be estimated from historical data, e.g., measures based on lagged differences between expected and actual prices. In principle, the concern is with adverse deviations from the producer's expected price (which may not treat positive and negative differences symmetrically).

Although risk may be difficult to measure, it is nonetheless possible that the level of a supply curve is influenced by producers' perceptions of price and yield risks. Assuming most producers are adverse to risk, the greater the risk, the less farmers are likely to produce at a given price, and vice versa. Risk-averse producers can shift risk by purchasing crop or revenue insurance and/or using forward contracts, futures contracts, or options contracts (to be discussed), but all these alternatives have some cost.[6] These instruments do not eliminate risk from the marketplace; rather, these vehicles allow producers to shift the risk to others who are willing to bear it for a price.

Empirical research supports the conclusion that increases in price risk reduce supply, *ceteris paribus.* This conclusion is robust to a variety of alternative measures of risk (Tronstad and McNeill 1989). In other words, if a measure of risk is included in the supply model, its coefficient is negative and statistically significant. Adding a variable to measure risk (e.g., standard deviation or variance of price) to the model also results in the coefficient of the price variable becoming larger. This implies that omitting risk from the model has the consequence of biasing the price coefficient downward; the net effect of price on supply is underestimated.

## Technology

Improvements in technology have been the principal source of long-term shifts in agricultural supply functions. An improvement in technology is defined as something that enables firms to produce more output with the same quantity of inputs as previously (or, equivalently, the same output with fewer resources). In technical terms, the production function shifts so that producers find it profitable to increase output at the same ratio of product to factor prices. The theory of the firm typically treats the production function (technology) as a given, and hence a change in technology is a structural change.

Among the more important technical changes that have increased agricultural supply are the development of high-yielding varieties of crops and breeds of



6. Government programs have also been justified as a way to reduce or shift price and yield risks. The alternative programs available to producers (and their rules) are known for any given year but can be altered with the passage of time. These changes may introduce a different source of price/revenue risk that may influence long-run investment decisions.

livestock; better methods of insect, disease, and weed control, including the need for less of such inputs, mechanization and global positioning systems (GPSs) that make it possible to plant and harvest in a more timely and precise fashion, and better tillage techniques. Developments in biotechnology are also making it possible to alter the attributes of commodities, thereby making them potentially more valuable for specialized uses.

With continued research, a stream of new technologies enters the market. Whether these technologies are adopted depends, of course, on the economics of adoption—namely, do the economic benefits to producers equal or exceed the costs of adopting and using the new technology? Thus, new methods and varieties are not adopted instantly (indeed, they may not be adopted at all); rather, there is a process of adoption. As a consequence of the flow of new technologies and the varying rates of adoption by farmers, the aggregate effect of new technologies is an upward trend in yields. Thus, models of yields often contain a time trend (Just and Weninger 1999). Likewise, models of total supply may include a trend variable to accommodate the effects of new technology.

## Institutional Factors

The supply of agricultural commodities is often influenced by institutional factors, especially various government programs and rules. In the United States and indeed throughout the world, commodity supplies have been influenced by a variety of governmental programs and rules; these have included acreage allotments, incentive payments to keep land idle, varying levels and types of commodity payments, tariffs and import quotas, zoning, land-use and environmental regulations, marketing-order rules, and bases and quotas for output. In much of the twentieth century, the supply of farm commodities grew faster than demand (especially in high-income countries), and government programs typically attempted to limit supply and support farm prices. This has been less true since the late 1980s; since then, prices have been mostly influenced by market supplies and demands. Thus, farm programs in the United States have evolved to provide a safety net for prices and revenue rather than to support prices at above-equilibrium levels. Since safety nets presumably reduce risk, such policies still can influence the level of supply.

The diversity of, and changes in, farm programs make the empirical analysis of supply difficult. In many countries, it is difficult to find price and supply data that have not been influenced by government programs. Thus, supply analysts face a difficult trade-off: using a long sample of data that requires modeling the effects of changing government programs versus using a short sample for a time when government programs had no effect (or were unchanged). In the former case, it may be difficult to specify the appropriate variables to fully capture the government effects, but in the latter case, a degrees-of-freedom problem may exist (i.e., too few observations for effective statistical analysis). Supply models exist in the literature that attempt to model government programs (e.g., Lin et al. 2000; Liu et al. 1991).

## Supply-Response Relation

The traditional supply function in economics, as previously discussed, relates the quantity supplied to price changes, all other factors held constant. Of course, in practice, a change in a product's price can start a series of changes such that the net change in quantity differs from the *ceteris paribus* change in quantity supplied. An idea can be developed on the supply side that is analogous to the total elasticity concept discussed for demand relations in Chapter 3. The net response to a price change is the outcome of movements along a supply curve and shifts in the level of the curve.

In this section, we concentrate on a particular supply-response relation, introduced by Cochrane (1955), that implies an asymmetric or irreversible relationship. This concept is based on the hypothesis that, when price increases, new technologies are more likely to be adopted. Under conditions of rising prices, firms have increased incentives to adopt new technologies. Also, higher prices may make it possible to finance investments in new technologies from retained earnings. Thus, the increase in price is expect to have two effects. First, it will cause farmers to increase output along the static supply curve; second, it will lead to a shift in supply as the new technology is used. The resulting increase in output will thus be greater than might have been anticipated based solely on the static concept of supply.

Once adopted, improved production practices may be retained even though the price of the product subsequently declines. This is especially true where the technology involves a new investment that cannot be changed easily. To the degree that this is true, the response to a price decrease can be depicted as a movement along the new supply function. It does not shift back to the original (old technology) level. This implies that the magnitude of the net response to the price decline is less than the magnitude of the net response to the price increase. It is in this sense that the supply response is irreversible. A hypothetical response relation is shown in Figure 4-7. At a price $P_1$, producers offer an output of $A_1$, but as the price increases to $P_2$, output expands along the diagonal between $S_1$ and $S_2$, ultimately reaching $A_2$. If the price thereafter declines to $P_3$, output declines along the new supply curve $S_2$, resulting in the supply of $A_3$.

Favorable or unfavorable prices obviously can have a marked influence on the rate at which new technology is adopted and, hence, on the rate of change of farm output. Farmers must have an incentive to use new techniques and access to sufficient funds to make the necessary investments.

## Constraints on Supply Responses

As already noted, the production of agricultural commodities is based on a biological process, and it is important to recognize that biological and physical constraints can limit supply response in the short run. In animal agriculture,

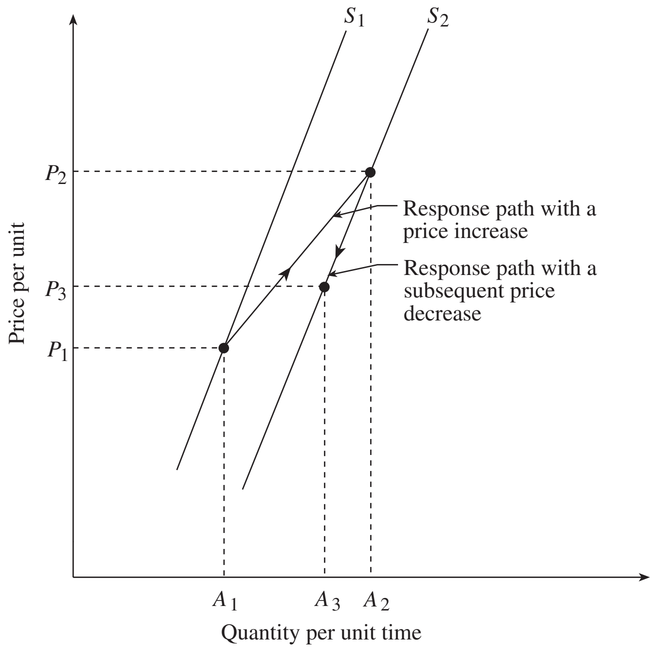



**Figure 4-7.** Hypothetical supply response paths

output at a point in time is limited by the size of the breeding herd. Specifically, $H_t = kB_t$, where $B$ is the size of breeding herd (or flock), $k$ is the number of offspring per herd member, $H$ is the output of breeding herd, and $t$ is the time period. For example, the number of beef calves born in year $t$ is necessarily limited by the number of sexually mature cows in the herd in that year.

Each year ranchers make decisions about the number of cows to cull and the number of young females (heifers) to retain for the breeding herd. Thus, the size of the next year's herd is given by the identity $B_{t+1} = B_t + b_t - c_t$, where $b$ is the number of heifers added to the herd and $c$ is the number of cows culled from the herd. Culling and replacement decisions are based on economic principles, but the point is that these decisions take place within the context of the inventory (constraint) of the available animals. Likewise, the placement of calves and yearlings into lots for grain feeding to produce grain-fed beef is dependent on the number of animals available in appropriate weight categories at the time that decisions are being made. And the marketing of finished (say, 1,100-pound) steers this month depends on the placement of a certain number of animals of various weight categories into feedlots in prior months; a 500-pound calf, if fed a high-energy ration, takes approximately 6 months to reach market weight and be slaughtered for beef.

Not only does the existing stock of animals place a constraint on decisions, but it also imparts dynamic behavior to supply and prices. Current supply is dependent on past decisions made about managing the herd (Rosen 1987). Thus, cyclical behavior can arise. We will say more about this in Chapter 9, when the time dimension of price behavior is further discussed.

The amount of land under cultivation is fixed at any point in time, and thus represents a type of physical constraint. This may not be especially important in the analysis of supply for a single crop. Land can be moved into and out of the production of alternative crops. But if one is considering a system of supply equations that pertains to all crops, the sum of areas planted to individual crops cannot exceed the total area available for crops in the region or country. That is, the constraint is that the sum of the individual crop areas must equal the total area available for crops. If $A_i$ is the area in $i$th crop, TA is the total land area, and $i = 1, 2, \ldots, n$ total number of crops, then by definition $A_1/\text{TA} + A_2/\text{TA} + \cdots + A_n/\text{TA} = 1$. The proportions of the total land area devoted to various uses must sum to 1.

The choice of crops is also limited by the climate and the quality of soil resources. In this context, allocation decisions about land use are based on economic principles, but in supply analysis, it is often important to take explicit account of the effect of resource constraints on production decisions.

## Aggregate Farm Output

Sometimes there is an interest in analyzing the aggregate output of the agricultural sector. For example, there is an interest in how total food output balances the aggregate demand for food. Will there be enough food to feed a growing world population? What is the effect of changes in the supply-demand balance on farm and consumer prices?

Changes in aggregate farm output over time are associated mainly with shifts in the aggregate supply function rather than movements along a static supply function. New knowledge and technology are the principal drivers of aggregate supply (Johnson 2000), and the experience of the last half of the twentieth century was that food supplies more than kept pace with growth in demand (Tweeten 1998). But, in the twenty-first century, the question remains, can food supplies keep pace with food demands, or will demand outpace supply with consequent increases in prices?

The aggregate supply relation for agriculture in most countries is highly price inelastic in the short run. This is due mainly to the fact that resources once committed to farming tend to remain in use. This phenomenon is sometimes call *asset fixity.* Land, buildings, and machinery often have low values in alternative uses outside agriculture. The salvage value of the buildings or equipment can be lower than the value of these resources used in agricultural production.

Individual farmers may go out of business, but their land and equipment can be resold for use in agriculture.

Of course, in the long run, by definition, resources like land may move completely out of agriculture or perhaps new land may be added via reclamation projections (see footnote 2). Old equipment may be depreciated and not replaced, or with high prices, new equipment may be added. But these adjustments take time.

Also, as noted previously (and see the appendix to this chapter), output and input prices are often correlated. If product prices decline, input prices may also decline. Of course, the ratio of input to output prices does change, but looking at changes in product prices in isolation from changes in input prices is potentially misleading about the net supply response. When product prices decline but supply changes relatively little, observers may incorrectly infer that price changes have no influence on supply. Such an observation misses several points—namely that resources are relatively fixed in the short run and input prices may have also changed so that the relative prices have changed very little.

## Concluding Remarks

The analysis of supply of agricultural products is firmly rooted in basic economic principles, but the unique characteristics of the agricultural production process must be taken into account in supply analysis. The biological nature of the production process introduces time lags and often precludes rapid adjustments in output. Also, random factors such as weather events influence output.

Farmers can alter production decisions, but the analysis of these decisions requires an understanding of the price expectations held by farmers. Likewise, concerns about price risk may influence production decisions. Further, asset fixity and constraints on adjustments may need to be taken into account. Much progress has been made in understanding the phenomena that influence supply relations in agriculture, but the analysis of supply decisions remains fraught with difficulty. Chavas, Chambers, and Pope (2010) review the production economics literature, and Just (2000) provides a comprehensive critique of empirical production research in agriculture.

## Appendix: Product Prices, Factor Prices, and Factor Use

This chapter has emphasized the potential interrelationships between product and factor prices, and this appendix provides some additional detail. The discussion is also related to the concept of derived demand introduced in Chapter 2. Here, the product is the farm commodity, designated $A$, and the input used to produce $A$ is $X$. For one firm $a = f(x)$, but for convenience, we use an aggregate function, $A = f(X)$ and consider market-level relationships. The derived demand for the factor $X$ depends on the price of $X$ relative to the price of $A$. Changes in product prices shift the demand for factors up or down, and this shift, in turn, affects both factor prices and factor use. Whether a given change in product prices will affect mainly the price or the quantity of the factor depends on the shape of the factor supply function.

Factor use and factor prices are, in theory, determined simultaneously (as with product prices and quantities) by the intersection of demand and supply curves. Factor demand is determined by the marginal productivity of the factor (the first derivative of the production function, $dA/dX$) multiplied by the prices of the product. Because the marginal product of a variable factor declines as more of it is applied to a fixed factor, the marginal productivity curve slopes downward and to the right. This relationship, $dA/dX$, when multiplied by the (fixed) price of the product, determines the marginal value product, i.e., the increment in total revenue associated with a 1-unit increase in the use of a factor.

A typical marginal value product curve is illustrated for input $X$ in Figure 4-8. This function represents the demand for input $X$ used to produce commodity $A$. The quantity demanded of $X$ is a function of the price of $X$ for a given price of $A$. An increase in the price of the product, $A$, shifts the factor (input) demand curve for $X$ upward and to the right; a decrease in the price of $A$ shifts the demand for $X$ in the opposite direction.

A factor supply curve ($S_X$) also is shown in Figure 4-8. It slopes upward and to the right, indicating that more of $X$ will be supplied at higher prices. The supply function is relatively flat in this example, which implies that the production of $X$ is relatively responsive to changes in the price of $X$.



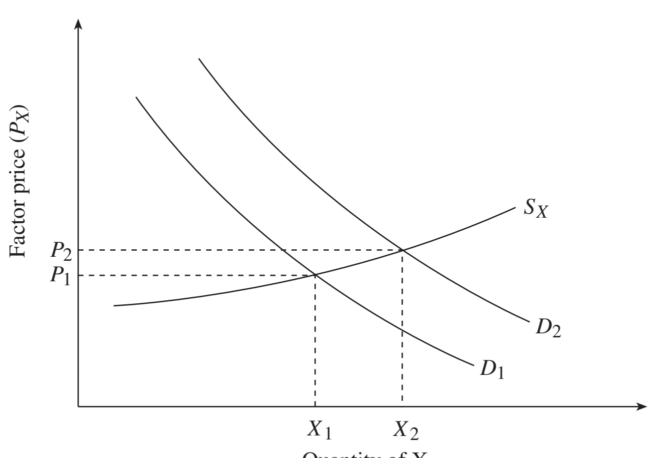



**Figure 4-8.** Effect of an increase in product price ($P$) on factor price and factor use with an elastic supply function

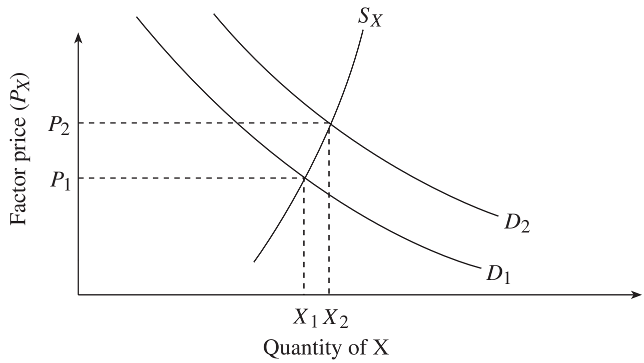



**Figure 4-9.** Effect of an increase in product price on factor price and factor use with an elastic factor supply function

The effect of an increase in the price of the product, *A*, is shown as a shift in the demand curve from $D_1$ to $D_2$. The effect of this shift in demand is to raise the price of *X*. In this case, there is a relatively small increase in the price of *X* because of the elastic supply curve. The optimum use of *X*, as shown in Figure 4-8, increases from $X_1$ to $X_2$. (In some theoretical models, the supply of *X* is assumed to be perfectly elastic; a shift in demand has no effect on price. This assumption is often a matter of convenience rather than necessarily being realistic.)

If, instead of being relatively responsive, the supply schedule for input *X* had been relatively unresponsive, the effect on the factor price and use of an increase in the price of *A* would be quite different. This situation is illustrated in Figure 4-9. Because of the steeper-sloped supply function, the increase in the price of *X* is larger than in Figure 4-8; however, the increase in factor use is less.

Figures 4-8 and 4-9 illustrate the importance of knowing something about the factor supply elasticity in attempting to estimate the potential effect of a change in product prices on factor prices and factor uses. If the factor supply is price inelastic (within the relevant range of prices), the major effect of a change in product prices is to change factor prices. However, if the supply function is relatively elastic, the major effect is to change factor use. In both illustrations, an increase in the price of *A* is associated with an increase in the price of *X*; the output and input prices are positively correlated. It is often unrealistic to assume that input prices will remain unchanged in the face of change in the price of the product. But it is true that, in some cases, factor prices are unresponsive to changes in product prices.

## References


Antle, J. M. 1999. "The New Economics of Agriculture," *Am. J. Ag. Econ.* 81:993–1010.

Ball, E. V. 1988. "Modeling Supply Response in a Multiproduct Framework," *Am. J. Ag. Econ.* 70:813–825.

Chavas, J-P, R. G. Chambers, and R. D. Pope. 2010. "Production Economics and Farm Management: A Century of Contributions," *Am. J. Ag. Econ.* 92:356–375.

Chavas, J-P., and S. R. Johnson. 1982. "Supply Dynamics: The Case of U.S. Broilers and Turkeys," *Am. J. Ag. Econ.* 64:558–564.

Cochrane, W. W. 1955. "Conceptualizing the Supply Relation in Agriculture," *J. Farm Econ.* 37:1161–1176.

Gardner, B. L. 1976. "Futures Prices in Supply Analysis," *Am. J. Ag. Econ.* 58:81–84.

Houck, J. P. 1967. "Price Elasticity and Linear Supply Curves," *Am. J. Ag. Econ.* 57:905–908.

Johnson, D. G. 2000. "Population, Food, and Knowledge," *Am. Econ. Rev.* 90:1–14.

Just, R. E. 2000. "Some Guiding Principles for Empirical Production Research in Agriculture," *Ag. & Res. Econ. Rev.* 29:138–158.

Just, R. E., and Q. Weninger. 1999. "Are Crop Yields Normally Distributed?" *Am. J. Ag. Econ.* 81:287–304.

Li, X., T. Takahashi, N. Suzuki, and H. M. Kaiser. 2011. "The Impact of Climate Change on Maize Yields in the United States and China," *Ag. Sys.* 104:348–353.

Liang, Y., J. C. Miller, A. Harri, and K. H. Coble. 2011. "Crop Supply Response under Risk: Impacts of Emerging Issues on Southeastern U.S. Agriculture," *J. Ag. & Applied Econ.* 43:181–194.

Lin, W., and R. Dismukes, 2007. "Supply Response under Risk: Implications for Counter-Cyclical Payments' Production Impact," *Rev. of Ag. Econ.* 29:64–86.

Lin, W., P. C. Westcott, R. Skinner, S. Sanford, and D. G. De La Torre Ugarte. 2000. *Supply Response under the 1996 Farm Act and Implications for the U.S. Field Crops Sector*. USDA Tech. Bul. 1888.

Liu, D. J., H. M. Kaiser, T. D. Mount, and O. D. Forker. 1991. "Modeling the U.S. Dairy Sector with Government Intervention," *West. J. Ag. Econ.* 16:360–373.

NASS. 2012. *New York Agricultural Statistics, 2011–2012*. Nat. Ag. Stat Ser., USDA, Albany, NY.

Rosen, S. 1987. "Dynamic Animal Agriculture," *Am. J. Ag. Econ.* 69:547–557.

Tronstad, R., and T. J. McNeill. 1989. "Asymmetric Price Risk: An Econometric Analysis of Aggregate Sow Farrowings, 1973–1986," *Am. J. Ag. Econ.* 71:630–637.

Tweeten, L. 1998. "Anticipating a Tighter Global Food Supply-Demand Balance in the Twenty-First Century," *Choices, The Magazine of Food, Farm, and Resource Issues*. Third Quarter: 4–7.

Varian, H. R. 2010. *Intermediate Microeconomics, a Modern Approach*, 8th ed. New York: W.W. Norton & Co.

Whipple, G. D., and D. J. Menkhaus. 1989. "Supply Response in the U.S. Sheep Industry," *Am. J. Ag. Econ.* 71:126–135.

CHAPTER 5

# Price Determination: Theory and Practice

This chapter is about the determination of product prices and hence about the economic forces influencing price behavior. How a particular product's price behaves, in terms of the level and frequency of change, may differ with the structure of the market for that product. The *market structure* depends on such characteristics as the number of buyers and sellers, their size distribution, barriers to entry in the industry, and the degree of product differentiation.

## Classification of Markets

Economists' commonly classify markets by assuming many buyers and then differentiating among markets by the number and distribution of sellers. In this context, markets may be classified as competitive (many sellers), monopolistic (a single seller), and oligopolistic (a few sellers). Another category that is sometimes used is monopolistic competition, which is defined as many firms selling similar but differentiated products.[1] In this section, we outline the characteristics of these four types of markets.

A common starting point in economics is to define a *perfectly competitive* market. There are four conditions for this type market.

1. There are many buyers and sellers; hence, each is so small in relation to the entire market that no buyer or seller can affect the product's price.

1. Markets are also classified by assuming many sellers and varying the number of buyers. A *monopsony* is defined as a single buyer, and *oligopsony* is defined as a few buyers. The oligopsony market structure is relevant for some commodity markets where many farmers are selling to a few buyers. Other market structures can be defined that depend on the varying combinations of numbers of buyers and sellers and of product differentiation. For example, a market with a single buyer and a single seller is called a *bilateral monopoly*. The literature on market structure is large (e.g., Carlton and Perloff 2004). This chapter can provide only an introduction.

2. The product is homogenous; the product of each seller is identical to that of other sellers.
3. All resources are completely mobile; it is costless for resources to enter or leave the market.
4. All of the buyers and sellers have perfect knowledge of the relevant forces determining prices.

Obviously, no such market exists in the real world. The notion of a perfectly competitive market is used as a benchmark for evaluating actual market structures. But, although no market is perfectly competitive, some markets come close to competitive conditions. We will define them as *purely competitive* or *atomistic* markets. In such markets, four conditions apply. (1) The number of buyers and sellers is sufficiently large that no individual has a perceptible influence on price; the individual seller is a “price taker.” (2) The product is sufficiently homogenous that the product of one firm is almost a perfect substitute for that of another. (3) The cost of entry or exit from the market, although not zero, is low; thus, resources are relatively mobile. This also implies that no governmental or other restrictions exist on resource use. (4) Information about market conditions, although not costless, is widely available to all market participants, and market participants are well informed about these economic conditions.

In a competitive market, it is assumed that every producer-seller seeks to maximize profits by selling at as high a price as possible and that every buyer seeks to maximize utility by obtaining the product at as low a price as possible. Since each buyer and seller cannot perceptibly influence price, the demand and supply relations appear horizontal to the respective individuals. For instance, the individual farmer views the (derived) demand function as being perfectly elastic (horizontal); she or he can sell any proportion of the firm’s output at the then-prevailing price without influencing it. Nonetheless, the collective actions of buyers and sellers aggregated into market demand and supply functions determine market prices.

A second type of market is designated as an *absolute monopoly.* The distinguishing characteristic of this type of market structure is a single seller. The firm’s demand schedule coincides with the industry demand schedule. Presumably a monopoly could not exist unless the firm’s product or service has unique attributes. Perhaps the firm has a patent on the product, which excludes other producers, or the firm has a monopoly in a particular location because it is very costly for a second firm to enter this location. True monopolies appear to be rare because, typically, substitutes (or the threat of substitutes) exist. Nonetheless, the theory underlying monopoly prices is useful in understanding some types of price behavior (discussed later).

*Oligopoly* refers to a market with a few large sellers. Each firm produces a large fraction of the industry’s total output, and consequently, the action of one firm in the industry can greatly influence other firms. In a *pure oligopoly,* the sellers produce a homogenous product. Aluminum smelting is an example. In a *differentiated oligopoly,* the firms produce similar but not identical products. The breakfast cereal industry is an example. Because each oligopolistic industry tends to develop unique interrelationships among firms, this type of market structure is the least susceptible to generalizations about price and output behavior. Or, put another way, economists have developed a number of alternative models of oligopolies, each assuming different types of behavior by firms in the industry.

The fourth type of market structure is *monopolistic competition.* This refers to a market in which a large number of sellers offer similar but differentiated products. The livestock feed industry in an example. These products are presumably close substitutes, but the individual sellers are able to differentiate their product on the basis of style, quality attributes (real or imagined), associated services, location, and so forth. This situation implies that the firm has some influence on price, but this pricing discretion is limited by the existence of close substitutes. The demand relation faced by the individual sellers, although not perfectly elastic (horizontal), is presumably very elastic at prevailing market prices.

In practice, actual market structures do not necessarily fall neatly into the foregoing categories. Many markets involve different degrees of product differentiation. For example, the farm-level market for apples is essentially competitive, but some apple producers are able to differentiate their product by providing a service (e.g., via a roadside stand or a "you-pick" orchard) or attributes desired by some consumers (e.g., organic and local-origin products). Thus, even though apples grown on different farms appear to be perfect substitutes, some product differentiation occurs.

Also, an industry may consist of a few large firms and a large number of small firms. This complicates the task of classifying a market by the number of sellers and thereby deciding whether an oligopoly exists or not. One approach is to use a measure of concentration such as the proportion of total industry sales made by the four largest firms. If the four largest firms in the industry account for 90 percent of total sales, the market may be classified as an oligopoly, but there is no precise level of concentration that separates an oligopoly from a competitive industry. Further, a small level of concentration at the national level may mask a possible high concentration in local market areas, which is often true of fluid milk markets. In practice, the analyst must examine the relationship of market behavior to market characteristics.

For some purposes, it is sufficient to classify markets as having flex or fixed price behavior. For many agricultural products, prices at the farm level are variable (flex). Prices may rise by 50, 100, or even 200 percent in less than 1 year and then decline to the original level a few months later. This behavior occurs when markets are relatively competitive. In fixed price markets, firms have some control over prices. Prices may rise more or less with inflation, or they may drop as new technologies are introduced. Such changes tend to be gradual and not dramatic. For example, auto manufacturers might raise prices by 5 percent from one model year to the next, or they may offer price discounts in the face of competitive pressures. But the price of autos does not double or triple and then decline by a like amount.

Although farm product markets are flex price markets, it is possible that such markets have imperfections such as asymmetrically held information or a few buyers having market power. Likewise, wholesale and retail markets for foods may have a combination of competitive and monopolistic attributes. Thus, we turn to a discussion of price determination in both competitive and monopolistic markets.

## Price Determination under Pure Competition

Models of farm-level prices for agricultural products often assume a competitive market structure, and unquestionably many agricultural products are produced by a large number of sellers. Two broad reservations exist, however, about the applicability of this model. One is that government programs may influence price behavior. This was true for much of the twentieth century in the United States, when farm programs influenced the prices of perhaps one-half of farm output, and also has been true for many other countries around the world. More recently, policy changes have made government programs less influential in the pricing of farm commodities in the United States. Nonetheless, it remains important to ask, is the price behavior of the particular product influenced by government programs?

A second reservation about the applicability of the competitive model for agricultural products relates to the number of buyers. There is a tendency in the U.S. economy toward increased concentration in food manufacturers, grain exporters, and other buyers of commodities.[2] Thus, in studying particular markets, it is important to ask whether the structure is an oligopsony (a market with few buyers) and whether this structure is influencing price behavior.

With these caveats in mind, we first explore price determination under competitive conditions. One of the concepts used is that of an equilibrium price. This is simply the price at which quantity demanded equals quantity supplied in the market. If the demand curve has a negative slope and the supply curve has a positive slope, the two functions will intersect at some price. At prices above equilibrium, the quantity that consumers are willing to buy is less than

2. In 1980, the four largest beef processors slaughtered 36 percent of all steer and heifers marketed; by 2010, the figure had grown to 85 percent. The comparable figures for the four largest hog processors are 34 percent in 1980 and 65 percent in 2010 (USDA 2012). These markets also contain many small processors. A question is, do the large meat processors have sufficient market power to pay producers prices that are below competitive levels (see Chapter 6)?

the quantity producers are willing to sell, while at prices below equilibrium, the quantity demanded exceeds the quantity supplied. Thus, statements such as "demand exceeds supply" are meaningless unless specified in relation to a particular price.

Also, the discussion that follows abstracts from the "discovery" of prices for individual transactions and concentrates on the determination of an equilibrium price.[3] In the real world of commodities, individual transactions often have unique attributes, which result in an array of transactions prices at any point in time. These attributes include variation in quality of individual lots, different locations, and differing terms of trade. In addition, individual traders may have imperfect information. We will explore the reasons for some of these price differences in subsequent chapters. Here we think in terms of price determination under relatively homogeneous conditions for the transactions, and in practice, prices reported for months, quarters, or years are averages of prices for individual transactions.

## Very Short Run

In the very short run, supply is perfectly price inelastic (Chapter 4). For nonstorable commodities, supply may consist of only the current season's production. (The vertical supply function implies that additional supplies cannot be imported within the time period in response to a price change.) The intersection of the vertical supply function with a sloping demand curve determines the equilibrium price, designated as $P_1$ in Figure 5-1. This price exactly clears the market of available supply. If the fixed supply is defined as $A$, then in equilibrium, quantity supplied equals quantity demanded, $A = Q$, and price is determined from the inverse demand function, $P = D^{-1}(Q)$.

The level of the very short-run supply function, of course, can shift from one harvest to the next. Poor growing conditions in one year would result in a small quantity available for sale, such as $S_1$ ( $= A_1 = Q_1$) in Figure 5-1, while good growing conditions would result in a large quantity for sale, such as $S_2$ in Figure 5-1.[4] In this model, quantity supplied varies because of random (e.g., weather) events, and price is determined by the level of demand.

The real world is obviously more complex than the simple model shown in Figure 5-1. Demand functions also shift, and a farm-level price will be related to a derived, rather than a primary, demand function. Nonetheless, this elementary model is a useful first step in understanding agricultural product price behavior. It demonstrates that the degree of price variability depends on the

3. A distinction is made between the theory of price determination and price discovery. *Price discovery* refers to the institutional method of arriving at a transaction price, such as an auction or private negotiations, and is discussed in Chapter 11.

4. Total production could be so large that, if the entire crop were harvested and marketed, the resulting price would be below the harvesting and marketing costs. In this case, a part of the crop would be abandoned. The model depicted in Figure 5-1 does not cover this situation.

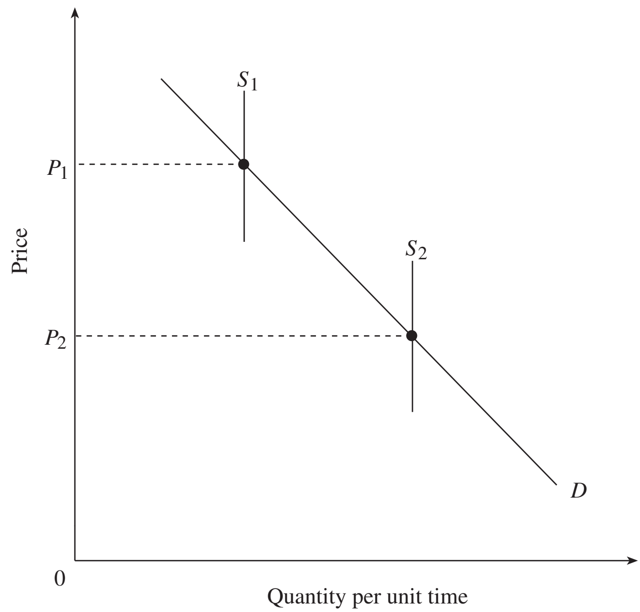



**Figure 5-1.** Equilibrium prices for perfectly inelastic supply functions

slope of the demand function and on the magnitude of the random shifts in supply. If demand is price inelastic and/or supply shifts are large, prices will have a large variance.

If a commodity is storable, this adds to the complexity of the model of price determination. Although current output of an agricultural commodity is determined by decisions made in the past, the ability to store means that current production can be augmented by inventories carried over from the previous period and that part of current production and inventories can be carried to the next crop year.

The carrying of inventories and determination of seasonal price variability can be modeled in several different ways. In Chapter 2, we made the distinction between the demand for current consumption and speculative demand. In this case, current supply can be viewed as allocated (by the level of current price relative to expected future prices) between consumption and inventories. We defer the discussion of seasonal prices and expected or futures prices to Chapters 9 and 12.

If the problem of modeling expected prices (for the forthcoming period) is ignored, a simple alternative model is to treat current-period demand as combined speculative and consumption demand, specified as $P = D^{-1}(Q)$, where $Q = A + I_{-1} - I$, and $A$ is the predetermined production, $I_{-1}$ is the beginning inventory, and $I$ is the ending inventory. The ending inventory is modeled as depending on current price, $I = I(P)$. In other words, $A$ and $L_{-1}$ are given by prior decisions, i.e., predetermined, as implied by the definition of the very short run. Then, the market determines price and the allocation of the given supply between current consumption and ending inventory. (In practice, of course, the levels of the demand and inventory depend on additional variables.)

Substituting the definition of $Q$ into the inverse demand function gives $P = D^{-1}(A + L_{-1} - I)$, and define the inventory equation as $I = I(P)$. This says that given $A$ and $L_{-1}$, the market determines the current price and ending inventory. In the terminology of economics, $P$ and $I$ are current endogenous variables (the variables determined by the model), and $A$ and $L_{-1}$ are predetermined, i.e., they are fixed by past decisions. Although price and ending inventory are, according to the economic model, simultaneously determined, a simplification sometimes used in empirical analysis is to make price a function of ending inventory (e.g., Westcott and Hoffman 1999).

A scatter plot of observations for the relationship between the marketing year average price for corn in the United States and the predetermined supply (beginning inventories plus production) of corn for 2007–08 to 2012–13 is shown in Figure 5-2. Although the scatter of observations is far from perfect, a clear relationship exists; a small supply is associated with high prices, and vice versa. This relationship has the advantage of making price a function of the level of supply that can be accurately estimated at the beginning of the marketing year (rather than ending inventory).

For empirical analyses, a common practice is to make price a function of the ratio of the ending stocks to total use, $P = f(I/U)$, where $U$ is the total use (Chapter 16). This equation is, however, a simplification of a far more complex model. The identity, above, should be disaggregated to define important uses, including exports. For example, $I = A + L_{-1} + M - Q - X$, where $I$ is the inventory, $M$ is the imports, $Q$ is the domestic consumption, and $X$ is the exports. Domestic consumption might be further subdivided into various types of uses. Equations may be required, for example, to explain export demand and domestic demands for various uses. In addition, if the sample period contains observations when government programs and stocks were important, this effect also should be modeled (Westcott and Hoffman 1999). A complete model highlights that commodity prices depend on a host of factors, including international events.

For our current purposes, however, two points are emphasized. First, a major determinant of prices is the available supply, $A + L_{-1}$, which is fixed for each very short-run time period, and rather well-known and influential theoretical models make this assumption (e.g., Deaton and Laroque 1992). Second, the existence of inventories has the potential to impart dynamic behavior to prices. This period's price depends, in part, on the inventory carried over from the prior period ($L_{-1}$). Note, the carrying of stocks can help reduce the amplitude of price fluctuations; stocks augment current production. However, if a sequence of

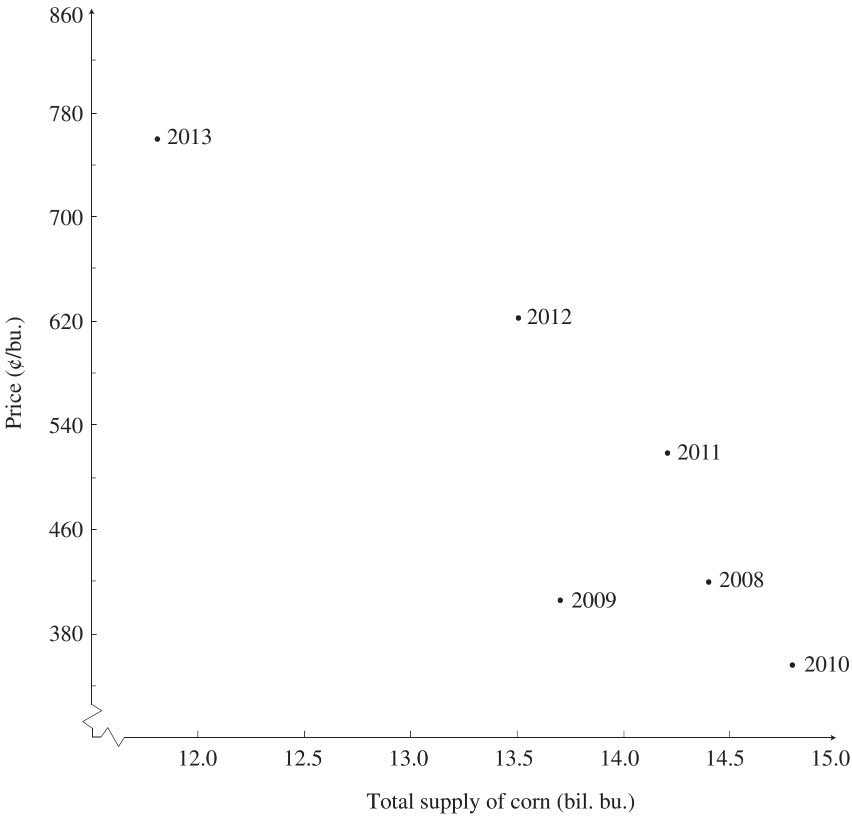



**Figure 5-2.** Price of corn versus total corn supply, United States, 2008–13

small crops results in tiny inventories, prices can become very high. In this situation, the relationship between prices in adjacent periods is reduced (or eliminated, if carryover is zero), and prices "spike," i.e., jump to very high levels relative to typical, historical experience.

The foregoing discussion has used very short-run periods of a crop year, a quarter, or a month; the principal idea is that the quantity supplied cannot change within the time period. Prices also vary from day to day and week to week under competitive conditions. These changes occur primarily in response to new information entering the market; this information may relate to current supply and demand or to expected supply and demand. Information about available inventories may be revised, expectations about the size of the new crop harvest may be changed because of new weather conditions that influence yield, a currency devaluation by a major importer may occur; etc. Current prices are linked to expected (not just current) economic conditions, and since news that influences expectations arrives in markets frequently, prices can change frequently.

## Short Run

Recall, the short run is defined as a sufficient length of time that producers can respond to price changes, and hence the short-run supply function has a positive slope. The short-run equilibrium price is determined by supply and demand schedules like those illustrated in Figure 5-3, which illustrates an equilibrium for a given set of short-run demand and supply conditions.

The conventional supply-demand diagram, depicting an equilibrium, does *not* imply that price and quantity are constant in a competitive market. An equilibrium is defined for supply and demand functions at a given snapshot in time (static). In reality, the forces determining supply and demand change with the passage of time. Hence, the equilibrium price changes through time. For example, an increase in income shifts the market demand function to the right. Likewise, the level of supply can change for a variety of reasons, such as changes in the prices of inputs.

If shifts in demand and supply are equal and in the same direction, the equilibrium price will remain constant. If the demand schedule shifts to the right more rapidly than the supply schedule, the equilibrium price will increase; however, if supply increases relative to demand, then the equilibrium price will fall (Figure 5-3).

The frequency and magnitude of price changes under purely competitive market conditions depend (1) on the frequency and magnitude of shifts in

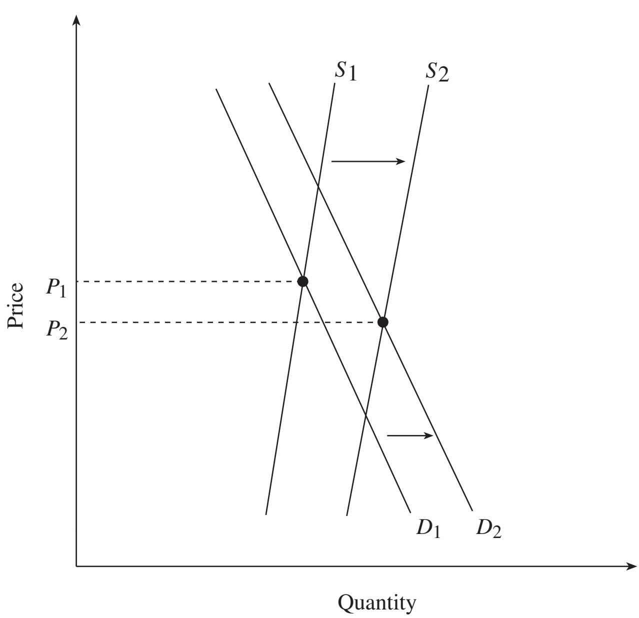



**Figure 5-3.** Increase in supply relative to increase in demand

demand and supply and (2) on the elasticities (more precisely, the slopes) of the demand and supply functions. Wide price fluctuations occur if both demand and supply are relatively price inelastic (under existing economic conditions) and if either demand or supply changes by a large amount from period to period.

Since farm-level supply and derived demand functions are usually price inelastic and since these functions can have major shifts, commodity prices tend to be variable. But different commodities have different degrees of price variability. If, for example, changes in supply and demand are relatively small, price changes will be relatively small.

The principles of short-run price determination are generally applicable to any competitive market. Instead of assuming supply is fixed, we now have a supply function, $A = S(P)$. As discussed in Chapter 4, however, the supply of an agricultural product is best modeled in terms of expected prices, and since expectations are formed before the output is realized and sold, time lags are introduced into the model. If producers use naive expectations, that is, use the actual price at the time the production decision is made as their measure of expected price, then a notation for the supply equation is $A_t = S(P_{t-1})$, where the subscripts indicate different time periods. In the simplest case, the model is completed by assuming $A_t = Q_t$ and, as above, $P_t = D^{-1}(Q_t)$.

In Figure 5-3, the supply and demand functions are assumed to be linear, and prices and quantities are simultaneously determined within the same time period. But, as just described, this need not be the case. When supply is determined by lagged price and this supply, in turn, determines price in the next period, the model is said to be recursive. The functional form may or may not be linear, and in addition, a complete commodity model probably needs to introduce equations to explain inventory adjustments, exports and imports, and other complexities of actual markets. These are more fully explored in Chapters 8, 9, 14, and 16.

Prices established under purely competitive market conditions should not be considered inherently superior and therefore sacrosanct. Society may prefer to maintain prices above or below those that would prevail with competition. A particular price can be judged "good" or "bad" only insofar as society as a whole considers the consequences of that price desirable or undesirable. Prices established in competitive market conditions do have desirable properties, not the least of which is the avoidance of problems associated with government programs designed to limit production (to produce above-equilibrium prices) or to ration available supplies (to produce below-equilibrium prices).[5]

5. Since consumers do not like either rationing or high prices, still another strategy is for the government to subsidize production. Producers receive the (low) market price plus a subsidy, and the combination provides the incentive for them to produce the quantity necessary to match the quantity demanded at a low market price. It is not clear that voter-consumers always understand the full implications of various policy choices.

Also, if prices approximate those that prevail under perfect competition, the total social benefit of the last unit purchased is just equal to the marginal social cost of producing that unit of output. However, the social product is maximized only if *all* industries operate under competitive conditions, a condition that is not met in reality. Firms operate within a framework of laws that attempt to regulate or prevent "price fixing," collusion, or other types of anti-competitive behavior. However, these laws vary from country to country, and in any case, it is clear that a variety of market structures exist in the United States within its current legal system.

## Price Determination under Monopoly

Pure, unrestricted monopoly is, in practice, a relatively rare phenomenon, and thus the discussion of pricing under monopoly conditions may seem academic. However, the theory of pricing under monopoly helps us understand price behavior by providing a contrast with competitive price behavior. And, occasionally, firms can have monopoly-like situations. A firm may develop a new product that has no close substitutes, or the firm may operate in a market whose size permits only one firm to operate profitably. Of course, limits exist on the extent of such a local monopoly, since customers may go elsewhere or, if the price is too high, another firm may find it profitable to enter the market despite the limited scale. Finally, governments sometimes permit monopoly-type pricing by groups or organizations in some sectors of agriculture; a farm organization or cooperative might be exempted from some laws so that farmers can organize to "collude" to obtain higher prices. In the United States, milk marketing orders sometimes use the principles of discriminatory pricing (discussed later).

The distinguishing feature of a monopoly is that the demand schedule faced by the monopolistic firm (or monopoly organization) coincides with the industry demand schedule. A monopolist can determine the price or the quantity produced and sold, but these two variables are not independent. If the monopolist sets the price, then consumers have the choice of the quantity, if any, to purchase at that price. This quantity is determined by the aggregate demand schedule. Alternatively, the monopolist can select the volume of output, but the demand schedule will dictate the price at which this output can be sold. In reality, of course, mistaken judgments occur; e.g., a temporary surplus in output can occur if the price set by the monopolist is higher than permitted by prevailing demand.

Assuming the monopolist's objective is to maximize profits, price will be set at the level at which marginal costs equal marginal revenue. Under monopoly conditions, however, price does not equal marginal revenue (as it does under competition). The monopolist faces a downward-sloping demand function, and price will be above marginal revenue in this situation. The relationship among marginal revenue (MR), price ($P$), and the price elasticity of demand ($E$) is expressed by the following equation:

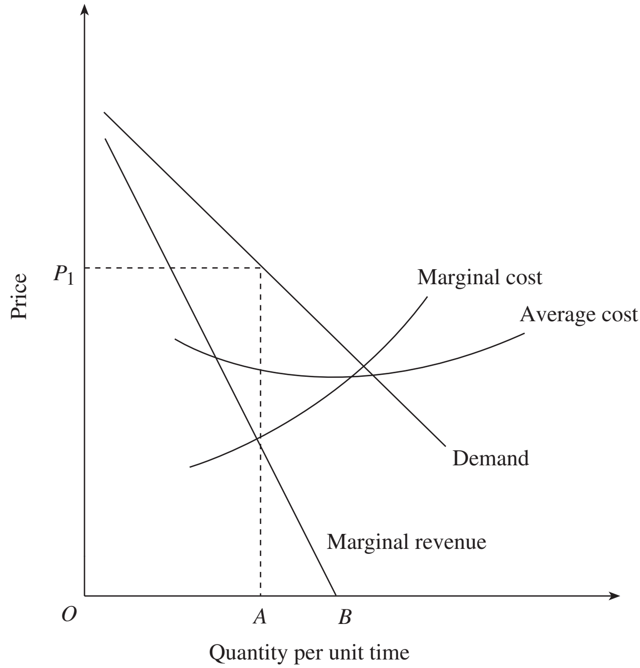



**Figure 5-4.** Price determination with a monopoly market structure

$$\mathrm{MR} = P(1 + 1/E).^{6}$$

The price elasticity is negative; so, the equation is sometimes written MR = $P(1 - 1/E)$.

If the elasticity approaches infinity (horizontal demand faced by a competitive firm), the last term approaches zero as a limit and the equation becomes MR = $P(1 + 0) = P$. Marginal revenue is positive only if demand is elastic; i.e., the absolute value of $E$ must be larger than 1. For example suppose $E = -2.0$, then MR = $P(1 - 1/2) = 0.5P$, or $P$ = 2MR. On the other hand, if $E = -0.2$, then MR = $P(1 - 1/0.2) = -4$P. It follows that a profit-maximizing monopolist is not going to set a price in the range of demand that is price inelastic.

The foregoing principles are illustrated in Figure 5-4 using a linear demand function, which gives a linear marginal revenue function. Marginal revenue becomes zero at point B on the horizontal axis at the midpoint between the origin and where the demand curve intersects the horizontal axis. Thus,

6. Total revenue is $R = PQ$, and the relation between marginal revenue and the price elasticity can be derived by taking the first derivative.

$$\mathrm{MR} = dR/dQ = P(dQ/dQ) + Q(dP/dQ) = P[(dQ/dQ + (Q/P)(dP/dQ)] = P[1 + 1/E].$$

production and sales will not exceed OB units. Given the demand and cost curves of Figure 5-4, the profit-maximizing output is OA, and the price that will clear the market of this quantity is $P_1$, as determined by the demand function.

The volume of output and the corresponding price that will maximize profits for the monopolist will change if either the demand or cost curves change. Thus, factors that shift demand or change costs result, in principle, in a price change. But, since monopolists typically have administrative control of prices, prices are less likely to change over short periods of time than under purely competitive conditions. The monopolist need not adjust to changes that are viewed as transient in nature; rather, the firm can respond to longer-term trends in demand and/or costs. That is, the monopolist can take a long-run view of profitability and not try to micro-manage changes in demand or costs that are viewed as temporary.

Also, although the monopolist has discretion in setting prices (or output), the firm must consider the long-run effect of pricing decisions on sales and net revenue. The more *potential* substitutes available in the long run, the more elastic is the long-run demand schedule facing the monopolist. A high price may provide incentives for the development of substitutes, and the greatest protection society has against the undue exercise of monopoly power is perhaps the potential availability of substitutes. In addition, monopoly pricing may result in governmental intervention under anti-trust laws, and consequently the threat of such intervention probably reduces abuses.

For the foregoing reasons, prices set by monopolists may not be at levels that maximize short-run profits, and they may change infrequently. Ultimately, however, a monopolist's prices are influenced by changes in demand and changes in costs.

## Price Discrimination

A monopolist can sometimes increase revenue and profits by charging some buyers higher prices for its product than others. The following conditions that must exist for a monopolist to gain from discriminatory pricing:

1. It must be possible to identify two or more separate groups of buyers (markets) with different price elasticities of demand and, hence, with differing marginal revenues at a particular price.
2. The markets must be effectively separated to prevent a flow of the product among markets; it should not be possible to buy the product in the low-priced market and sell it profitably in the high-priced market.

If these conditions exist, it is possible to increase revenue by charging a higher price in the market with the less elastic demand and a lower price in the market

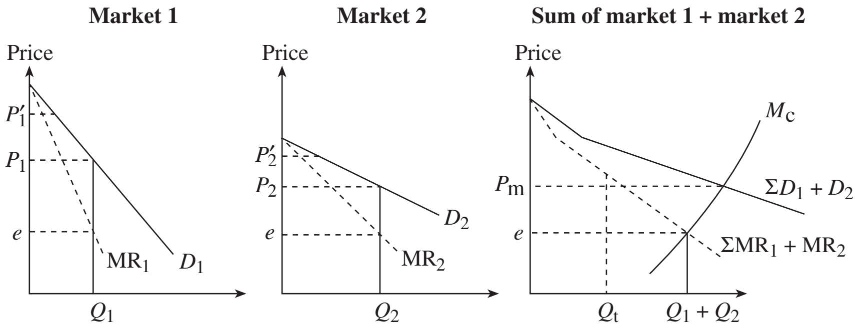



**Figure 5-5.** Optimum total output under price discrimination and profit maximizing allocation of total output between markets

with a more elastic demand.[7] Profit is also increased provided that the marginal cost of discrimination does not exceed the marginal revenue from discrimination.

The principles involved in maximizing returns by practicing discriminatory pricing can be illustrated simply by assuming just two groups of buyers with different price elasticities of demand. Initially, assuming a fixed supply of the product to be sold and that the cost of allocating the supply between markets is zero, marginal costs are zero and can be ignored. The profit-maximizing rule is to shift quantities between the two markets up to the point where the marginal revenue (MR) obtained from the sale of the last unit in each market is equal. This may be done (provided there is sufficient quantity) up to the point that $MR_1 = MR_2 = 0$. It would not, of course, pay to sell quantities such that the marginal revenue became negative in either market.

Successful price discrimination requires differences in price elasticities, which implies different demand relations and hence different marginal revenue curves. If the same price were charged to the two groups of buyers, then the marginal revenues would be different. The equalization of marginal revenues increases total revenue by transferring part of the available supply from the market with the lower marginal revenue (raising the price in this market) to the market with the higher marginal revenue (lowering the price in this market). For example, if reducing sales by 1 unit in the primary market increases revenue 25 cents per unit and if the added quantity in the secondary market decreases revenue only 8 cents per unit, there is a gain of 17 cents per unit from the transfer. It would pay to continue transferring units until the gain in revenue in market 1 equals the loss in market 2, i.e., until marginal revenues are equal.

The profit-maximizing solution for a monopoly facing two separate markets is illustrated in Figure 5-5. The optimal output, as already indicated, is given

7. Economists sometimes define three degrees of discrimination (e.g., Varian 2010: chapter 24). We discuss here third-degree discrimination, which is the degree of practical importance.

by the point at which the aggregate marginal revenue curve (which is the horizontal sum of the marginal revenue curves for the two markets) intersects the monopolist's marginal cost curve (the right-hand panel in Figure 5-5). For simplicity, the demand relations in the two markets are linear in price and quantity, and hence the elasticities change as one moves along each demand curve. But the demand equation in market 1 has a steeper slope than in market 2, and consequently, for any given quantity marketed, the demand is less price elastic in market 1 than in market 2 (recall that, since marginal revenue is positive at the prices and quantities shown, the demands must be price elastic).

Given that the demands differ in the two markets, profit is maximized by allocating total output (and establishing the corresponding prices) so that marginal revenues are equal in the two markets (point *e*). In Figure 5-5, quantity $Q_1$ is sold at price $P_1$ in market 1, and the larger quantity $Q_2$ is sold at the lower price $P_2$ in market 2. By definition, $Q_1$ plus $Q_2$ equals the optimal total output.

The total revenue for the price-discriminating monopolist ($P_1Q_1 + P_2Q_2$) is greater than it would have been in a competitive market. With competition, the market price would have been dictated by the intersection of the aggregate market demand with the aggregate market supply curve (the industry marginal cost curve), as shown in the right-hand panel of Figure 5-5. The resulting price is $P_m$. A larger quantity is sold and the average price is lower under competitive conditions than under monopoly conditions (assuming, of course, the demand and costs are identical in the two cases).

The rules for profit maximization, outlined above, can be used to determine the optimum allocation of total output among markets regardless of how much is available for sale. Actual output may deviate from the profit-maximizing output for a variety of reasons. In an agricultural context, for example, the output of individual farmers may not be controlled, even though they are organized into a monopoly-sales organization. Moreover, variations in weather conditions, diseases, or pests can result in the actual output deviating from planned output. In this case, the organization faces the task of maximizing returns subject to the quantity available for sale. Assume that in year *t* production is equal only to $Q_t$. Then the point at which $Q_t$ just touches the aggregate marginal revenue curve (right-hand panel of Figure 5-5) dictates the counterpart of point *e* (previously described). Moving horizontally, the point on each marginal revenue curve will then identify the optimum allocation of $Q_t$ between market 1 and market 2. These points of intersection are not shown (to avoid excessive clutter in the graphs), but as before, they can be obtained by drawing a line parallel to the horizontal axis, intersecting the point where $Q_t$ just touches the aggregate marginal revenue function. The optimum allocation would lead to prices $P'_1$ and $P'_2$ in the respective markets; these are higher than the original prices because total output is less.

The optimum can be determined mathematically as well as graphically. A mathematical solution requires, however, a knowledge of calculus, and some additional details are provided in the appendix to this chapter.

Since the derived demands for agricultural commodities typically are price inelastic at prevailing prices, complications arise in applying price discrimination principles. Namely, marginal revenue is negative when demand is price inelastic. Thus, if the firm (or farmer organization) were truly a monopoly, total revenue could be increased by reducing supply and raising prices in both markets. Presumably, as with a linear demand function, demand would become price elastic as prices rose. As previously noted, maximum revenue (not profit) occurs when marginal revenues are zero. If a group of farmers is producing an aggregate output such that the market clearing prices occur at levels at which demand is inelastic, then it still pays to reallocate supplies among markets as long as the marginal revenue is less negative in one market than in another.

Although price discrimination appears to be appealing for a monopoly, the conditions necessary for success are often difficult to achieve (at least at low cost). It may be difficult to effectively separate markets (i.e., prevent arbitrage among the markets). If buyers can purchase supplies in the lower-priced market and resell them in the higher-priced market, price differences obviously cannot be maintained. Furthermore, price discrimination may be prohibited by law, although in the United States marketing orders for farm commodities have been permitted to use the principles of discriminatory pricing for a few commodities.

Successful separation of markets for agricultural products is usually based on location or use of the raw product. Discrimination based on location is commonly the distinction between domestic and export markets, assuming that the demand is more elastic or less inelastic in the export market. Price discrimination based on use assumes that the derived demand for alternative uses differs. For example, milk used as fresh fluid milk can be differentiated from the same quality milk that is processed into butter, cheese, ice cream, and other dairy products, where the demand for fluid milk is more price inelastic than the demand for processed dairy products. Under the rules of marketing orders for milk in the United States, farmers have been paid a blend price, i.e., an average of different prices for milk used for fluid and for processing purposes, with the fluid price being the highest. In theory, a monopoly organization of farmers could sell a commodity at different prices for each different use; e.g., three prices for apples, distinguishing among sales as fresh fruit, sales for processed products like applesauce, and sales for apple juice. This would increase revenue, provided that each use has a different price elasticity.

Occasionally, sellers may be able to discriminate among buyers on the basis of time of sale. For example, film distributors often charge more for the first run than for the same film shown later, and theaters often charge less for matinee than for evening shows. Presumably demand is more price elastic for matinee showings than for evening showings. In theory, it would be possible to charge more for a product with a less elastic demand during a holiday period, such as cranberries for the Thanksgiving holiday in the United States, than at other times of the year.

Still another way in which a monopolist might gain from price discrimination is to charge consumers different prices based on their income levels. It is a plausible hypothesis that consumers with different incomes have different elasticities of demand for various products and services (Chapter 3). Perhaps a professional service provider (e.g., a medical doctor) can charge a higher fee to a high-income buyer, with a relatively less elastic demand, than to a low-income consumer, even though both are receiving the same service. Services are, of course, not easily transferrable among consumers. Price elasticities for foods probably vary by income group (see Chapter 3), but the potential for arbitrage between low- and high-income groups probably limits the ability of suppliers of food to discriminate by income class. Low-income buyers could profit by reselling the product, purchased at the low price, to high-income consumers facing the higher price.

The actual use of price discrimination principles in agricultural product markets is limited. One of the reasons for this is the lack of accurate information on marginal revenue and marginal cost functions. Obtaining reliable estimates of these relationships is not a trivial task. Marginal revenue functions are especially difficult to determine where markets are potentially interrelated. For example, processed products derived from lower-priced raw products may undercut sales in the fresh-product market. Raw cotton exported at low prices (relative to domestic prices) may come back to the domestic market in the form of cheap cotton products, altering the derived demand of domestic buyers of cotton and hence altering the marginal revenue function.

Even if separate markets can be identified, gains in net revenue will be small unless considerable differences exist in elasticities between markets and a relatively large proportion of total output is sold in the higher-priced market. Furthermore, benefits to producers may decline with the passage of time because of induced changes in both demand and supply. Consumers confronted by high prices may gradually shift to substitutes. That is, the long-run price elasticity in the primary market may turn out to be much larger than expected, thereby eroding the expected gains from the higher price.

Supply effects, likewise, can limit long-run gains from price discrimination. If returns are increased, farmers will expand output in the absence of supply-control programs. Larger output will depress prices in both markets or will force the monopoly organization to sell a larger proportion of the total output at the lower price prevailing in the secondary market. If demand in the secondary market becomes price inelastic as sales increase and prices decline, total revenue will decline. In any event, selling a higher and higher proportion of total output at the lower secondary-market price will reduce the average price paid to producers.

In sum, policymakers and farmers have been intrigued by the potential benefits from using price discrimination principles in selling products, but a number of practical problems have limited the success of such programs.

## Price Behavior under Oligopoly

Since an oligopoly by definition consists of a few firms, the distinguishing feature of oligopolistic markets is recognized interdependence among the suppliers. Each must take account of what other firms will do in response to a particular price (or quantity) decision. Various kinds of interrelationships can exist, and therefore, it is difficult to generalize about price behavior when the number of firms is small. A firm may be a leader or a follower, or the firms may make decisions simultaneously. Consequently, alternative models of oligopoly pricing exist, and the solutions for price depend on the assumptions made about the decision behavior of the individual firms. In this section, we outline selected alternative models, but it is beyond the scope of this book to provide a detailed discussion of each model (for more details, see intermediate microeconomics books, such as Varian 2010; or see textbooks on industrial organization, such as Carlton and Perloff 2004 or Kaiser and Suzuki 2006).

We note, first, that one possible outcome is for the sellers to collude. Collusive behavior is an example of a *cooperative game*. In this case, the optimal solution is to set a price that maximizes industry net revenue. This can be done by limiting total sales to the amount that would be produced by a monopolist, i.e., by letting total output and sales be determined by the point where industry marginal revenue equals industry marginal costs. Thus, price is determined by, and changes with, the level of market demand and costs.

Producers are still confronted with the task of deciding on market shares, and it may not be easy to agree on how the total market will be allocated. Even after agreement is reached, incentives may exist for one or more firms to cheat on the cartel agreement. The Organization of the Petroleum Exporting Countries (OPEC) agreement is perhaps the best-known cartel, involving major oil-producing countries. In this case, the agreement is among countries regarding the total quantity to produce and the allocation of this production among countries. The countries do not, however, have identical costs of oil production, and the optimal output and corresponding price are a matter of judgment. As of 2012, it appeared that OPEC members were attempting to set production at levels that would maintain petroleum prices in the $90–110 per barrel range; this level may be adjusted through time to approximate growth in inflation faced by large buyers.

Collusion to fix prices is illegal under U.S. law (Robinson-Patman Act of 1936). In October 1996, Archer Daniels Midland and three corporate co-conspirators admitted to fixing prices for lysine (an amino acid produced from corn) and allocating the volume to be sold by each firm. Connor (1997) estimates that the direct overcharges to lysine buyers during 1992–95 amounted to at least \$70 million. In his analysis of this case, Connor notes that many of the conditions for collusion existed. In particular, an oligopolistic market existed: there were four large producers in the world; they were producing a relatively homogenous product, feed-grade lysine; and the technical barriers to entry were high.

Setting aside the possibility of collusion, the literature on oligopoly price determination commonly starts with the simplifying assumptions of a duopoly (two firms) producing a homogenous product. Models of such a market determine the price and quantity for each firm (four variables). This setup gives four possible models: quantity leadership, price leadership, simultaneous quantity setting, and simultaneous price setting. As the terms imply, quantity leadership means that one firm chooses its output first and the second firm follows; in contrast, with price leadership, one firm chooses its price first and the second firm follows. The leader presumably knows that its actions will influence the choice of the follower; therefore, the leader's decision is influenced by the expected reaction of the follower. This process is sometimes called a *sequential game.*

It is also possible that, when each of the firms makes its choice about price or quantity, it does not know the choice made by the other firm. Decisions are made in the context of a *simultaneous game.* Again, it is common to differentiate between the simultaneous choice of price and the simultaneous choice of quantity. The specific model most nearly applicable to a particular oligopoly must be based on an underlying knowledge of the industry, and in any case, none of the models is likely to be exactly applicable to real-world situations. Nonetheless, basic models can be modified to give insights into the behavior of industries with oligopolistic attributes (e.g., Sexton 2000).

In the various models, the basic assumption, that individual firms maximize profits, is maintained, and thus the profit-maximizing rule is still for a firm to equate its marginal revenue and marginal costs. The issue for a firm in an oligopoly is how other firms' reactions will influence its marginal revenue (or for the oligopsonist, how other firms' reactions will influence marginal costs).

As indicated previously, a monopolist's marginal revenue (MR) equals $P(1 + 1/E)$, and marginal costs are a function of output, $Q$. The expression for MR can be generalized (e.g., Sexton 2000) for varying degrees of market power to be $\text{MR} = P(1 + K/E)$, where $P$ is the market price, $E$ is the price elasticity of market demand, and $K$ is the a market conduct parameter, sometimes called a conjectural elasticity (see below). If $K = 0$, $\text{MR} = P$, which is the competitive case, while if $K = 1$, this represents the case of monopoly. Thus, values of $K$ between 0 and 1 represent varying degrees of market power. If $E = -2$ and $K = 0.8$, then $\text{MR} = P(1 - 0.8/2) = 0.6P$, or $P = 1.667\text{MR}$. In other words, given $E = -2$, the optimal price for a monopoly ($K = 1$) would be twice the marginal revenue (as in the example in the monopoly section), while for $K = 0.8$, the optimal price would be 66.7 percent above marginal revenue.

One way to derive $K$ is to assume that the firms in an oligopolistic industry behave as in the *Cournot model.* In this model, a firm selects its optimal output conditional on its estimate of its competitors' outputs. Taking the perspective of firm 1 in a duopoly, its profit is $Pq_1$ minus the costs of producing $q_1$, where $P$ is the market price and $q_1$ is firm 1's output. The total output of the duopoly is $q_1 + q_2 = Q$. Thus, firm 1 is maximizing its profit subject to its estimate of the output of firm 2, which together with its own output gives the industry output and determines the price that will clear the market of $Q$. An equilibrium is reached when each firm chooses to produce the amount of output that the other firm expected it to produce, where both are equating marginal revenue to marginal costs.

In this context, a functional relationship can be specified between industry output and the output of the $i$th firm; this "reaction" function represents the relationship between the conjectured total industry output and the $i$th firm's own output: $Q = N_i(q_i)$. Associated with this function is a conjectural elasticity, $K_i = (dQ/dq_i)(q_i/Q)$. These equations merely formalize an idea contained in the Cournot model that a firm will base its optimal output decision on its conjecture about its competitors' outputs and hence about total industry output.

If the $i$th firm assumes that the other firms' quantities will remain fixed, then $Q$ will change only by the amount of the change in $q_i$. That is, $dQ/dq_i = 1$, and hence $K_i = q_i/Q$. In other words, the conjectural elasticity or market conduct parameter, used above, can be defined for the $i$th firm merely as its market share. If we assumed further that each firm in the industry is identical in size, then $K = q/Q$, or equivalently $K$ merely depends on the number of firms in the oligopoly. For a duopoly, $K = 1/2$, or more generally for $m$ firms in the industry $K = 1/m$. This is the assumption used by Holloway (1991) in his study of imperfect competition in the food industry.[8] If, for example, $E = -2$ and $m = 4$ ($K = 0.25$), then MR $= P(1 - 0.25/2) = P(1 - 0.125) = 0.875P$, or $P = 1.143$MR.

This approach to oligopoly price determination has the advantage of being analytically tractable and therefore convenient to use in simulations (e.g., Holloway 1991; Sexton 2000). Also, an analogous approach can be taken for evaluating prices paid by oligopsonists that exercise market power on the buying side (Sexton 2000). The foregoing model implies, however, that prices will be approximately at competitive levels for a rather modest number of firms in the industry. For $m = 20$, $K = 0.05$. Thus, if $E = -2$, then MR $= P(1 - 0.05/2) = P$

8. Varian (2010: chapter 26) derives a similar result as follows. Marginal revenue for the $i$th firm can be defined as $P + (dP/dQ)q_i$, where $P$ is a function of $Q$ (and marginal cost of the $i$th firm depends on its level of output $q_i$), $Q$ is industry output, and $P$ is the price received by all firms in the industry. The marginal revenue expression can be rewritten as:

$$P[1 + (dP/dQ)(Q/P)(q_i/Q] = P[1 + s_i/E],$$

where $s_i$ is $q_i/Q$. Note that this derivation makes clear that, under the Cournot model, $E$ is the market-level price elasticity of demand.

$(1 - 0.025) = 0.975P$, or $P = 1.026$MR. In other words, in this example with 20 firms, the market price is less than 3 percent above marginal revenue.

Of course, this result also varies with the assumed price elasticity of demand. If demand is price inelastic, say –0.5, and assuming 20 firms, the result is MR $= P(1 - 0.1) = 0.9P$, or $P = 1.11$MR. In this case, the optimal price is 11 percent above marginal revenue. Demand is probably price inelastic for some food products at retail.

An alternative model, sometimes called *Bertrand competition,* assumes that firms set their prices and let the market determine the total quantity sold, $Q$. In this case, when a firm chooses its price, it must estimate the prices set by other firms in the industry. In other words, simultaneous price setting is assumed, and in a duopoly, the problem is to find the pair of prices such that each price is the profit-maximizing choice for the respective firms, given the choice made by the other firm. If the assumption that the firms are selling a homogenous product is maintained (as, indeed, we have been assuming), then there is a surprising result. Namely, the Bertrand equilibrium equals the competitive equilibrium, where the market price equals marginal cost.

The Bertrand result may not be intuitively satisfying, but it is analytically valid. Indeed, the foregoing models do little to narrow the range of possible results, implying that prices established in industries characterized by a few sellers of a homogenous product can range from the competitive level (the Bertrand result) to the pure monopoly level (the collusive behavior result). The results merely establish that optimal price levels depend on demand and costs, and will change as costs and demand change. Within this framework, prices will further depend on the nature of the competitive interactions among firms. Since firms have control of quantity or price, however, price changes will not be frequent. There are costs associated with repricing (see Chapter 6), and firms do not wish to make small adjustments in prices due to changes in demands or costs that may be transient in nature.

Another model of oligopolistic behavior is also consistent with infrequent price changes. In this model, the current price is maintained if each firm in the industry believes that a price reduction will be matched by other firms. In contrast, an attempt to raise prices may not be matched by other firms. If so, the firm seeking to raise prices could suffer a substantial loss of market share. These assumptions are the basis for what is commonly referred to in the economics literature as a *kinked demand curve.* Such a curve is illustrated in Figure 5-6. The kink (change in slope) occurs at the established price, say $P_a$. Above this price, the demand curve confronting each firm is assumed to be relatively price elastic because other firms in the industry are not expected to go along with a price increase. On the other hand, if each firm assumes others will match any price reduction it initiates, the increase in sales will be limited by the shape of the aggregate demand curve. The demand faced by the individual firm at prices below the current level will be relatively less elastic.

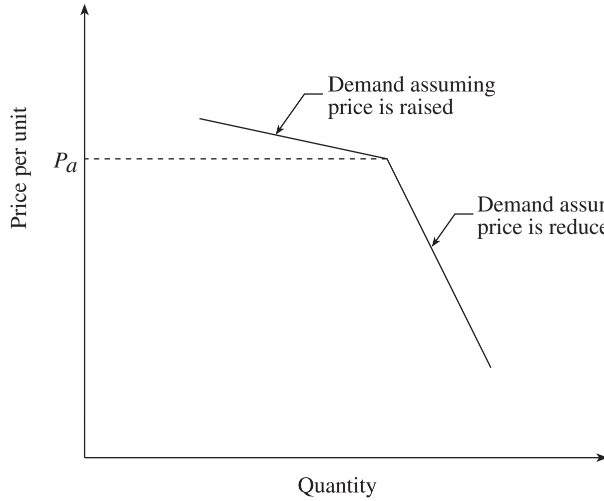



**Figure 5-6.** Demand confronting an individual firm that initiates a price increase or decrease from the established price

The kink in the firm's demand curve creates a discontinuity in the firm's marginal revenue function. Thus, it is possible for the marginal cost curve to shift without causing the price to change. This model, like the others discussed earlier, may have limited applicability, but it is consistent with the desire of firms to avoid "price wars," especially if they are selling a relatively homogenous product.

The analysis of price determination in oligopolistic markets is complicated by additional factors. First, the size and diversity of firms can make a difference in cost structures. A large firm may achieve economies of scale, size, and scope that cannot be achieved by a small firm. Economies of scope occur if a firm is producing more than one product and if this allows the firm to produce the products at lower cost than if each of the products was being produced by separate firms. Economies of scale occur when increasing the scale of the firm reduces costs; this would occur, assuming fixed input prices, if the doubling of all inputs more than doubled the output (for an in-depth discussion of economies of scope and scale, the reader may consult an intermediate microeconomics textbook). Thus, a few large firms may have a lower cost structure than would exist in many small firms. If so, prices in an oligopoly may not be higher than those in a competitive market.

Second, the products of an oligopolistic industry may be differentiated, perhaps with well-established brand names, as in the auto industry. The differentiation is likely to permit some degree of control of prices, although, if the brands of different firms are close substitutes, the firms still must be conscious of the effects of their price and output policies on the decisions of other firms in the industry, and vice versa. A plausible hypothesis is, however, that non-price competition in such industries has the effect of raising costs. Packaging, advertising and promotion, executive compensation, and other costs are probably larger than they would be in a purely competitive industry.

In the food marketing sector, some evidence supports both of the foregoing points, which is to say the evidence depends on the particular industry under analysis. The growing concentration in the U.S. beef-processing sector has apparently improved economic efficiency and lowered costs (Azzam and Schroeter 1995; Morrison-Paul 2001). This conclusion, however, does not necessarily contradict the finding that "welfare losses are higher in industries that also have higher advertising intensities . . . " (Bhuyan and Lopez 1998: 263; also see Cotterill and Franklin 1999). The food marketing sector in modern economies is complex, and this point is discussed further in the next chapter.

## Price Behavior under Monopolistic Competition

Monopolistic competition is essentially a competitive market structure with product differentiation. Many variants of a product exist that are close but imperfect substitutes, e.g., livestock feed. Thus, the demand curve facing a firm is likely very elastic, although not perfectly elastic. Firms producing similar but differentiated products are assumed to have similar costs, but not necessarily identical costs. The varying attributes of the differentiated products can create small differences in costs, but firms with high costs relative to others in the industry presumably will not be able to survive the competition.

Assuming profit maximization, the firm's optimal price is determined by equating marginal revenue and marginal costs, and this price is likely to approximate the average total costs. With monopolistic competition, it is assumed that the costs of entry and exit are small (as in the competitive model). Under such circumstances, any excess profits that might exist temporarily will help attract new entrants into the industry, and this will have the effect of shifting the demand functions for existing firms to the left. In other words, competition will tend to keep prices near the minimum average total cost level.

Still, prices could be higher with monopolistic competition than with pure competition, particularly in the short run. This occurs if the cost structure is higher with monopolistic competition. As with an oligopoly, producers of differentiated products may have larger costs related to added packaging expenses, advertising, etc. They also may attempt to increase market shares by making changes in design or ingredients, thereby providing the product with new and slightly different attributes.

Markets may have varying degrees and types of product differentiation. Consumers have choices among many brands and types of manufactured products such as soaps, toothpaste, and other toiletries. These products are produced by firms of varying sizes, and the differences in attributes may be modest and difficult to evaluate. In contrast, the products in a farmers' market come from numerous small firms, and the vegetables, fruits, baked goods, and other products may be very similar. Nonetheless, sellers are still able to differentiate their goods to some degree. Clearly, sellers in a farmers' market have limits on their ability to vary prices, but they often can establish a premium for their product based on its attributes (real or imagined). This premium may be necessary to cover the higher cost of providing the service or attribute that differentiates it from its competitors. For example, some consumers may be willing to pay a higher price for organically grown vegetables than for conventionally produced vegetables.

A part of imperfect competition arises from imperfect knowledge about attributes.[9] It can be costly for consumers to evaluate the differentiated products. Thus, price differences for essentially identical products may persist because consumers are not fully informed about the similarities.

## Concluding Remarks

It is clear that prices in competitive commodity markets change more frequently and by larger amounts (over a given time interval) than do prices established in oligopolistic or monopolistic markets. Competitive markets have flexible prices; they respond quickly to new information about current and expected economic conditions. Markets with monopolistic attributes have relatively inflexible prices; they respond more gradually to changing economic conditions.

It is less clear whether the average price level would differ between a competitive and an oligopolistic market (if one could observe the same product being priced under the two market structures). In general, economists prefer competitive pricing to monopoly pricing because, in theory, monopolies lead to smaller outputs and higher prices than do competitive markets. But, in practice, large firms may achieve economies of scale and be able to undertake research and development that improve product quality and lower costs.

In sum, the issue is whether the benefits associated with size, when they exist, are more or less offset by higher costs and non-competitive profits. Therefore, it is not surprising that generalizations are not possible on purely theoretical grounds. One must undertake empirical analyses of specific markets.

## Appendix: Price Discrimination

The principles of price discrimination can be illustrated by assuming two separate markets with linear demand functions. The two demand equations are written in inverse form, and the subscripts identify the markets. We will also

9. The pricing of commodities with varying characteristics is covered in detail in Chapter 7.

assume that the total quantity to be allocated among the markets is fixed, i.e., predetermined, by prior decisions of the suppliers, as could be the case in an agricultural market. Thus,

$$P_1 = a - bQ_1$$

$$P_2 = c - dQ_2, \text{ and}$$

$$Q_1 + Q_2 = Q.$$

Total revenue is $R = R_1 + R_2 = P_1Q_1 + P_2Q_2$. The optimization problem can be characterized as maximizing $R$ subject to the restriction that $Q_1 + Q_2 = Q$. In this characterization, allocation costs are zero, and the entire quantity available, $Q$, is to be allocated.

Substituting the $P$'s out of the revenue function,

$$R = (a - bQ_1)Q_1 + (c - dQ_2)Q_2,$$

the objective function to be optimized can be written

$$L = aQ_1 - bQ_1^2 + cQ_2 - \mathrm{d}Q_2^2 + \lambda(Q_1 + Q_2 - Q).$$

The optimal allocation of $Q$ between the two markets is obtained by solving the first-order conditions for an optimum for the $Q$'s. The first-order conditions are obtained by taking the partial derivatives of $L$ with respect to $Q_1$, $Q_2$, and $\lambda$. These are

$$a - 2bQ_1 + \lambda = 0$$

$$c - 2dQ_2 + \lambda = 0$$

$$Q_1 + Q_2 - Q = 0.$$

With some algebra, the optimal $Q$'s are

$$Q_1{}^* = [d/(b+d)]Q + (a-c)/(2b+2d), \text{ and}$$

$$Q_2{}^* = Q - Q_1{}^*.$$

The optimal quantities to be allocated to each market depend on the parameters of the demand equations and on the size of the crop, $Q$, to be allocated. Given the optimal quantities, the demand equations can be used to obtain the corresponding prices, $P_1{}^*$ and $P_2{}^*$. The total revenue corresponding to these prices and quantities is the maximum revenue for the given conditions.

A variety of extensions of the model are possible. The cost of allocation from market 1 to market 2 might be modeled as equal to $kQ_2$. Hence, the net revenue for market 2 is

$$R_2 = P_2Q_2 - kQ_2.$$

The objective function can be redefined accordingly, and the optimum obtained.

If the markets for $Q_1$ and $Q_2$ are related (i.e., the two quantities are substitutes for each other in the respective markets), then the appropriate demand equations are

$$P_1 = a - bQ_1 - eQ_2, \text{ and}$$

$$P_2 = c - fQ_1 - dQ_2.$$

Note, the parameters $a$, $b$, $c$, and $d$ in this set of equations will not be same values as in the previous set of equations. The ability to substitute will change the parameters. For further discussion of these topics, the reader may consult Waugh (1964: Appendix 6).

A practical problems in implementing a discrimination scheme is that the demand parameters are unknown and must be estimated from market data. Obtaining high-quality estimates of such parameters is often difficult. These difficulties are discussed further in Chapters 14 and 15.

## References

Azzam, A.M., and J. R. Schroeter. 1995. "The Trade-off between Oligopsony Power and Cost Efficiency in Horizontal Consolidation: An Example from Beef Packing," *Am. J. Ag. Econ.* 77:825–836.

Bhuyan, S., and R. A. Lopez. 1998. "What Determines Welfare Losses from Oligopoly Power in the Food and Tobacco Industries?" *Ag. & Res. Econ. Rev.* 27:258–265.

Carlton, D. W., and J. M. Perloff. 2004. *Modern Industrial Organization*, 4th ed. Reading, MA: Addison-Wesley.

Connor, J. M. 1997. "The Global Lysine Price-Fixing Conspiracy of 1992–1995," *Rev. of Ag. Econ.* 19:412–427.

Cotterill, D. W., and A. W. Franklin. 1999. "An Estimate of Consumer Benefits from the Public Campaign to Lower Cereal Prices," *Agribusiness* 15:273–287.

Deaton, A., and G. Laroque. 1992. "On the Behavior of Commodity Prices," *Rev. of Econ. Studies* 59:1–23.

Holloway, G. J. 1991. "The Farm-Retail Price Spread in an Imperfectly Competitive Food Industry," *Am. J. Ag. Econ.* 73:979–989.

Kaiser, H. M. and N. Suzuki, eds. 2006. *New Empirical Industrial Organization and the Food System*. New York: Peter Lang Publishing.

Morrison-Paul, C. J. 2001. "Market and Cost Structure in the U.S. Beef Packing Industry," *Am. J. Ag. Econ.* 83:64–76.

Sexton, R. J. 2000. "Industrialization and Consolidation in the U.S. Food Sector: Implications for Competition and Welfare," *Am. J. Ag. Econ.* 82:1087–1104.

USDA. 2012. *2011 P&SP Annual Report*. Washington, DC: Grain Inspection, Packers and Stockyards Administration.

Varian, H. R. 2010. *Intermediate Microeconomics, A Modern Approach*, 8th ed. New York: W.W. Norton & Co.

Waugh, F. V. 1964. *Demand and Price Analysis: Some Examples from Agriculture*. USDA Tech. Bul. 1316.

Westcott, P. C., and L. A. Hoffman. 1999. *Price Determination for Corn and Wheat: The Role of Market Factors and Government Programs*. USDA Tech. Bul. 1878.

PART II

# PRICE DIFFERENCES AND VARIABILITY

This section focuses on the factors that help explain price differences associated with the provision of marketing services, with quality attributes, with location, and with the passage of time. Differences between prices paid to farmers and those paid by consumers, commonly referred to as *marketing margins,* are analyzed in Chapter 6. This chapter also covers the transmission of price information among farm, wholesale, and retail markets. Agricultural product prices have unique behavioral characteristics, and the behaviors associated with quality and location attributes are discussed in Chapters 7 and 8. Systematic behavior with the passage of time, such as seasonal patterns, is covered in Chapter 9. In Chapter 10, changes in the general level of all farm prices, including the discussion of changes in prices received by farmers relative to those paid by farmers, are considered.

CHAPTER 6

# Marketing Margins

Price theory in its simplest form assumes that many buyers and sellers meet directly. Equilibrium prices are determined by the market demand and supply schedules, which aggregate the behavior of these buyers and sellers. Although farmers and consumers sometimes do meet directly in farmers' markets, most foods move through a complex processing/distribution system. Indeed, the product on the grocery store shelf may be far different than the original farm commodity; the commodity is just one of many inputs used to produce the retail product. Thus, it is not surprising that the price of bread in a bakery is much higher than the value of the flour (and the price of wheat paid to farmers). There is, however, a strong interest by farmers, policymakers, and consumers in the connection between farm prices for commodities and retail prices for foods.

The difference between the price received by farmers and that paid by consumers is a marketing margin, also called a price spread. Both producers and consumers have expressed concerns about the size of margins and have asked why marketing margins change. Moreover, if a marketing margin increases, what is the incidence of the change on farm and retail prices? Consequently, substantial research has been done in agricultural economics on questions related to price differences between the farm and retail levels. Are marketing margins too large? Why do margins differ among products? How have they changed with the passage of time? Is the size of the margin influenced by the size (hence, price) of the crop? If a marketing cost, such as a wage rate, increases, does this result in a higher retail price or a lower farm price, or some combination of the two? What is the impact of market power on marketing margins?

The principal objective of this chapter is to provide tools of analysis and evidence that will help answer such questions. Attention is focused on spreads among prices at different levels of the marketing chain. Most of the analysis is based on the marketing margin for a single product and assumes a purely competitive market structure. The potential effects of imperfect competition are, however, discussed, especially the empirical evidence on the effects of market concentration. Since marketing margins are about prices at different levels of the marketing chain, a natural question arises about how information is transmitted among these levels. Thus, this chapter also introduces the topic of price transmission. The chapter concludes with a discussion of the relationship between own-price elasticities at different market levels.

## Models of Margin Behavior

### A General Approach

Concepts developed in prior chapters can help model the behavior of marketing margins, and to keep concepts clear, we introduce some notation. The behavior of consumers is represented by the primary or retail demand function, $Q = D(P, N)$ where $Q$ is aggregate market quantity for a single product, $P$ is its retail price, and $N$ is a determinant (shifter) of the level of the demand function. The equation is written in general functional notation.

The quantity, $Q$, is assumed to be produced by many competitive marketing firms using the agricultural input, $A$, and a marketing input (say, labor), $B$. These firms are profit maximizers and use the same production technology. Thus, the concepts developed in Chapter 4 to model farmers' production and supply behaviors can also be used to model the behavior of marketing firms. Letting lowercase letters represent an individual firm, the firm's production function can be written $q = f(a, b)$; for $n$ identical firms, the total output is $Q = nq$ and the total input uses are $A = na$ and $B = nb$. The quantities of $a$ and $b$ used by each firm will depend on the prices of these inputs as well as the price of the output. As discussed in Chapter 4, one can obtain the derived demands for $A$ and $B$ as functions of the prices ($P$, $P_A$, and $P_B$), as well as obtaining the supply of $Q$ as a function of these same prices.

To complete the model, a supply function is specified for each input. These are written in general notation as $P_A = A^{-1}(A, W)$ and $P_B = g^{-1}(B, T)$. The subscripted prices are for the respective inputs, and each equation is written in inverse form, i.e., with price on the left-hand side. The letters $W$ and $T$ represent factors that shift the level of the respective functions. For example, $W$ might be a measure of the weather, where good weather increases commodity supply; T might represent, for example, improved production technology, and likewise, an increase in $T$ raises the supply of $B$.

In sum, the model consists of six equations that explain six variables: three prices ($P$, $P_A$, and $P_B$) and three quantities ($Q$, $A$, and $B$). The six variables are determined simultaneously. From our setup, the six equations are the retail demand function for $Q$; the production function for $q$, where $Q = nq$ (or the supply function for $Q$); the derived demands for the two inputs, $A$ and $B$; and the two supply functions for $A$ and $B$. (Gardner 1975 is a basic reference.)

From the perspective of farm-to-retail margins, interest centers on $P$ and $P_A$. The margin is $M = P - P_A$. It is also sometimes stated as a ratio $P/P_A$. Given the model, we can ask, how does the margin change as the variables $N$, $T$, and $W$ change? For example, if good weather (the variable $W$) shifts the supply function for $A$ to the right, more $A$ is supplied to the market and the price of $A$ decreases. Consequently, it is profitable to use more $A$ to produce more $Q$, and the price of $Q$ will decrease. At the same time, additional $B$ is required to produce more $Q$, and with an upward sloping supply function for $B$, the price of $B$ will increase. Thus, the margin, $M$, should increase (Gardner 1975: 402). Farmers are often annoyed when farm prices fall by more than retail prices, but this model shows that it can happen in a competitive market. The precise magnitude of the change in the margin will depend on the magnitudes of the parameters (elasticities, in the Gardner specification) of the equations.

The foregoing model is a simplified version of the marketing system. It assumes that the firms are producing one product, $Q$, with identical technology. All the inputs used in manufacturing, packaging, distributing, and retailing the products are aggregated into one quantity, $B$, with price $P_B$. Nonetheless, it is difficult to visualize the interactions of six variables in graphical terms (see Wohlgenant and Haidacher 1989: figures 2 and 3), and we turn to some simpler alternatives in the next subsection. The simplifications should help us understand the role of selected assumptions and thereby understand the issues involved in modeling marketing margins.

### Specific Model

The foregoing model can be modified as follows. First, the supply of the farm commodity, $A$, is treated as perfectly price inelastic; this is the very short-run supply function of Chapter 4. For an annually produced crop, the very short run is 1 year and the crop size is fixed for the year; of course, the size of the crop can change (the function can shift) from year to year. Second, the retail product, $Q$, is assumed to be produced as a fixed proportion of $A$. For example, a choice-grade steer weighing 1,150 pounds is transformed into 479 pounds of salable beef at retail; i.e., each pound of live animal produces 0.41667 pound of retail meat. Mathematically, $Q = kA$, where in the beef example $k = 0.41667$. Sometimes, the same fixed-proportions assumption is made for the marketing input $B$ . Here, we assume that the relationship of $Q$ to $B$, the marketing input, is a continuous function. Thus, at the individual firm level, the total variable costs (TVC) of producing $q$ from $b$ can be written as a continuous function of $q$; TVC $= P_B b = C(q)$. Several alternative assumptions about costs are discussed later in the chapter (also see Helmberger and Chavas 1996: 134f).

A third common simplification is to treat the price of $B$ as a given (exogenous) from the marketing firm's perspective. That is, the supply of $B$ is perfectly price elastic; the marketing firm can buy any quantity of $B$ at the given price $P_B$. Of course, this still permits $P_B$ to change with the passage of time; the assumption merely implies that marketers can obtain their desired quantities of the input at the then-prevailing price.

Given these assumptions, the profit function for a representative marketing firm can be written as $Pq - P_A a - P_B b$ (recalling that $a$ and $b$ are the firm-specific quantities of the inputs). From the fixed-proportions assumption, $q = ka$, and for simplicity, let $k = 1$. Thus, the profit equation of the representative firm can be written $(P - P_A)q - P_B b$. The term in parentheses is the marketing margin. The cost function, $C(q)$, associated with the use of $b$ can be written in a variety of forms. The simplest case is where total variable costs are a linear function of $q$, $C(q) = cq$, where $c$ is a constant. This specification implies that, when $q = 0$, no $b$ is used and costs are zero.

Using the three assumptions given above and assuming further that retail demand is linear and that marginal costs are constant, a model of the corresponding marketing margins can be created, shown graphically in the upper panel of Figure 6-1. This model is in terms of the aggregate market relationships ($Q = A = na$). If the model is modified to allow for quadratic variable costs and hence a linear marginal cost function, we obtain the equations represented in the lower panel of Figure 6-1. The algebra is shown in the appendix to this chapter.

In the upper panel, the difference between the retail and farm prices depends on the constant marginal cost of the marketing input. Shifts in the supply of the farm commodity do not influence the margin; the farm and retail prices change by the same amount as the level of $A$ shifts. Likewise, a shift in the retail demand function will result in a corresponding shift in the derived demand function, and the margin does not change. Margins do change when the marginal cost of the marketing input, $B$, changes, and this in turn depends on the price of $B$. Since the price of $B$ is exogenously given in this model, the margin is determined solely by this price. For this simple model, $M = f(P_B)$. Thus, for example, if wages in the processing or retail food sectors rise, margins will rise.

In the variant of the model shown in the lower panel of Figure 6-1, where $C(q)$ is a quadratic and marginal costs vary linearly with $q$ (hence, $Q$), shifts in the supply of $A$ will influence the margin. A larger supply of $A$ reduces both the retail and farm prices, but because marginal costs increase as $Q$ increases, the margin widens. Note that the larger margin resulting from the increase in the supply of $A$ is not a function of a larger price of $B$, which is assumed constant in this illustration. Rather, the rising marginal cost is a consequence of the assumed underlying production technology and the implied diminishing returns associated with the increased use of input $B$ in the production of $Q$ (Chapter 4).

With the passage of time, the price of $B$ can change. Hence, with a quadratic cost function, $M = f(P_B, Q)$. Recalling $Q = kA$, the units on the horizontal axis in Figure 6-1 are adjusted to retail equivalent values. The model implies that the margin will vary directly with changes in the price of marketing inputs and with the quantity supplied (and demanded) of $Q$. For example, even if $P_B$ does

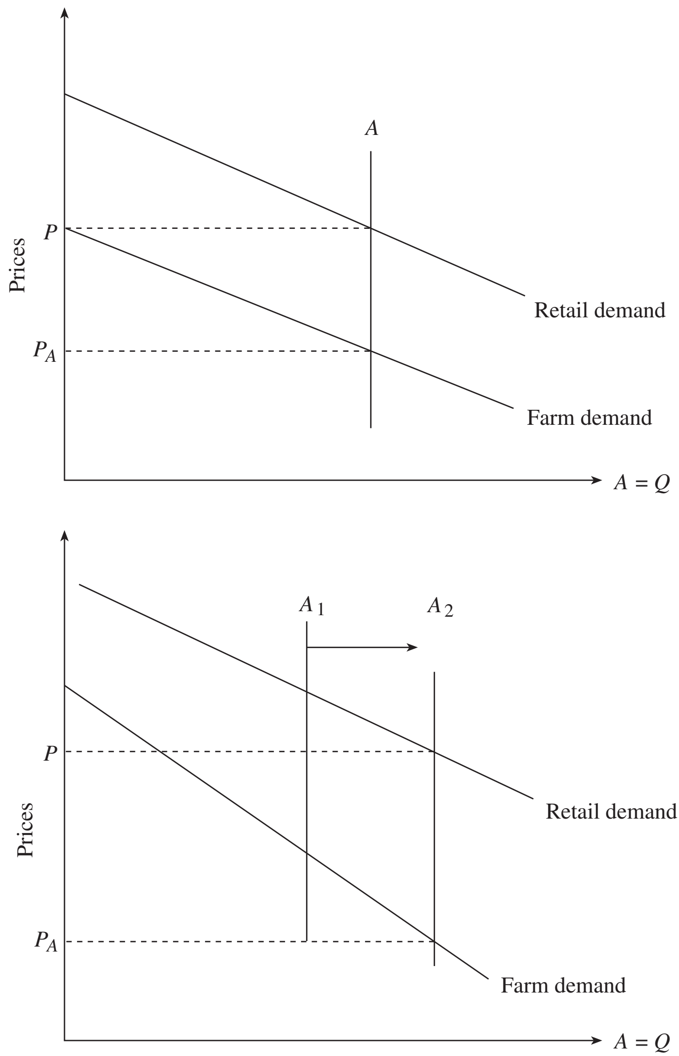



**Figure 6-1.** Alternative models of margin behavior



not change, the margin will decrease if the quantity, $Q$, decreases. Unlike the more general model in the previous section, shifts in retail demand do not influence the margin.

## Other Variants of the Conceptual Model

Many marketing margin models exist in the literature. We mention a few here to emphasize the range of possibilities. In one alternative, Azzam (1996) maintains the fixed-proportions and exogenous input-price assumptions but allows the supply of $A$ (hence, $Q$) to vary with the price of $A$, a positively sloping supply function. In this case, Azzam shows that the marketing margin will vary with the level of $Q$, even if the marginal cost associated with $B$ is a constant. If this model is appropriate, the margin depends on the magnitudes of both the input prices and quantity sold. This gives an additional possible justification for including $Q$ in the model, but it implies that the prices $P$ and $P_A$ are simultaneously determined with $Q$.

In another model, Holloway (1991) modifies the Gardner (1975) specification to allow for imperfect competition. In Holloway's model, each firm is assumed to face the same production function, $q = f(a, b)$, but the number of firms is varied from one (monopoly) to many (perfect competition). In the monopoly case, the firm produces the total industry output $Q$. He assumes, however, that the supply of $B$ is perfectly elastic and that the supply of $A$ is perfectly inelastic (analogous to the discussion in the previous subsection). As a consequence, $M = f(N, P_B, A)$, where $N$ represents retail demand shifters; i.e., shifts in retail demand also influence the size of the margin.

Wohlgenant and Haidacher (1989) provide another model that can be used for analyzing the competitiveness of the food marketing sector. They make the point that even if individual firms produce retail products in a fixed proportion to the farm input, the heterogeneity of these firms means that the aggregate output of the industry should be viewed as produced in variable proportions. In contrast to Gardner and Holloway, the Wohlgenant-Haidacher model permits each marketing firm to have a different production function.

We discuss the possible effects of reduced competition on margins in a subsequent section, but note in passing that Holloway's empirical results suggest that the departures from competition in the major food markets of the United States were small, at least for the 1955–83 period used in the analysis. Also, a study (Reed and Clark 2000) based on the Wohlgenant and Haidacher model, supports this conclusion.

A third category of models includes either farm ($P_A$) or retail prices ($P$) as explanatory variables in the margin equation. A justification for including retail price as a variable is provided by Lyon and Thompson (1993). If the variable-proportions (rather than fixed-proportions) assumption is used, then, as in Chapter 4, the derived demand for $A$ can be written as $A = f(P_A/P, P_B/P)$. If, however, $A$ and $P_B$ are assumed to be exogenous variables, then from a statistical viewpoint it is preferable to rewrite the model as $P_A/P = f^{-1}(A, P_B/P)$. After some algebra, the general expression for the margin is $M = f^n(P, PA, P_B)$. This model implies that one of the explanatory variables is the product $PA$, called an interaction term. The conceptual model also implies that $M$ and $P$ are simultaneously determined.

In still another model, Heien (1980) hypothesizes that retail food prices can be viewed as a markup over costs, i.e., over wholesale prices. Retail food stores sell a large number of different products, which implies that a fairly simple approach is required for establishing prices, and managers of food retailing firms do have the discretion to set prices via some rule(s). Since the various retailers face the same or approximately the same wholesale prices, they know the wholesale costs of their competitors. Thus, the competitive position of particular retail stores (and chains) depends on their pricing rules.

In this context, Heien suggests that it is reasonable to think that a retail price is based on a markup rule over the wholesale price. The same argument is not applicable for farm prices to wholesale prices, but some empirical evidence, including that in Heien, suggests that causality runs from farm to wholesale prices. Therefore, still another potential model specification makes the retail price or margin a function of the wholesale (or farm) price. The model also includes the costs of the marketing inputs, like $P_B$ above. In sum, this model is $M = f^m(P_A, P_B)$.

The model illustrated in the top panel of Figure 6-1 is consistent with a simple markup rule. The retail price equals the wholesale price plus a constant, which is based on the cost of the marketing inputs. As is shown later in this chapter, other markup rules can be used, and the rule used influences the relationship between elasticities at the retail and farm levels.

## Variables in Empirical Analyses

The research literature contains a large number of empirical analyses of marketing margins for agricultural products. Individual studies start with a theoretical foundation, such as those discussed above, but they typically add other variables to the statistical model. These variables depend on the research problem and the particular characteristics of the commodity market under analysis.

One research question is, does increasing concentration in the food marketing system increase marketing margins? In a study of the lamb market, the question was, has increased concentration in the meat-packing sector increased the margin for lamb (Capps, Byrne, and Williams 1995)? The analysts made the wholesale-farm price spread a function of a four-firm concentration ratio, a time trend, and variables suggested by theory. The statistically most important variable in the model was a measure of marketing costs ($P_B$). The coefficient of the four-firm concentration ratio was positive, but with only marginal statistical importance. The trend variable, which is intended to measure the effects of other, unmeasurable factors, like improving technology in meat processing, had a negative coefficient but also had little statistical importance. Thus, weak empirical evidence exists that the margin widened as meat-packer concentration increased but that the margin decreased for other reasons associated with the passage of time.

Price risk is another factor that may influence marketing costs. A processor, for example, might buy a farm commodity at a known current price but be uncertain about the price at which the processed product will be sold. Brorsen et al. (1985) suggested that this may be the case for flour millers. Input prices, such as wheat, energy, and labor, are known at the time of processing, but the miller may not have forward contracts for all the flour that is produced. If so, risk is associated with unpredictable variability in forthcoming flour prices. As discussed in Chapter 4, risk is similar to a cost of production for the risk-averse producer. Thus, an increase in price risk is expected to increase the marketing margin, other factors held constant. Brorsen et al. (1985) considered the price spread between the farm price of wheat and the wholesale price of flour, and they modeled the spread as a function of the quantity of wheat consumed for food, milling costs, and a measure of the riskiness of flour prices. They found that all three of these variables had a positive effect on margins. Larger costs of milling, larger consumption of wheat, and larger risks all meant larger margins between the farm price of wheat and wholesale price of flour.

Other research has considered the effects of particular events on margins. The event might be the elimination of a particular government program or regulation. Or it might be unionization in a particular marketing sector, etc. It is possible to construct variables that represent the presence (or absence) of the event and, hence, to estimate the effect of the event, net of other factors influencing the margin (e.g., Thompson and Lyon 1989).

## Summary of Models

The conceptual literature suggests that marketing margins depend on, among other things, the supplies and prices of inputs to the marketing sector. The factors affecting these supplies may be measured in a variety of ways, but it is common to treat the quantity of the agricultural input and the prices of the marketing inputs as exogenous explanatory variables. These models make the assumption that the retail commodity is produced as a fixed proportion of the farm commodity input. The variable-proportions models imply that the margin will also depend on factors that shift the retail demand function. Whether or not the fixed proportions assumption is appropriate is controversial, but it is important to understand that two quite different approaches exist to modeling marketing margins. Moreover, analysts have used many variants of these models.

Since the marketing system is complex, variables in addition to prices and quantities may be important in models of margins. These may include improvements in technology in the marketing system, increased concentration of firms in the marketing system, varying levels of price risk, and perhaps other factors like the elimination of a government program. Consequently, although the margin literature is based on well-accepted economic principles, a large number of models exist in the literature that differ in the individual assumptions that they make about the nature of the food marketing system. These differing assumptions can, in turn, result in differences in empirical results concerning the behavior of marketing margins.

## Empirical Measures of Margins

The U.S. Department of Agriculture (USDA) publishes two types of measures of marketing costs: farm-retail price spreads for individual foods and the food dollar series, which is composed of three primary data sets that include aggregate food marketing costs. Further, the data for individual price spreads have been aggregated into an index of price spreads for a fixed market basket of foods. (The aggregate market-basket series was discontinued after 2010, but data are still published for sub-aggregates like a milk and dairy basket.) Farmers' share statistics can be estimated for either the price spread or the total marketing bill series.

The farm-retail price spreads are calculated for selected foods produced from farm commodities of domestic origin. The foods used in the computations tend to be basic or common products, such as a dozen eggs or a 1-pound loaf of bread, where the computations can rely on readily available prices. The retail prices are for precisely defined products so that the estimated spread is not influenced by changes in the product's characteristics. For a composite commodity, like choice beef, the retail price is a weighted average of prices of the component cuts of beef.

The margin computation is made by subtracting the farm value of the raw commodity (adjusted for the value of any by-products) from the price of the corresponding retail product. The quantity of the commodity embodied in the typical consumer unit must be estimated and then valued at the average farm price. This procedure assumes that the retail product is produced in a fixed proportion from the farm commodity, an assumption made in some of the conceptual models discussed in the previous section. The spreads are estimated for months as well as annually.

Price spreads, such as those for beef and milk, have been used to analyze public policy issues such as New York State's anti-price-gouging law for milk products, which prohibits the retail price from being more than double the farm price. Hence, the construction of price-spread measures is the subject of controversy. The retail prices for most products are obtained by the Bureau of Labor Statistics via sampling procedures and are subject to sampling error. In addition, critics have questioned whether special sales prices, say, for beef, are captured by the sampling procedures. Further, when the retail price is a weighted average, as in the case of beef, the correctness of the weights can be questioned. The fixed weights may not be the same as the actual proportions of cuts obtained from the beef carcasses. Because of these criticisms, the U.S. Congress mandated that the USDA construct new price series for meats that include additional cuts of meat and price discount effects. (The USDA data come from supermarket scanners.) Thus, two alternative price-spread series are available for beef, pork, poultry, lamb, and veal, using the two different estimates of retail prices.

Farm prices are obtained by market reporters at selected locations. For beef, for example, the farm price is a simple average of the weekly prices of live choice steers in five regions. Since the observations at the farm level are obtained over the month but the retail prices are sampled at specific times during the month, the degree to which they are observed contemporaneously is uncertain. Indeed, since it takes time for beef to be processed and move through marketing channels, it is not clear that the current retail price should be compared with the current farm price.

Also, the quality definition for the farm commodity may not correspond exactly with the quality specification of the retail product. Still another issue is the correctness of the fixed-proportions factor used to adjust the farm price. For beef, it is assumed that it takes 2.4 pounds of live animal to produce 1 pound of retail meat.

Data could be improved with added resources. The question is, however, are the benefits of improved estimates of price spreads equal to the cost of improving the estimates? In principle, one can estimate the trade-offs, but in practice the question is answered as part of a political budgeting process.

If the spreads are accurately measured, they provide good estimates of the aggregate costs, including profits, of all processing and marketing activities between the farm gate and the retail grocery store sale. The spreads are estimates of a total margin for specific foods, and since the definition of each food item is held constant through time, the changes in the spread should be related to the costs and profitability of processing and marketing the particular foods. Most empirical studies of marketing margins in the United States have used the spread data.

The marketing chain from farmer to consumer provides many services. Farm commodities are stored, transported, processed, packaged, and advertised and promoted; use brokerage and wholesaling services; and finally arrive at retail outlets. The costs of these activities are dependent on the prices of numerous inputs and are affected by changes in technology. The prices of inputs, like labor, materials, and energy, are themselves determined in a variety of markets, like labor markets.

A commodity is frequently used for numerous final products. Thus, the derived farm-level demand for a commodity depends on the final demands of numerous products. Beef is sold as fresh meat in grocery stores, is processed into a variety of frozen and canned products, and is sold in restaurants, but the price spread for beef is estimated from fresh choice beef sold in retail outlets.

The value of the farm input is often a small proportion of the retail price. For example, the farm share of every dollar spent on food in the fixed market basket in the United States was 23 cents in 2010. It was only 7 cents for the market basket of cereals and bakery products. This implies that changes in the price of wheat should have little impact on the price of bread. If the price of wheat doubles (increases 100 percent) from 5 to 10 cents per pound and if only the ingredient cost is fully passed through, the implied increase in a $1.50

**Table 6-1.** Indexes of retail price and farm value of a market basket of foods, United States, selected years, 1950–2010, 1982–84 = 100

| Year | Retail price | Farm price | Farm value as a share of retail (%) |
|---|---|---|---|
| 1950 | 30 | 40 | 47 |
| 1955 | 31 | 36 | 41 |
| 1960 | 34 | 38 | 39 |
| 1965 | 35 | 40 | 38 |
| 1970 | 42 | 46 | 37 |
| 1975 | 64 | 76 | 40 |
| 1980 | 88 | 97 | 37 |
| 1985 | 104 | 96 | 32 |
| 1990 | 134 | 113 | 24 |
| 1995 | 149 | 103 | 24 |
| 2000 | 171 | 97 | 20 |
| 2005 | 198 | 122 | 22 |
| 2010 | 226 | 145 | 23 |

*Source:* Elitzak (2012).

loaf of bread is about 3 percent (0.05/1.50). Since non-farm input prices have tended to increase relative to commodity prices, the relative importance of the farm commodity's value in the retail product has tended to decrease (Antle 1999); however, there are exceptions to this generalization, at least over short periods of time. The "farmer's share of the consumer's dollar" statistic is based on market-basket data aggregated from the individual spread data (Table 6-1). This statistic is sometimes interpreted as the cents received by farmers per dollar spent on food by consumers (for the specific market basket). This value was 47 percent in 1950 and just 23 percent in 2010. Since the retail items in the basket are held constant, this change is for the most part not a reflection of added services. Rather, it is a reflection of changing prices of the various inputs used in producing and marketing the retail product. The farmer's share of the food dollar has stabilized in the 20–24 percent range since 2000.

The per unit margin (spread) statistics and especially the related concept of the farmer's share of the consumer's dollar are subject to misinterpretation. The tendency is to use the number to indicate the well-being of farmers or to indicate that marketing costs are too high. In fact, the farmer's share statistic has little to say about either issue. For instance, beef producers get a larger proportion of the retail value of beef than wheat farmers obtain from the value of bread, but this does not mean that beef producers are better off than wheat farmers. Also, since the spread is measuring a total margin, it alone cannot identify whether some firms in the marketing chain are obtaining "excess profits." At most, one can use statistical methods to try to explain changes in margins.

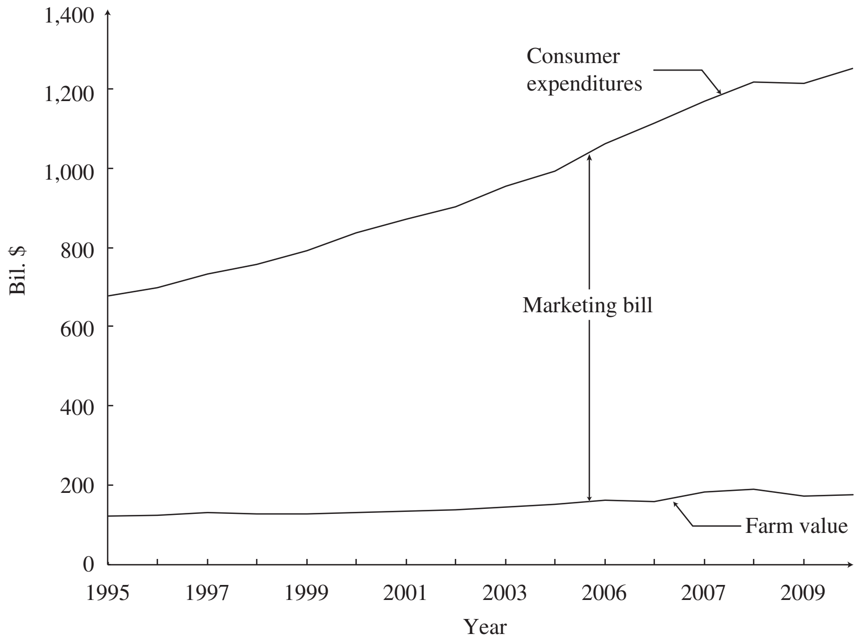



**Figure 6-2.** Consumer expenditures, marketing bill, and farm value for U.S. farm foods, 1995–2010

The USDA also computes an aggregate food-marketing bill based on the cost of all marketing services added to purchases from U.S. agriculture (Canning 2011). (Thus, this particular measure of total consumer expenditures relates to domestically produced foods, not to all food consumed in the United States.) The marketing bill reflects changes in the volume of food marketed, changes in the services provided, changes in the product mix marketed, and changes in costs of existing services. As depicted in Figure 6-2, the marketing bill is trending upward, and the farm value share is a decreasing percentage of consumer expenditures. In the early 1970s, the farm value was about 33 percent of total consumer expenditures on food; by the mid-1990s, the farm share was about 18 percent; and since 2000, this share has ranged between 16 and 14 percent. This is a consequence of a variety of factors: higher prices for marketing inputs, the addition of more services to products, and an increase in the proportion of sales of products with these services. Basically, the composition of the demands for foods has been changing. Consumers have choices among buying ingredients and preparing foods at home, buying prepared foods, and eating away from home, and with changing incomes and lifestyles, consumers in the United States and elsewhere are opting for the items with the added services. Thus, it is understandable that the farmers' share for a fixed market basket is larger than the farmers' share for a changing market basket.

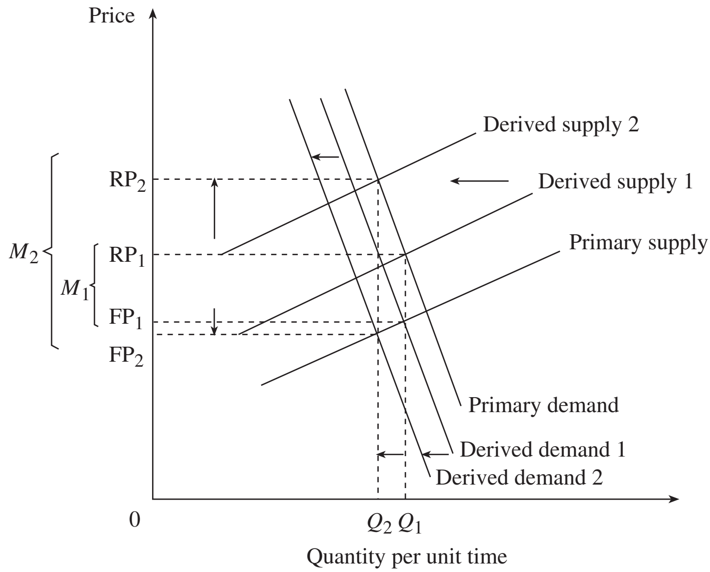



**Figure 6-3.** A change in margin and the incidence of change on retail prices (RP) and farm prices (FP)

## Incidence of Changes in Marketing Costs

If the price of a marketing input, such as wages, increases, what is the consequence for the prices of a retail product and its farm input? The general answer is that both farm and retail prices are affected and that the relative magnitudes depend on the own-price elasticities of supply and demand at the products' prevailing prices and quantities.

For simplicity, we consider first the very short-run situation depicted in the top panel of Figure 6-1. In this case, the quantity supplied ($Q$) is fixed, as is the retail demand function. An increase in a marketing cost will, *ceteris paribus,* cause the farm-level derived demand to shift backward by the amount of the per unit marketing costs. The entire incidence of the margin increase is reflected as a lower farm price; the retail price is not affected.

With the passage of time, however, the lower farm price will reduce the quantity supplied because the short-run supply curve has a positive slope. The smaller supply will increase both the farm and retail prices. Thus, ultimately, the retail price will adjust upward. The initial increase in the price of the marketing input triggers a chain of events that results in a lower farm price and a higher retail price. This is depicted as backward shifts in the derived supply and demand functions while assuming that the primary supply and demand functions are unchanged. An examination of Figure 6-3 helps visualize that differing slopes of the supply and demand functions will result in differing incidences on farm and retail prices.

The foregoing analysis has several qualifications. First, it applies to the changing costs of providing existing services for a given commodity. We are not discussing changes in costs associated with adding new services (attributes). If the retail product changes, this should be depicted by a new retail demand function specific to that product.

Second, the analysis shown in Figure 6-3 is an example of partial-equilibrium theory. An equilibrium situation for a single product is shown, and one variable is changed with other factors held constant. Although useful for illustrating the general nature of the changes in a margin for farm and retail prices, it may be misleading when applied to real-world situations where other things are *not* actually constant. For example, a lower transportation cost for apples grown in Washington State will result in higher prices for Washington growers and lower retail prices for eastern consumers. However, this analysis would be incomplete unless it took into account the substitutability between western-grown and eastern-grown apples. A lower retail price for western apples will result in a sympathetic price decrease for substitutes. Thus, the net effects on prices will not be as large when the secondary effects are considered.

The foregoing conceptual analyses also assume that price changes are transmitted instantly through the price system, but this is not usually the case. The topic of price adjustments is addressed in the next section.

## Price Transmission

New information, which influences prices, can originate from consumers (changes in primary demand), from the farm level (changes in commodity supply), or from the marketing sector (changes related to marketing inputs). In a perfectly competitive market, prices would adjust instantly and correctly to the new information, but of course, markets are not perfect.

The topic of price transmission has to do with the nature of price adjustments in the marketing system. Market analysts are interested in such questions as, how long do price adjustments take? Are adjustments symmetrical; i.e., are the length of lags and the magnitudes of adjustment the same for price increases and decreases? What are the reasons for this behavior, and to what degree is it a consequence of market imperfections?

For the most part, the analyses have been empirical and have concentrated on the relationship between farm and retail prices. These analyses lead to three generalizations. First, causality usually runs from changes in farm prices to changes in retail prices; it is uncommon to find these prices being simultaneously determined. Second, time lags are months in length, even for perishables like milk, meats, and fresh fruits and vegetables. Third, retail prices appear to respond asymmetrically, with adjustments to increases in farm prices occurring faster and more fully than adjustments to decreases in farm prices.

It has been difficult, however, to reach definitive conclusions about competing hypotheses to explain the observed price adjustments. As Azzam (1999) has pointed out, explanations of the "why" of asymmetry have been casual. This section provides a framework for understanding the issues, and as we will see, the various hypotheses are not mutually exclusive.

The marketing margin models depicted in Figure 6-1 are used as a point of departure. In these models, the margin depends on prices of marketing inputs but also on the quantity marketed (if, as is assumed in the lower panel, the marginal costs of marketing vary with the quantity marketed). Since the spread equals retail price minus farm price, the model can be specified so that retail price is a function of the farm price, marketing input prices, and quantity marketed. Suppose unusually good weather causes the farm supply function to shift to the right and farm prices to decline; then, in practice, retail prices do not adjust instantly. Thus, if the analyst is using weekly or monthly observations, the current retail price is made a function of lagged farm (or wholesale) prices. It is possible, as previously noted, that farm prices will decline more than retail prices, and the statistical model may need to allow for asymmetric responses, a topic outside the scope of this chapter.

Two questions arise: why do lags occur, and why are responses asymmetric? The answers are interrelated. One important explanation for lags is the costs of information transfer, e.g., the costs of repricing products in retail stores. Levy et al. (1997) estimate that the average menu costs in U.S. supermarket chains are 0.7 percent of revenues, 35 percent of net margins, and 52 cents per price change. The cost of repricing was especially clear when individual items on store shelves had to be re-marked by hand, but it apparently remains important even with the ability to make price changes in computerized price lists. A related explanation is possible inertia in information transfer. The pricing process used by the retailer may have built-in lags; retailers do not necessarily adjust their prices instantly to changes in their wholesale costs. Also, retail prices may remain rigid over some range of changes in wholesale prices. Retailers may be concerned about negative reactions from consumers to frequent repricing. Asymmetry also can be explained by technical arguments related to the costs of repricing (Azzam 1999).

Another explanation for lagged and asymmetric price responses is related to the nature of consumers' expectation formation and inventory adjustments to price changes. Inventories consist of stocks held both by consumers and merchants. In this context, price increases may cause consumers to increase their inventories in anticipation of still further price increases, which has the effect of reducing inventories held by merchants, thereby accelerating the price increase. If consumers do not slow their purchases when prices are falling as much as they accelerated them when prices were rising, then retail prices will fall more slowly than they have risen. For retailers and food manufacturers, the costs of adjusting to inventory changes may not be symmetric; the short-run costs of adjusting to a decrease in inventory are perhaps larger than the costs of adjusting to an increase in inventories. (These inventory arguments are made by Borenstein, Cameron, and Gilbert 1997 in the context of asymmetric price responses for gasoline.)

A third possibility is that market power allows marketing firms to delay price adjustments and to capture additional profits by passing cost increases on to consumers more rapidly than the cost decreases. This is a popular explanation of asymmetric price responses but one that is difficult to prove empirically.

A fourth possible explanation of lags relates to the questions, should the estimated margin be based on contemporaneous farm and retail prices, and if so, are the farm and retail prices collected contemporaneously? If monthly observations are being used, but the retail prices are collected earlier in the month than are farm prices, then the resulting lag is merely a function of the timing of the collection of data.

In addition, long lags may result from the biological and institutional adjustments inherent in the production and marketing system for foods. As discussed earlier, an increase in supply that reduces farm prices will, *ceteris paribus,* reduce the quantity supplied in the next production period. As discussed in Chapters 4 and 9 on commodity supply and on the time dimensions of commodity prices, the biological production process is likely to induce autocorrelations in prices (i.e., the current price depends on past prices).

To the extent that labor unions working in food manufacturing, wholesaling, or retailing have cost-of-living adjustment (COLA) clauses in their contracts, they can create lags in price adjustments (Lee 1988). If food prices increased, then wages would increase, which in turn would increase marketing costs, which would then increase retail food prices while reducing farm prices. This would occur only with time lags because of the institutional nature of the COLA process. Few studies have been conducted, however, of possible lags in adjustments of retail prices to changes in non-farm input costs, like wages.

None of the foregoing arguments is, in our view, completely satisfactory in explaining the length of lags for perishable produce that seem to exist in the data. By definition, perishables, like fresh meats, have a relatively short shelf life. Grocery stores receive these products frequently; i.e., they have continuous updates of wholesale prices. In this context, it is unclear why retail prices seem to require months to adjust to changes in farm prices.

## Market Structure and Margins

The structure of markets probably influences the behavior of marketing margins. This section considers trends in market structure and the implications of these trends for performance. Some empirical evidence is introduced. The conclusions of these studies are, however, conditioned by the assumptions of the underlying models used in the analyses. This emphasizes the importance of understanding the nature of the alternative economic models (discussed above) that can be used to study marketing margins.

## Framework

The food and fiber marketing system consists of a complex mixture of types and sizes of firms. It is not possible, nor is it the purpose of this book, to describe the food marketing system in detail, but several generalizations are relevant, two of which are considered here. First, one often finds a core of very large firms surrounded by a fringe of small firms. Second, mergers and acquisitions by large firms have tended to increase the concentration in food manufacturing and food retailing.

Mergers have influenced many parts of the marketing chain, even the transportation sector. The resulting firms are large and often have diverse outputs. A food manufacturer may, for example, produce a combination of producer goods (products sold to other manufacturers), unbranded or private-label consumer products, and brand-name consumer products. A large processor may buy farm commodities in spot markets from a variety of suppliers and/or make forward contracts with farmers; the latter are sometimes called "captive supplies." In addition, a processor may vertically integrate back to the farm level, thereby producing a portion of its ingredient needs.

Increased concentration can be measured from the viewpoint of both the purchase of inputs and the sale of output. Since 1995, the top four beef packers have slaughtered 79 percent or more of the steers and heifers processed in the United States (USDA 2012: 31). The four-firm concentration ratio for hog processing was 34 percent in 1980 and has stabilized at just over 60 percent since 2003 (USDA 2012: 31). Presumably the sales of meats have similar levels of concentration.

Unquestionably the food marketing system does not fulfill all the requirements of a purely competitive model. This raises the question, what are the consequences of the changing structure of the system? Presumably, mergers and acquisitions improve the economic well-being of the resulting firm (although, in some cases, the short-run consequences have been negative). There may be economies of size, and large firms probably have a wider range of alternatives in financing their operations and managing risks. With effective competition, these benefits would be shared with consumers and farmers; i.e., *ceteris paribus,* the margin between farm and retail prices would decrease.

Large size also has the potential to create monopoly or monopsony power that can be used to influence the prices paid for inputs procured from farmers and others, as well as output prices. As explained in Chapter 5, the price set by a monopoly for its output is likely to be higher than the competitive price. A related concern is that large firms will make non-competitive allocations of resources to advertising, packaging, compensation for executives, etc. Thus, increased concentration in food processing and retailing may result in larger margins.

General appeals to logical reasoning cannot, however, predict the effect of changes in market structure on marketing margins (and other performance measures) in the food marketing system. Although a majority of economists would probably associate increased concentration with larger marketing margins, the empirical evidence is mixed.

## Empirical Evidence

Studies of the effects of market structure have used a variety of models. One set of models analyzes retail prices and consumer welfare. A second set of studies focuses on farm prices and farmer welfare. A third set has examined both retail and farm prices, i.e., the effects of market structure on the performance of margins. Fourth, there is the limited evidence for the effects of market structure on price transmission.

*Retail Prices and Consumer Welfare.* Research on welfare losses by consumers requires estimates of retail demand functions (elasticities) and of marginal costs of food manufacturers and/or other marketers. In addition, the research must make assumptions about the nature of the oligopolistic behavior of the marketing firms. This information can be used to estimate welfare losses and income transfers due to monopolistic pricing (for a discussion of methods, see Hines 1999 or Kaiser and Suzuki 2006). Since the researcher has many choices in modeling and estimation, the estimates of welfare losses by consumers vary widely. Peterson and Connor (1995) summarize 47 studies of U.S. food manufacturing industries. They find (1) that welfare losses do exist, (2) that the estimated magnitude of losses varies widely among studies, and (3) that, nonetheless, the studies have a high degree of agreement in the ranking of industries. For example, the breakfast food and soft drink industries were typically in the top five in terms of estimated welfare losses, no matter which model was used in the analysis.

The breakfast food industry is interesting because it was subjected to well-publicized calls for lower prices in the mid-1990s, and apparently as a consequence, some major firms in the industry reduced their shelf prices. Cotterill and Franklin (1999) estimate that consumer benefits from the reduced prices over a 35-month period, starting in April 1995, were perhaps \$2.6 billion. Among other things, this estimate of consumer savings depends on what one assumes prices would have been if there had been no campaign to lower prices.

Another line of research has attempted to estimate the effect on a market basket of food prices of varying levels of concentration in retail food firms in different locations. This research requires data on food prices as well as information on concentration and other factors that may influence prices for individual firms in numerous different locations. A typical model tries to explain the variability of an index of prices of a market basket of food collected from stores of different firms in different cities. The explanatory variables include measures of market share held by the individual firm in the particular city and the degree of concentration in the market (such as a four-firm concentration ratio). Other variables in the model attempt to capture differences in costs and efficiencies of the firms, including the possibility that different grocery stores provide different levels of services. From the viewpoint of marketing margins, this is an attempt to disaggregate the costs of retailing for individual firms at various locations.

Some research has found a positive relationship between retail food prices and market concentration; other research has not (Kaufman and Handy 1989). The results are sensitive to the data and model used in the analysis. Clearly it is difficult to hold other things constant in the statistical model and hence to obtain an estimate of the net effects of market-structure variables.

*Farm Prices and Farmer Welfare.* Oligopsony is a potentially important issue at the farm gate. Farm commodities are often bulky and/or perishable, and farmers are specialized in the production of particular commodities, implying that they have little short-run flexibility in the use of their resources. In this context, high buyer concentration in particular locations may give buyers considerable market power relative to farmers, thereby influencing the prices paid to farmers. In other words, oligopsony power may result in larger marketing margins (Rogers and Sexton 1994); however, as noted in Chapter 5, empirical studies have found that efficiency gains from consolidation in meat packing dominated any possible losses from market power (e.g., Morrison-Paul 2001). Also, farmers sometimes organize into large marketing cooperatives or bargaining associations to counter buyers' market power.

A limited amount of empirical research has addressed whether monopsony power has lowered prices received by farmers and whether farmer cooperatives have been effective in maintaining or raising farm prices. Rogers and Petraglin (1994) studied a cross section of 134 industries in the food sector (e.g., wheat flour and fresh fruit juices) using 1982 census data. Their model included a variable to measure the percentage of the value of shipments in the market accounted for by the 100 largest cooperatives in each industry and a measure of four-firm concentration in the industry. The variable to be explained was the output price relative to average materials and labor costs (i.e., is not a direct analysis of prices paid farmers). Their research found that cooperatives tended to reduce output prices (relative to costs), while larger concentration increased prices.

Most analyses of farm prices are related to the livestock sector. What is the effect on prices paid to farmers of the increased concentration observed in meat packing? To answer this question, the research requires observations on transactions prices for individual lots of the commodity as well as on the attributes inherent in the particular lots. The effect of market structure is just one of many factors influencing price of the commodity.

Ward (1992), for example, obtained live-weight sales prices for fed cattle from 173 feedlots in the southern plains in June 1989. This data set was for a particular location and a relatively short time period. Nonetheless, Ward had to model differences in the attributes of each lot, including percentage of choice-grade cattle, the sex of the animals, and the number of head in each lot (for a discussion of the effects of quality characteristics on prices, see Chapter 7). In this context, he estimated the effect of the number of buyers bidding and of the size of the packer; the "big three" packers in the region were differentiated from smaller processers. His analysis suggested that, *ceteris paribus,* the larger the number of buyers, the higher the price paid. Also, the evidence suggested that the three largest firms, as a group, paid lower prices than their competitors. Nonetheless, Ward notes that the conclusion that monopsony power lowers fed cattle prices is tenuous.

*Performance of Margins.* Numerous studies have examined the effect of oligopoly and oligopsony on marketing margins. This research has involved varying levels of aggregation, from individual markets, like beef, lamb, and pork, to more aggregate markets, like all meat. Some of this research has tried to isolate the effects of particular types of concentration, such as in meat packing, while other research has considered more aggregated approaches to market-structure effects. Capps, Byrne, and Williams (1995) found evidence, although statistically weak, that increased concentration in meat packing increased the marketing margin for lamb. They used bi-monthly national data.

Koontz, Garcia, and Hudson (1993) analyzed the difference between choice steer and box beef daily prices in selected locations. They conclude "that market power appears to have been exercised in cattle purchases . . . " (1993: 547). Reed and Clark (2000) used annual data for seven commodity groups in the United States, and they found that changes in marketing margins were consistent with competitive behavior. They attribute the differences in results in the literature to the differences in models. "Studies based on fixed-proportions consistently reject the competitive model, whereas [models that permit variable-proportions technology] just as consistently fail to reject the competitive model for U.S. food industries" (Reed and Clark 2000: 2).

As emphasized earlier, the literature on margin behavior contains a variety of conceptual models, and it is important to understand the implications of alternative models. Several related caveats are applicable to all statistical analyses, including those in Reed and Clark. First, a hypothesis test should be viewed as a joint test of the particular null and the entire model specification. Model specification includes choices of concepts, like fixed- or variable-proportions technology, and also practical issues of functional form, excluded variables, lag structure, and sample period (Chapter 15). Second, the conclusion "cannot reject the null hypothesis of a competitive market" is not equivalent to concluding that the market is competitive.

*Asymmetric Price Transmission.* Finally, as discussed above, market concentration is one possible explanation for asymmetric price transmission, and numerous studies have found asymmetry. It is difficult to test whether the asymmetry is related to changes in market structure; this would require a model that could test whether varying degrees of asymmetric responses were associated with varying levels of market concentration. Bernard and Willett (1996) provided indirect evidence that concentration in the broiler industry may have influenced price transmission. Wholesale prices for broilers appear to have become the central, causal price in this market, and Bernard and Willett found that downward changes in wholesale prices are passed on more fully to farmers than are increases in wholesale prices. However, no asymmetries were found between wholesale and retail prices.

## Marketing Margins and Elasticities of Demand at Various Market Levels

As noted above, marketing margins between the farm and retail levels can be viewed in terms of a difference between primary and derived (farm-level) demand. Thus, in this section, relationships between price elasticities of demand for primary and derived demand relations for the same commodity are discussed. We first cover the relationship between elasticities at various market levels, especially the retail level versus the farm level. In addition, we discuss the relationship between the elasticities for joint products and the commodity from which they are derived.

### Market Levels

Much of the empirical research on the demand for foods has used retail-level data, but sometimes analysts want estimates of farm-level elasticities, e.g., for a supply-control policy to be effective in raising farm income, the derived demand elasticity needs to be inelastic. The elasticities at the two levels will usually be different, and if the researcher is willing to make some assumptions about the nature of the marketing margin between the two levels, the farm-level elasticity can be approximated from the retail-level estimate. In this context, the typical question is, given the retail price elasticity of demand, what is the corresponding farm-level elasticity? This implies that the analyst is interested in comparing elasticities for a given quantity (see Figure 6-1) for a commodity that is relatively unchanged as it passes through the marketing system, e.g., eggs, apples, or beef.

The exact relationship between elasticities depends, however, on how the primary and derived demand curves are related. One approach in the literature is to assume that the derived demand function can be obtained by subtracting a schedule of marketing costs from the primary demand function (explained more fully in the appendix to this chapter). A simple, but somewhat unrealistic, assumption is that the margin is a constant, absolute amount regardless of the amount marketed (as in Figure 6-1). This is equivalent to assuming that the marginal cost of all processing and marketing is a constant over the range of quantity marketed. Thus, if the two demand functions are straight lines, they will be parallel. In this case, the elasticity at the farm level can be estimated from the elasticity at the retail level as follows:

$$E_{\mathrm{d}} = E_{\mathrm{r}}(P_{\mathrm{d}}/P_{\mathrm{r}}),$$

where the subscript d indicates the derived (farm) level and subscript r indicates the primary (retail) level, $P$ is the price, and $E$ is the own-price elasticity.[1]

From the foregoing, the margin, $M$, equals a constant, $c$; $c = P_{\mathrm{r}} - P_{\mathrm{d}}$. The ratio $P_{\mathrm{d}}/P_{\mathrm{r}}$ is always less than 1, and consequently the retail-level elasticity is being multiplied by a number less than 1. This implies, of course, that the farm-level demand is more price inelastic than retail demand. A large margin results in a large difference in the elasticities at the two levels. For instance, if the ratio is 0.5 and the retail-level elasticity is −1.5, the farm-level elasticity is −0.75 (= −1.5 × 0.5). Thus, the primary demand for bananas in the United States can be price elastic while the derived demand in Central America is inelastic. The two market levels, in this case, are separated by a large transportation cost, implying a large margin.

If the assumption of $M = c$ is relaxed, then the comparison of elasticities will depend on the assumption made about the nature of $M$ as $Q$ varies. The alternative assumptions imply that the slopes of the primary demand function and of the derived demand function differ. Thus, it may still be possible to make a qualitative judgment that demand at the farm level is more inelastic than at the retail level, but an estimate of the farm-level elasticity now requires additional information (beyond the ratio of prices). For example, if the marginal cost of marketing is increasing as $Q$ gets larger, this implies that $|dQ/dP_{\mathrm{r}}| > |dQ/dP_{\mathrm{d}}|$ (see the appendix to this chapter). Given that the price at the retail level is above the price at the farm level, plausible arithmetic examples show that demand is more inelastic (less elastic) at the farm level.

An assumption used in the literature on comparing elasticities is that the margin is a combination of a constant absolute amount and of a percentage

1. Assuming a constant margin between parallel demand equations, the only thing that differs between the two elasticities is the price. Both equations have identical slopes, $dQ/dP_{\mathrm{r}} = dQ/dP_{\mathrm{d}}$, and the elasticities are computed for a given level of $Q$. Thus, in

$$E_{\mathrm{d}} = E_{\mathrm{r}}(P_{\mathrm{d}}/P_{\mathrm{r}}) = (dQ/dP_{\mathrm{r}})(P_{\mathrm{r}}/Q)(P_{\mathrm{d}}/P_{\mathrm{r}}),$$

the $P_{\mathrm{r}}$'s cancel, and given the equality of the slopes, the farm-level elasticity is obtained. As noted previously in this chapter, this computation also assumes that the retail quantity is produced as a fixed proportion of the farm quantity and, thus, that the $Q$ used in the computation equals $kA$, where $k$ is the constant proportion and $A$ is the farm-level quantity.

markup. The markup is, however, specified as a percentage of the retail price because this specification permits a convenient derivation of a farm-level elasticity from a retail-level elasticity, which is the more commonly known coefficient. In practice, of course, markups are made by store managers as some percentage of the farm-level or the wholesale-level price, and this percentage is unknown. Thus, although it is possible to show the relationship between the retail and farm elasticities under this assumption, it has relatively little practical importance because the researcher does not have the necessary information to do the computations. Again, it is possible to reach the qualitative conclusion that demand is likely to be more price inelastic (or less elastic) at the farm level,[2] but it should be noted that the assumption of a fixed percentage markup has the consequence of implying that the marginal cost of marketing declines as the quantity marketed increases, which may not always be true.

### Joint Products

If joint products are obtained in fixed proportions from the basic commodity and if the elasticities are all computed at the same market level, then the price elasticity of the underlying basic commodity ($E_X$) is a weighted harmonic average of the price elasticities of the joint products, say $E_1$ and $E_2$ (Houck 1964). Specifically, the fixed-proportions assumption is defined as:

$$X_1 = w_1 X \quad \text{or} \quad w_1 = X_1/X, \text{ and}$$

$$X_2 = w_2 X \quad \text{or} \quad w_2 = X_2/X$$

where $X$ is the basic commodity, $X_1$ and $X_2$ are the joint products obtained from $X$, and $w_1$ and $w_2$ are the fixed yields per unit of $X$. Also, if $P_1$ and $P_2$ are the prices per unit for the respective joint products, then the relationship among the three elasticities is

$$\text{(1)} \qquad E_X = [P_1 w_1 + P_2 w_2]/[(1/E_1)(P_1 w_1) + (1/E_2)(P_2 w_2)].$$

This example can be generalized to the case of $n$ joint products (Houck 1964), but the case of two major products, such as wool and meat, produced from one commodity, sheep, is common.

The weights in the expression are proportions of $X$'s average value (per unit) attributable to the sales of the joint products. Thus, taking $E_1$ and $E_2$ as constants, $E_X$ will vary as the value weights vary. The extreme cases are either $P_1 w_1$ or $P_2 w_2$ equals zero; that is, $X_1$ or $X_2$ represents the entire value of $X$ and the

2. The third and earlier editions of this book provided a detailed derivation of the derived-level elasticity in terms of the primary-level elasticity under the assumption that the margin can be viewed as a combination of an absolute amount and a percentage of the retail price. We have come to view, however, that this is an algebraic exercise with little practical importance.

remaining component is discarded.[3] If $X_1$ has zero value, then $E_X = E_2$. Or, in a more realistic case, if $X_2$ represents the major proportion of the value of $X$, then the two elasticities are approximately the same. An example is the division of a beef animal into the joint products meat and hide. The hide represents a relatively small proportion of the total value of the animal. Hence, the farm-level elasticities for the animal and for beef are similar.

Soybeans, which are processed into the joint products soybean meal and oil, provide another example. In this case, $X$ is a 60-pound bushel (27.2 kilograms) of soybeans, $X_1$ is 47.6 pounds of meal, and $X_2$ is 11.1 pounds of oil, with 1.2 pounds assumed to have zero value (USDA 1992). The weights are 0.793 (47.6/60) and 0.185 (11.1/60). For the purposes of the example, we assume that the price of soybean meal is 25 cents per pound and that the price of soybean oil is 50 cents per pound; we further assume that the respective price elasticities are –0.9 and –2.5. Then, using the Equation (1),

$$\begin{aligned} E_X &= [25(0.793) + 50(0.185)]/[(1/-0.9)(25)(0.793) + (1/-2.5)(50)(0.185)] \\ &= -1.13. \end{aligned}$$

If the hypothetical elasticities for the joint products are correct, then the derived demand for soybeans is price elastic at the assumed prices for the joint products. Clearly, the prices of the products can change, which is to say that the weights are not constants even if the products are obtained in a fixed proportion from the underlying commodity.

## Concluding Remarks

A marketing margin is the difference between prices at two market levels. But, the margin for a particular commodity, such as between bread and wheat prices, is difficult to measure since the form of the commodity changes as it is combined with other inputs in the marketing system.

Farm-wholesale-retail price spreads are estimated by the USDA for selected products. These spreads attempt to measure the cost of providing a set of marketing services that transform the farm commodity into a retail product. They do not measure the effect of adding new services to an existing product or the effect of introducing new products.

Theoretical models of margin behavior suggest that they depend on changes in factor prices and efficiency in the marketing sector, changes in the supply of the farm commodity, and changes in retail demand. Empirical evidence clearly supports the relationship between margins and changes in factor prices. The

3. In reality, throwing away waste is costly, and this case is ignored. Each of the joint products typically has some value, although as noted in this chapter, the value of one of the products may be a tiny proportion of the total value.

extent to which changes in retail demand and in farm output effect margins is less clear.

The empirical evidence is also mixed in terms of the effects of imperfect competition on marketing margins in the food industry. Changes in market structure, through mergers and acquisitions, can result in cost savings, especially for the firms involved. These potential savings may be offset wholly or in part by costs associated with non-price competition such as higher advertising and promotion costs. There is the additional question of whether large firms engage in oligopolistic or oligoposonistic pricing practices that reduce consumer and producer welfare. Given the numerous theoretical and measurement issues in estimating the possible effects of imperfect competition, it is not surprising that diverse results have been obtained.

Likewise, the academic literature contains many studies of price transmission in food markets. These studies typically find substantial lags as well as asymmetries in price transmission. These studies have been less successful in understanding the fundamental sources of this price behavior.

## Appendix: Marketing Margin Model

This appendix formalizes the discussion of the marketing margin model presented in Figure 6-1 of the chapter. A marketing firm's short-run profit equation is $Pq - P_A a - P_B b$, where $P$ is the price of (retail) output, $q$ is the quantity produced, $P_A$ is the price of farm commodity input, $a$ is the quantity of input $a$, $P_B$ is the price of marketing input, and $b$ is the quantity of input $b$. In this model, $q$ is assumed to be produced as a fixed proportion of $a$, so that $q = ka$; for simplicity, let $k = 1$. Hence, the profit function can be rewritten as $(P - P_A)q - P_B b$, where $P - P_A$ is the margin.

The margin between the farm price and retail price depends on the assumption made about the nature of the short-run variable costs related to the use of input $b$, the generic marketing input. If these costs are assumed to be a linear function of $q$, then $P_B b = c_1 q$, where $c_1$ is a parameter of the cost equation. Substituting into the profit ($\pi$) equation,

$$\pi = (P - P_A)q - c_1 q.$$

The first-order condition for profit maximization is obtained from the first derivative as

$$d\pi/dq = (P - P_A) - c_1 = 0, \text{ or}$$

$$P - P_A = c_1.$$

In this specification, the marketing margin equals a constant, $c_1$, which is independent of the quantity ($q = a$) produced and sold.

If, instead, the variable cost is specified as a quadratic function of $q$, then the short-run profit equation is

$$\pi = (P - P_A)q - c_1 q - c_2 q^2,$$

and the first-order condition for profit maximization is

$$d\pi/dq = (P - P_A) - c_1 - c_2 q = 0, \text{ or}$$

$$P - P_A = c_1 + 2c_2 q.$$

Marginal costs and the margin increase as $q$ (= $a$) increases.

The model is in terms of an individual firm. Assuming that marketing sector consists of $n$ identical firms all facing the same prices, total output is $Q = nq$, and the inverse supply equation for $Q$ can be written as

$$P - P_A = c_1 + 2c_2(Q/n) = \alpha_1 + \alpha_2 Q,$$

where $\alpha_2 = 2c_2/n$. To complete the model, an aggregate retail-level demand equation is specified as

$$P = \beta_0 + \beta_1 Q, \quad \beta_1 < 0.$$

The farm-level (derived) demand equation is obtained by subtracting the supply equation from the demand equation. This follows from our assumptions, including, especially, the assumption that $Q$ is produced in fixed proportion to $A$. The derivation, therefore, differs from the one that assumes variables proportions (Chapter 4).

$$P_A = (\beta_0 - \alpha_1) + (\beta_1 - \alpha_2)Q.$$

The margin in this model varies with $Q$, which is shown in the lower panel of Figure 6-1. It is possible in an efficient competitive market for the margin to widen as the supply of the farm commodity increases. A key assumption is the quadratic cost function; in practice, this is unknown.

The model also assumes that the price of the input $B$ is constant. If $P_B$ changes, then the level of the cost function will change, and hence the margin will change. In sum, for this model, the size of the margin depends on exogenous changes in the supply of $A$ (= $na$) and on the price of $B$, $P_B$.

The fixed-proportions assumption is used by the USDA in constructing some retail consumption series, such as the consumption of beef, but of course $k$ is not 1. When $k$ is not 1, the foregoing model is merely modified by changing the profit function as follows:

$$Pq - P_A(1/k)a - P_B b = [P - (1/k)P_A]q - P_B b.$$

The marketing margins reported by the USDA are computed as

$$M = P - (1/k)P_A - r,$$

where $P$ is the retail price, $P_A$ is the farm price, and $r$ is a by-product allowance. For beef in the United States, $k$ = 0.41667, or $1/k$ = 2.4. So, the margin for choice beef depends on the retail price of choice beef minus 2.4 times a farm price of choice steers minus an allowance for the value of the by-products (the hides).

## References


Antle, J. M. 1999. "The New Economics of Agriculture," *Am. J. Ag. Econ.* 81:993–1010.

Azzam, A. 1996. "Testing the Monopsony-Inefficiency Incentive for Backward Integration," *Am. J. Ag. Econ.* 78:585–590.

———. 1999. "Asymmetry and Rigidity in Farm-Retail Price Transmission," *Am. J. Ag. Econ.* 81:525–533.

Bernard, J. C., and L. S. Willett. 1996. "Asymmetric Price Relationships in the U.S. Broiler Industry," *J. Ag. & Applied Econ.* 28:279–289.

Borenstein, S., A. C. Cameron, and R. Gilbert. 1997. "Do Gasoline Prices Respond Asymmetrically to Crude Oil Prices?" *Quart. J. Econ.* 112:305–339.

Brorsen, B. W., J-P. Chavas, W. R. Grant, and L. D. Schnake. 1985. "Marketing Margins and Price Uncertainty: The Case of the U.S. Wheat Market," *Am. J. Ag. Econ.* 67:521–528.

Canning, P. 2011. *A Revised and Expanded Food Dollar Series: A Better Understanding of Our Food Costs*. USDA Econ. Res. Rep. ERR-114. February.

Capps, O., Jr., P. J. Byrne, and G. W. Williams. 1995. "Analysis of Marketing Margins in the U.S. Lamb Industry," *Ag. & Res. Econ. Rev.* 24:232–240.

Cotterill, R. W., and A. W. Franklin. 1999. "An Estimation of Consumer Benefits from the Public Campaign to Lower Cereal Prices," *Agribusiness* 15:273–287.

Elitzak, H. 2012. *Market Basket Indexes of Retail Price, Farm Value, Farm Value Share, 1950–2010 (Excel file)*. Econ. Res. Ser., USDA, December.

Gardner, B. L. 1975. "The Farm-Retail Price Spread in a Competitive Food Industry," *Am. J. Ag. Econ.* 57:399–409.

Heien, D. M. 1980. "Markup Pricing in a Dynamic Model of the Food Industry," *Am. J. Ag. Econ.* 62:10–18.

Helmberger, P. G., and J-P. Chavas. 1996. *The Economics of Agricultural Prices*. Upper Saddle River, NJ: Prentice Hall.

Hines, J. R., Jr. 1999. "Three Sides of Harberger Triangles," *J. Econ. Persp.* 13(2):167–188.

Holloway, G. J. 1991. "The Farm-Retail Price Spread in an Imperfectly Competitive Food Industry," *Am. J. Ag. Econ.* 73:979–989.

Houck, J. P. 1964. "Price Elasticities and Joint Products," *J. Farm. Econ.* 46:652–656.

Kaiser, H. M., and N. Suzuki, eds. 2006. *New Empirical Industrial Organization and the Food System*. New York: Peter Lang Publishing.

Kaufman, P. R., and C. R. Handy. 1989. *Supermarket Prices and Price Differences: City, Firm, and Store-Level Determinants*. USDA Tech. Bul. 1776.

Koontz, S. R., P. Garcia, and M. A. Hudson. 1993. “Meatpacker Conduct in Fed Cattle Pricing: An Investigation of Oligopsony Power,” *Am. J. Ag. Econ.* 75:537–548.
Lee, D. R. 1988. “Labor Market Dynamics in the U.S. Food Industry,” *Am. J. Ag. Econ.* 70:90–102.
Levy, D., M. Bergen, S. Dutta, and R. Venable. 1997. “The Magnitude of Menu Costs: Direct Evidence from Large U.S. Supermarket Chains,” *Quart. J. Econ.* 112:791–825.
Lyon, Charles C., and Gary D. Thompson. 1993. “Temporal and Spatial Aggregation: Alternative Marketing Margin Models,” *Am. J. Ag. Econ.* 75:523–536.
Morrison-Paul, C. J. 2001. “Cost Economies and Market Power: The Case of the U.S. Meat Packing Industry.” *Rev. of Econ. & Stat.* 83:531–540.
Peterson, E. B., and J. M. Connor. 1995. “A Comparison of Oligopoly Welfare Loss Estimates for U.S. Food Manufacturing,” *Am. J. Ag. Econ.* 77:300–308.
Reed, A. J., and J. S. Clark. 2000. *Structural Change and Competition in Seven U.S. Food Markets*. USDA Tech. Bul. 1881.
Rogers, R. T., and L. M. Petraglin. 1994. “Agricultural Cooperatives and Market Performance in Food Manufacturing,” *J. Ag. Coop.* 9:1–12.
Rogers, R. T., and R. J. Sexton. 1994. “Assessing the Importance of Oligopsony Power in Agricultural Markets,” *Am. J. Ag. Econ.* 76:1143–1150.
Thompson, G. D., and C. C. Lyon. 1989. “Marketing Order Impacts on Farm-Retail Price Spreads: The Suspension of Prorates on California-Arizona Navel Oranges,” *Am. J. Ag. Econ.* 71:647–660.
USDA. 1992. *Weights, Measures, and Conversion Factors for Agricultural Commodities and Their Products*. Ag. Hdbk. 697.
———. 2012. *2011 P&SP Annual Report*. Washington, DC: Grain Inspection, Packers and Stockyards Administration.
Ward, C. E. 1992. “Inter-Firm Differences in Fed Cattle Prices in the Southern Plains,” *Am. J. Ag. Econ.* 74:480–485.
Wohlgenant, M. K., and R.C. Haidacher. 1989. *Retail to Farm Price Linkage for a Complete Demand System of Food Commodities*. USDA Tech. Bul. 1775.

CHAPTER 7

# Price Differences Associated with Quality

Specific lots of an agricultural commodity can differ in terms of characteristics such as size, color, moisture level, protein and fat content, fineness of fibers, and proportion of defects or impurities. Wheat is sometimes viewed as a generic commodity, but there are in fact numerous varieties and grades of wheat, with varying attributes. For example, hard red spring wheat is milled into flours for making breads, while soft wheats are milled into flours for cookies and crackers.

Changes in the prices of different grades, attributes, and classes of a commodity tend to be correlated, but price differences exist among the various grades and can fluctuate as the levels of the various attributes vary. The economic rationale, meaning, and justification for price differences related to quality variation are explored in this chapter. (The next chapter will discuss price differences related to location (space), which also could be considered an attribute of a commodity.) To simplify the exposition, the term *grade* is used interchangeably with quality differences based on attributes (characteristics); however, in practice, an individual grade consists of a bundle of attributes. Therefore, the use of the term *grade* abstracts from the issue of valuing individual characteristics. It should also be noted that quality differences can involve undesirable attributes (e.g., pesticide contamination) as well as desirable attributes (e.g., high protein content).

The analysis of relationships among the prices of different grades is, for the most part, a special case of price relationships among substitutes. The prices of close substitutes are highly related in competitive markets, and because different grades of a commodity are often close substitutes, their prices often have large correlations. Large differences in prices can occur, however, precisely because an individual lot of a commodity has one or more attributes that differentiate it from other lots. A related issue is the availability and distribution of information among buyers and sellers about the differences in attributes; moreover, there is the question of the costs of obtaining this information and the incidence of this cost on buyers and sellers. Thus, in this chapter, we first discuss the demand for and supply of alternative grades and their relationship to price differentials. Then, issues related to defining grades and hedonic price models[1] and to possible market imperfections are discussed.

## Demand by Grades

We start with the assumptions that the attributes (hence, grades) of the commodity have economic significance, that they are measurable, and that buyers and sellers have equal access to this information. Given that grades have economic significance, it follows that a separate demand schedule exists for each grade. Each such schedule has the usual determinants of demand.

In general, there is great substitutability among grades of the same commodity, even though each has some differing levels of individual characteristics. Thus, the main demand shifter for a particular grade is the change in the price of its closest substitutes, which are usually the other grades and varieties of the product.[2] The various grades typically have large positive cross-price elasticities of demand with each other; hence, the demand for each grade is usually more price elastic than is the demand for the aggregate product. For example, Arzac and Wilkinson (1979) obtained price elasticities of −1.86 and −2.97 for grain-fed and non-grain-fed beef, respectively; the aggregate demand for beef is, in contrast, price inelastic.

Shifts in demand associated with income changes probably differ by grades. The income elasticity is thought to be highest for the "best" or preferred grade and smaller for lower grades. The reasoning is by analogy with the superior-inferior product concept. In principle, one might define grades on the basis of their relative income elasticities.

New uses for a commodity can change the relative demands for the attributes of that commodity. For example, processing uses have become the most important source of demand for corn (e.g., for processing corn into fuel ethanol, sweeteners, and starch). Thus, corn varieties designed to have a high starch content are of interest to corn processors. Similarly, preferences may shift, e.g., toward meats with a lower fat content. Also, demand shocks can occur if a product is found to have an undesirable attribute. An example would be a health concern related to the discovery of a pathogen in a particular food.

Wheat provides an example of the price relationships that exist among grades. Dark hard spring wheat is preferred for milling the flour used in the commercial

1. A *hedonic price model* is a model that explains the price of a commodity in terms of its characteristics. These characteristics are not completely separable from one another, but conceptually one could think of demand and supply functions for the individual attributes, with implicit prices for each of the attributes. Rosen (1974) and Waugh (1929) are original basic references.

2. This explains the high positive correlation in price changes among grades, but the unique characteristics of each grade explain why this correlation is not perfect.

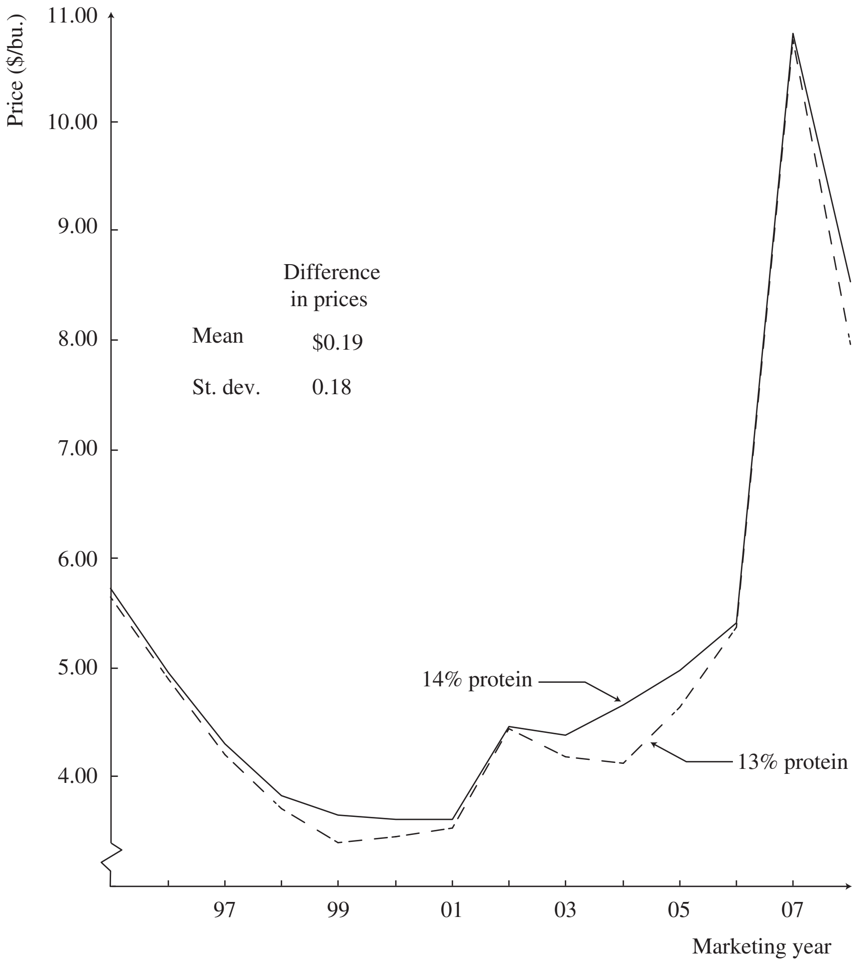



**Figure 7-1.** Prices of two protein levels of number 1 dark spring wheat, Minneapolis, 1995–96 to 2008–09 crop years

bread in the United States, and the relatively high protein content of these types of wheat is one of the desired attributes. For this reason, high-protein spring wheat customarily sells at a premium over other wheats, but the size of the premium varies from year to year and sometimes within a season.[3] To illustrate this point, the annual average prices of number 1 dark spring wheat for two levels of protein are plotted in Figure 7-1. These two grades of wheat are, of

3. Variability in premiums and discounts through time is of more than academic interest. For instance, a grain merchant storing grade-one grain and hedging this inventory in a futures contract specifying the delivery of grade-two grain is assuming that the prices of the two grades are closely correlated. Otherwise, the hedge will not work. (Hedging is discussed in Chapter 13.)

course, substitutes, and their prices are correlated. The wheat with 14 percent protein has a higher price than the wheat with 13 percent protein, and it is clear that the differential varies from year to year. For the crop years 1995–96 to 2008–09, the price of the 14 percent wheat averaged 19 cents per bushel higher than the price of the 13 percent wheat. Nevertheless, the difference varied substantially over the sample period, with a standard deviation of 18 cents per bushel.

The type of price behavior illustrated in Figure 7-1 is consistent with the demand function for the higher-grade wheat having a sharply changing slope as quantity varies. Bread flour requires a minimum protein content, and flour millers demand wheat to meet this requirement. When the supply of high-protein wheat is small, the lower-protein wheats are poor substitutes. The demand for high-protein wheat is, therefore, relatively price inelastic. When the quantity of high-protein wheat is large, the demand for protein is readily met from this supply, and the wheats with varying protein contents are close substitutes in meeting other residual demands for wheat. Thus, over the range of these larger quantities, demand is relatively price elastic.

If we generalize the wheat example, the demands for individual grades are likely to be price elastic within the range of quantities where the grades are close substitutes. This is associated with large positive cross-price elasticities, and it is in this context that the prices of the various grades are highly correlated. It is only when the supply of an important attribute is small—no alternative sources exist—that the demand for the individual grade is price inelastic. In this case, the price of this grade can spike above other grades.

## Supply by Grades

While demands for alternative grades can and do change, most year-to-year variations in price premiums are the result of changes in relative supplies. With the passage of time, producers can alter the varieties planted and change cultural practices to produce the attributes demanded by the market. For example, with the passage of time, pork producers have responded to consumer demands for less fat by producing leaner hogs. Thus, in principle, one can specify supply functions for alternative grades as a function of the relative (expected) prices of the grades and various determinants of supply.

Such responses, of course, require economic incentives, and the supply response may take considerable time. Expected profitability is influenced not only by price premiums (or discounts), but by the relative per unit costs of production. Clearly, other things being equal, a price premium for a particular grade or variety should tend to shift resources toward the production of that variety. But the variety could be more costly to produce and market. Changes in relative yields may in some cases offset the effects of price premiums. Or the choice of variety may influence the timing of harvest and, hence, influence the distribution and cost of labor requirements. It can be difficult to evaluate the relative benefits and costs of changing the variety produced.

The supply of attributes of farm commodities at any point in time is influenced importantly by random factors such as weather conditions and insect and disease damage. For example, the prices of individual lots of cotton depend on such attributes as trash content, color, fiber length, fiber strength, and fiber fineness (measured in microns and called micronaire). According to Bowman and Ethridge (1992), the supplies of these attributes depend on the variety of cotton, which is associated with the location of production, the level and timing of rainfall, and the level and timing of temperatures. In other words, given the variety of cotton planted, the supply of attributes is influenced in important ways by weather conditions. Of course, as noted previously, relative prices could influence the varieties planted, and with the passage of time, new varieties with improved qualities may be developed.

Analogous to the discussion of quantity supplied in Chapter 4, the bundle of attributes inherent in the available supply is fixed (predetermined) by past events, by farmers' decisions made months in the past as well as the subsequent growing conditions. And it is the fluctuations in these factors that cause variability in the supply of alternative grades of the commodity. The relative supplies of attributes from annually produced commodities may also vary within the year if quality declines during the storage period. Seasonality in the supplies of various qualities, sizes, and varieties of a product can result in seasonal patterns in quality premiums and discounts.

## Price Differences

Based on the foregoing discussion, the prices at any time are expected to vary among the different grades, reflecting the different sets of attributes. The supplies of attributes are predetermined by prior events, and given these supplies, prices are determined by the varying demands for the grades at that point in time. If data are available for transactions prices for individual lots of the commodity containing varying attributes, which are themselves known, the price for each lot is a function of that lot's attributes. This function is sometimes called a *hedonic price model* because buyers are thought to be deriving utility from consuming the attributes of the basic commodity.

One can visualize a base or reference price corresponding to a base grade. Then, the remaining prices, for the other grades, vary around the reference price because of the differences in attributes among the different grades of the commodity. When $P_b$ is the price of the base grade product and $P_j$ is the price of grade $j$ product, then $P_j - P_b = f(Z_j - Z_b)$, where the $Z$'s represent the characteristics of the respective grades. In words, the price differences at a point in time depend on the differences in attributes of the varying grades. It is perhaps worth emphasizing that this discussion assumes a competitive market structure but permits individual lots of the product to have different attributes. In this case (with a given output), producers have no control over the prices they receive; the market determines the prices of the different grades. But, in the short and longer run, farmers can influence the prices received to the degree that they can influence the quality of their output.

With the passage of time, demands and supplies change, and consequently, the price differences change. The size of the changes in quality discounts and premiums depends on the size of the shifts in functions and on the algebraic form and consequent slopes of these functions. One can visualize the overall price level rising and falling with changes in aggregate supply and demand and further visualize the relative prices of the different grades changing around the average level as the supplies and demands for the attributes vary.

Generalizations about price relationships among grades are difficult to make because of the many possible combinations of changes in relative supplies and demands by grades and of different possible functional forms. For instance, an increase in the supply of one grade depresses the price of that grade and thereby reduces the premium or widens the discount of that grade relative to others. However, the lower price for one grade leads to reduced demands for the other related grades and hence tends to lower prices of these other grades. These lower prices, in turn, feed back to the demand for the grade with the initial price change. The larger supply for one grade, then, tends to lower all prices as well as to change the price differences among grades. The precise size of the changes depends on the size of the increase in supply, the size of the relevant own-price elasticity of demand, and the sizes of the relevant cross-price elasticities with the other grades.

As noted above, large changes in quality premiums can occur. A decrease in supply of a grade with attributes demanded by processors for a specialized use (so that the demand is price inelastic for this use) can push the premium to a high level. Conversely, an increase in supply may reduce the premium to nearly zero.

Empirical analyses of premiums and discounts associated with the attributes of commodities are increasingly common based either on secondary data, usually retail scanner data, or on experimental auctions. The retail scanner data approach to analyzing the relationship of prices to attributes uses data collected on transactions prices and on the attributes of the transactions. These data are usually cross-sectional and obtained over a short period of time so that the general factors influencing the overall price level are more or less constant. If the data are obtained through time, then the model needs to be modified to capture common factors influencing the price level of all the grades of the commodity. As already discussed, the characteristics can be treated as predetermined, and hence the variability in transactions prices can be viewed as a function of the variability in attributes:

$$P_i = f(Z_{1i}, Z_{2i}, \ldots, Z_{Ni}),$$

where the $P_i$ are the transactions prices for $i = 1, 2, \ldots, I$ transactions, and $Z_{in}$ are the characteristics affecting the prices for a transaction $i$ with $n = 1, 2, \ldots, N$.

There are numerous examples of hedonic models using scanner data, and such models are necessarily commodity specific because each commodity has its own set of relevant attributes. Carew (2000), for example, modeled the weekly prices of apples to be used as fresh fruit over 3 marketing years. Characteristics considered in the model included the apple variety (e.g., Red Delicious), grade, package size, fruit size, season of the year, and type (method) of storage. Since the overall price level is likely influenced by the quantity of apples being sold, the model also included weekly shipments of apples. The general point is that the characteristics used in the model are those thought to be important in influencing the prices of apples. The challenge in empirical analysis is the appropriate definition of the relevant attributes (for an application to beef, see Hahn and Mathews 2007; for an application to dairy products, see Chavas and Kim 2005; for an application to fluid milk, see Buccola and Iizuka 1997; and for an application to feeder cattle, see Coatney, Menkaus, and Schmitz 1996).

If prices and quantities are available by grade, then in principle a system of demand equations could be estimated for the individual grades. In this case, the emphasis is on estimating elasticities. Beare and Meshios (1990), for example, estimated a set of equations for wool in Australia, in which the equations are for eight diameters of the wool fibers (19, 20, . . . , 26 microns). Data were obtained for 10 years, and with 8 fiber categories, the analysts had 80 observations. The analysts obtained own-price elasticities that ranged from −1.0 to −2.0. Most cross-price elasticities were positive and statistically significant. For example, the cross elasticity for the quantity of 19-micron fiber wool with respect to the price of 20-micron fiber was estimated to be 0.52.

An alternative and popular approach to using transactions price and attribute data is to use experimental auctions to elicit consumers' willingness to pay (WTP) for a product. WTP is not the same as a transactions price; rather, it represents the maximum that a person is willing to pay for a good, which may be higher or lower than the actual price in the market. Experimental and behavioral economists have generally elicited WTP using either hypothetical surveys or auctions conducted in a laboratory. Using the WTP method, hedonic analyses are typically conducted by dividing different product attributes into separate "treatments" and then testing whether subjects in each treatment have significantly different WTPs. For example, suppose one was interested in determining whether consumers would pay a premium for Fair Trade coffee that was identical in quality to a regular blend of coffee. One could test this by comparing the

WTP elicited from one set of subjects where Fair Trade coffee is auctioned to the WTP elicited from another set of subjects where regular coffee is auctioned.[4]

Such methods have been used to look at a variety of issues of importance in agricultural markets. Many these studies have focused on food items containing genetically modified organisms (GMOs), which have varying levels of acceptance by buyers. Lusk et al. (2005) conducted a meta-analysis based on 25 studies encompassing 57 food items containing GMOs in 12 different countries; they found that, on average, consumers' WTP for foods containing GMOs was 23–28 percent lower than for their non-GMO counterparts. Fox, Hayes, and Shogren (2002) investigated the impact of irradiation on meat demand under positive, negative, and mixed (both positive and negative) information treatments and found that WTP for pork was significantly impacted by the framing of the information. The authors found that favorable information about meat irradiation increased consumers' WTP for pork, while negative and mixed information decreased consumers' WTP. Bernard and Bernard (2009) used experimental price auctions to decompose consumers' WTP for various attributes of organic milk, with an emphasis on the recombinant bovine somatotropin (rBST)–free and no-antibiotic characteristics. They found that consumers place a significant value on both of these attributes of organic milk.

## Defining Grades and Market Imperfections

Two important decisions are required in establishing a grading system. The first involves determining the attributes of the product to use as a basis for defining grades. The second decision involves, given information on the attributes, determining how should they be used and reported. A grading system is usually based on numerous characteristics; moreover, these attributes often vary continuously. Nevertheless, a relatively small number of grades must be established for practical use. What attributes should be used to define grades, and how should boundaries between grades be established (e.g., see Zusman 1967)?

If markets were perfectly competitive, a grading system would not be necessary. By definition, all buyers and sellers would have equal access to the available information, and each transaction would price the item correctly. In practice, information is not costless and may not be equally available to all market participants. In this context, a grading system can potentially make markets work more efficiently (e.g., Hudson, Ethridge, and Segarra 1998).

Assuming the grading is done by an unbiased inspector, buyers and sellers will have equal, fair information, and uncertainty about the quality of item is

4. There are numerous types of auctions that are used to derive WTP, including the English auction, Dutch auction, Vickrey auction, and Becker-DeGroot-Marschack auction. Auctions are described in detail in Chapter 11.

reduced. This situation is in contrast to the case of information being held asymmetrically. For example, the seller may be aware of a defect, but the buyer is not. This is the potential problem of "lemon" automobiles. Also, contracts can be based on grade specifications; buyers may not need to personally inspect specific lots of the commodity before consummating a transaction. Thus, grades may reduce marketing costs.

For grading to serve this purpose, the attributes used to define the grades must be related to the demands for the product. The design of a grading scheme must consider those quality attributes and defects that are economically important to buyers. But when a commodity has a variety of uses and buyers place different values on particular attributes, a simple, consistent grading scheme may be difficult to design. Moreover, some characteristics are relatively inexpensive to measure (e.g., percentage of foreign matter), while other attributes may be costly to measure (e.g., protein content or amount of pesticide residue).

As the foregoing discussion implies, trade-offs exist between the benefits of a grading system and its costs. Improving the information available about the characteristics of individual lots of a commodity reduces transactions costs associated with trading commodities with uncertain quality attributes. The benefits, however, must be compared with costs of improved information and grading. The costs of running a grading program depend partly on the measurement costs of characteristics as well as the costs of the distribution of this information. In principle, the grading/information system should balance marginal benefits with marginal costs, but this is easier said than done (e.g., Chvosta, Rucker, and Watts 2001).

A related concern is the incidence of costs and benefits of a grading program. Historically in the United States, official grades for agricultural commodities were promulgated by the federal government and the costs were borne by taxpayers. Official grades are difficult to change, however, and sometimes buyers develop their own standards. In this case, the buyers think that the benefits of the private system exceed the costs, but sellers relying on the private system must trust that the grading is being done fairly and correctly. The issue is further complicated by the development of new varieties of a commodity using GMOs. It may be important to segregate one variety from another. If so, it is not only a matter of identifying the separate varieties but also of not commingling them in the marketing system. Maintaining separate identities is costly, and cheating can occur.

The literature in economics on these topics is growing rapidly, but it is beyond the scope of this book to elaborate. The general point for our purposes is that appropriate prices for individual lots of a commodity depend on high-quality information. This information can be costly to obtain. A potential source of imperfect market behavior is poor-quality information and/or the asymmetric distribution of information between buyers and sellers.

## Price Discrimination and Government Programs

Differences in demands among grades may provide the basis for a price discrimination scheme (Chapter 5). Agricultural cooperatives in the United States, using the provisions of marketing orders, are occasionally able to do this. For example, quantities sold in the higher-price fresh fruit market may be restricted, with the remaining fruit going to processing markets (that presumably have a more elastic demand).

Marketing orders also permit the culling of low-quality produce (i.e., setting minimum quality standards for the sale of fruits and vegetables). A part of the rationale for minimum standards is to assure consumers of a high-quality product, but the main objective appears to be increased returns for producers (for a review, see Bockstael 1987). Nguyen and Vo (1985) discuss the conditions under which discarding part of the supply will lead to increased producer returns. It is possible for returns to increase even when demand is price elastic because culling low-quality product, in effect, shifts the demand function for the remaining fruit. Thus, the reduction in supply does not result simply in a new point of equilibrium on a static demand curve.

If, however, consumers are fully informed about the alternative qualities available, minimum grade standards result unambiguously in a welfare loss (Bockstael 1984, 1987)—consumers have been deprived of an opportunity to buy low-quality produce (at a low price), and the gains in farmer returns are not sufficient to compensate for this loss. Minimum quality standards perhaps can be justified where information is held asymmetrically and, hence, consumers are uncertain about the quality of the produce.[5] Even in this case, the tendency will be to set the minimum marketable grade too high (from a consumer welfare viewpoint) when the standards are set by producers (Bockstael 1987).

Chalfant and Sexton (2002) point out that errors can occur in enforcing the grading standards that are promulgated under market orders. Because the demand for high-quality produce is likely to be more price inelastic than the demand for low-quality produce, purposeful grading errors have the potential to increase the collective profit of the industry. Even if the errors are not purposeful, Chalfant and Sexton's article reminds us that grading methods are subject to error. As noted in the previous section, grading systems involve both benefits and costs.

Price differentials serve an important function provided that they reflect the relative demands of consumers for different grades and varieties. Premiums or

5. For example, if consumers cannot tell immature from ripe fruit before purchase, then it is in a producer's self-interest to harvest immature fruit to market the fruit prior to the main harvest and thereby obtain a higher price. Such individual decisions collectively will result in a reduction in the overall quality of fruit coming to market. In effect, lower-quality fruit drives good fruit out of the market, and regulation of minimum quality becomes a way to avoid the market failure associated with the lack of information available to consumers.

discounts for various grades do not perform this function effectively if they are arbitrarily maintained on the basis of historical relationships, such as may occur under government price-support programs. Thus, price-support programs are another type of "imperfection" for some agricultural commodities. At times, lower grades of tobacco, cotton, and wheat have been overpriced relative to higher grades (i.e., the discount on such grades has not been sufficiently wide). As a result, the lower grades have ended up in government warehouses, while the higher grades have moved into consumption. By over-pricing the lower grades, the government also provides an incentive for farmers to continue producing excess quantities of the inferior grades of the product (i.e., quantities supplied exceed the quantities demanded at the regulated prices).

## References

Arzac, E. R., and M. Wilkinson. 1979. "A Quarterly Econometric Model of the United States Livestock and Feed Grain Markets and Some of Its Policy Implications," *Am. J. Ag. Econ.* 61:297–308.

Beare, S., and H. Meshios. 1990. "Substitution between Wools of Different Fibre Diameter," *Australian J. Ag. Econ.* 34:56–66.

Bernard, J. C., and D. J. Bernard. 2009. "What Is It about Organic Milk? An Experimental Analysis," *Am. J. Ag. Econ.* 91:826–836.

Bockstael, N. E. 1984. "The Welfare Implications of Minimum Quality Standards," *Am. J. Ag. Econ.* 66:466–471.

———. 1987. "Economic Efficiency Issues of Grading and Minimum Quality Standards," in *Economic Efficiency in Agricultural and Food Marketing*. R. L. Kilmer and W. J. Armbruster, eds. Ames, Iowa: Iowa State Univ. Press. Chapter 12.

Bowman, K. R., and D. E. Ethridge. 1992. "Characteristic Supplies and Demands in a Hedonic Framework: U.S. Market for Cotton Fiber Attributes," *Am. J. Ag. Econ.* 74:991–1002.

Buccola, S., and Y. Iizuka. 1997. "Hedonic Cost Models and the Pricing of Milk Components," *Am. J. Ag. Econ.* 79:452–462.

Carew, R. 2000. "A Hedonic Analysis of Apple Prices and Product Quality Characteristics in British Columbia," *Canadian J. Ag. Econ.* 48:241–257.

Chalfant, J. S., and R. J. Sexton. 2002. "Marketing Orders, Grading Errors, and Price Discrimination," *Am. J. Ag. Econ.* 84:53–66.

Chavas, J. P., and K. Kim. 2005. "Cointegration Relationships and Hedonic Pricing of Differentiated Commodities: An Application to Price Dynamics in the US Dairy Sector," *Applied Econ.* 37:1813–1827.

Chvosta, J., R. R. Rucker, and M. J. Watts. 2001. "Transactions Costs and Cattle Marketing: The Information Content of Seller-Provided Presale Data at Bull Auctions," *Am. J. Ag. Econ.* 83:286–301.

Coatney, K. T., D. J. Menkhaus, and J. D. Schmitz. 1996. "Feeder Cattle Price Determinants: An Hedonic System of Equations Approach," *Rev. of Ag. Econ.* 18:193–211.

Fox, J. A., D. J. Hayes, and J. Shogren. 2002. "Consumer Preferences for Food Irradiation: How Favorable and Unfavorable Descriptions Affect Preferences for Irradiated Pork in Experimental Auctions," *J. of Risk & Uncertainty* 24:75–95.

Hahn, W. F., and K. H. Mathews. 2007. "Characteristics and Hedonic Pricing of Differentiated Beef Demands," *Ag. Econ.* 36:377–393.

Hudson, D., D. Ethridge, and E. Segarra. 1998. "Incorrect Price Information for a Heterogeneous Commodity: A Conceptual Synthesis," *Rev. of Ag. Econ.* 20:365–376.

Lusk, J. L., M. Jamal, L. Kurlander, M. Roucan, and L. Taulman. 2005. "A Meta-Analysis of Genetically Modified Food Valuation Studies," *J. of Ag. & Res. Econ.* 30:28–44.

Nguyen, D., and T. V. Vo. 1985. "On Discarding Low Quality Produce," *Am. J. Ag. Econ.* 67:466–471.

Rosen, S. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *J. Pol. Econ.* 82:34–55.

Waugh, F. V. 1929. *Quality as a Determinant of Vegetable Prices*. New York: Columbia University Press.

Zusman, P. 1967. "A Theoretical Basis for Determination of Grading and Sorting Schemes," *J. Farm Econ.* 49:89–106.

CHAPTER 8

# Spatial Price Relationships

An important attribute of individual lots of agricultural products is their location, and thus one should not be surprised that prices of a commodity vary regionally. The price of a particular grade of corn in Ames, Iowa, need not be—indeed probably will not be—the same as the price of the same quality corn in Batavia, New York. Likewise, the price of raisins in the United States will generally not be the same as it is in Japan. This chapter discusses the factors that cause prices to differ among regions, within and among countries, and that cause these relative prices to change with the passage of time.

## Background

Given a competitive market structure, spatial price relationships are determined by transfer costs among regions. Transfer costs, as discussed below, include a variety of costs beyond transportation rates, and indeed, the definition of costs can be broadened in an international context to include trade barriers such as tariffs, quotas, and subsidies. One of the potential problems in analyzing spatial price differences is the lack of a complete understanding of transfer costs and the possible changes in these costs with the passage of time. The relevant costs are the full costs of spatial arbitrage.

Transfer costs can be large relative to the farm value of the commodity. Wheat straw is an extreme example; it has a low on-farm value and is bulky and costly to transport. Intuitively, the transfer costs for straw are an important barrier to trade over long distances. Although straw is an extreme example, it is generally the case that the value of agricultural commodities per unit of weight is less than the value of manufactured goods per unit of weight.

As noted in Chapter 6, transportation costs are a component of marketing costs. If transportation costs are large relative to the farm value and if these costs are fixed, then a consequence is that small percentage changes in retail prices can be magnified into large percentage changes at the farm. In this chapter, however, we limit our discussion to regional and international prices for a specific commodity at a given (usually the farm) level of the marketing system.

In this context, a model is used to illustrate the impact of changes in demand, supply, or transfer costs on price differentials among regions and countries. This *spatial equilibrium model* also provides a convenient analytical framework that can be used to determine the quantities and directions of trade among regions. In addition, this model is used to illustrate the price effects of relaxing or increasing trade barriers between countries. Another dimension of this literature relates to the efficient market hypothesis and market integration, and we discuss the implications of a model of spatial price relationships for the empirical study of market efficiency. The principles that determine spatial price differences within a country apply equally to international prices when no trade barriers exist to the movement of commodities among countries. However, modifications in the spatial equilibrium model are necessary when trade barriers are present; this is examined in this chapter as well.

The principles that underlie price differences between regions (assuming a competitive market structure with a homogenous commodity and no trade barriers) can be summarized as follows:

1. Price differences between any two regions that trade with each other will just equal transfer costs.
2. Price differences between any two regions that do not engage in trade with each other will be less than or equal to the transfer costs.

The first principle is commonly referred to as the *law of one price* in a spatial dimension, i.e., prices between regions can only differ by the magnitude of the transfer costs and they change together with the passage of time.

Under competitive conditions and no trade barriers, price differences between regions cannot exceed transfer (arbitrage) costs. The reason for this should be obvious: any time the price difference is greater than transfer costs, buyers will purchase the commodity in the low-priced market and ship it the higher-priced market, thereby raising prices in the former and reducing them in the latter. This is a form of arbitrage across space, and it will continue until the price difference between markets no longer exceeds the costs of arbitrage.

Using these principles, theoretical spatial price relationships, sometimes called the *structure of prices*, can be determined. The structure of prices is a function of the pattern of trade (i.e., who ships to whom) and transfer costs per unit of product between regions that engage in trade. Where no trade exists between regions, an upper, but not a lower, price boundary can be established since, as pointed out above, the difference in price cannot exceed transfer costs between any two regions. In the absence of trade, the precise structure of prices

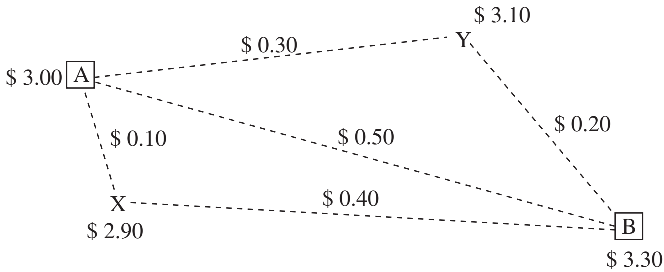



**Figure 8-1.** Hypothetical markets (A and B), production areas (X and Y), transfer costs (on dashed lines), and prices for corn

cannot be determined solely on the basis of transfer costs. They are easily determined, however, if all surplus-producing areas ship to one central market (see later in the chapter). The determination of spatial price relationships becomes much more complicated when there are many regions or points of origin for surplus commodities and several markets to which the surpluses may be sent. The optimum pattern of trade in that case may not be intuitively obvious, and until the pattern of trade is identified, the structure of prices cannot be determined.

The principles involved in determining the geographical structure of prices can be illustrated assuming two marketplaces, A and B, to which corn is shipped (Figure 8-1). Two points of surplus corn production are identified in Figure 8-1 as X and Y. Transfer costs between each location and each market are indicated along the dashed lines that connect all consumption and production areas. Using simple arithmetic, the prices that would prevail at every location can be determined once the price is given at any one location. If the market price is $3.00 per bushel at location A and if corn can be shipped from A to B at a cost of $0.50 per bushel, the maximum price that will prevail at B is $3.50 per bushel. But corn can be purchased at location X for $2.90 per bushel (the price at A less the $0.10 transfer cost). Corn purchased at X can be delivered to B at a price of $3.30 per bushel ($2.90 plus $0.40 transfer cost). The price of corn at B is dictated by the least-cost procurement price ($3.30 rather than $3.50 per bushel). Similarly, producers located at Y have the alternative of marketing their corn in either market A or B at a price competitive with that prevailing in each market. If they ship to A, the price will be $3.00 less $0.30 (i.e., $2.70 per bushel), while if they ship to market B, the price will be $3.30 less $0.20 (i.e., $3.10 per bushel). Thus, shipping to market B will be more profitable and will dictate the price at location Y.

The ideas illustrated by this example can be summarized as follows:

1. The lowest-cost source determines the price prevailing in each deficit market.

2. Producers sell in whatever market yields the highest net price.
3. The price prevailing in each surplus-producing area is the deficit-market price less the cost of transferring a unit of product to that market.

Of course, the foregoing assumes a competitive market structure, with a homogeneous commodity, informed traders, and no barriers to trade.

## Price Relationships for Commodities Sold in One Central Market

From the foregoing discussion, it follows that if producers ship homogenous units of the same commodity to a single central market, the price each producer receives under perfectly competitive conditions is the central market price less the cost of transferring a unit of that commodity to the central market. This is based on the reasonable assumption that buyers are indifferent to the source of supply of a homogeneous product and will not therefore pay more for a unit of product from one area than from another. If producers in a particular region offered their product for less, this would cause prices for the same commodity produced in other regions to fall by an equal amount. Such price adjustments occur because of competitive arbitrage.

The impetus for price changes can come either from supply or demand changes, and central markets can be the locus for the formation of prices. In this case, the price offered to farmers by local buyers equals the central-market quote minus the transfer costs. In an economy like the United States, central cash markets no longer play an important role in price discovery for most commodities, and the geographic structure of prices is typically more complex than the one just discussed.[1] Nonetheless, inter-regional price differences, as already pointed out, cannot exceed the cost of moving commodities between regions for very long since the competitive market will quickly equalize prices net of transfer costs. In the next several sections, we consider increasingly complex models.

## Market Boundaries

Market boundaries define the geographic scope of the producers and buyers of a particular commodity and provide a natural definition of what is a particular market. Determining the boundaries of a market is important for policies such as federal and state milk-marketing orders in the United States, since these orders are instituted in regional milk markets. Market boundaries are also

1. As will be discussed in Chapter 11, *price discovery* refers to the institutional arrangements for arriving at specific prices. For some agricultural commodities in the United States, commodity futures markets are an important component of price discovery (Chapter 12). These are central markets for the trading of contracts for future delivery of the commodity; they typically are not central markets for the physical delivery of the commodity.

important for determining the most efficient distribution system in supply chain management. If producers have the option of shipping to different markets, the boundary between supply areas is determined by the price at each destination less the cost of transferring the product from each point of origin to each destination. Natural barriers such as rivers or mountain ranges and artificial barriers such as political boundaries may determine the dividing line between supply areas for different markets. But where such barriers do not exist and the prices paid to producers decline continuously as the distance from each market increases, the boundary between supply areas can be determined by drawing concentric circles around each market and noting the points at which producer prices (net of transfer costs) are the same. Given free choice, producers will always ship to the market offering the highest net price. The locus of points where the price is the same determines the market boundary.

The boundary between two markets will shift if there is a change in the relative prices in the two markets or if transfer costs change. Differential rates of growth in demand, for example, may cause producer prices to rise in one market relative to another. This will tend to widen the market area serving the market with the increased demand while increasing prices in both markets. This type of effect is illustrated in Figure 8-2. The sloping lines originating at each vertical axis show the net price that would be paid to a producer located at varying distances from markets A and B. At the initial price of $6 per unit in market A and $5 in market B, and with transfer costs equal to 50 cents for each 100 miles in both markets, the boundary point would be located 400 miles from market

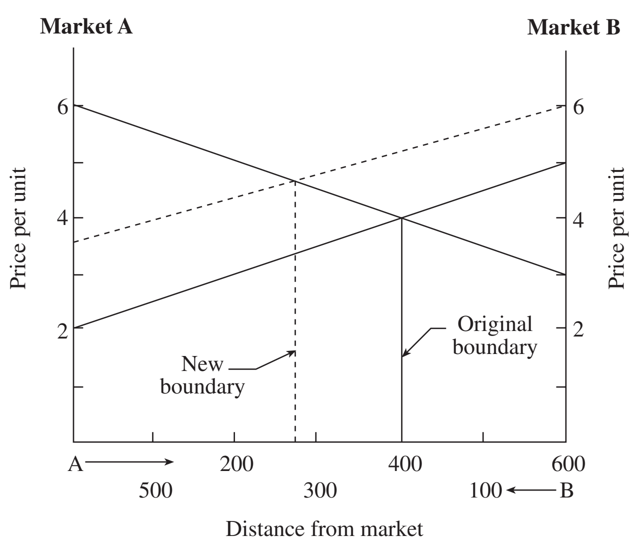



**Figure 8-2.** Effect of changes in market prices and transfer costs on the boundary between markets (price in dollars, distance in miles)

A and 200 miles from market B. The price at this location would be \$4 per unit [\$6 − (4 × 0.5) = \$5 − (2 × 0.5)].

If the price in market B rises to \$6 per unit, and if at the same time a more efficient transfer system is introduced in that market but not in the other, thereby reducing transportation costs to market B by 20 percent (i.e., to 40 cents per 100 miles), the new net price line for market B will be shifted upward and will be less steeply sloped. The new position is indicated by the dashed line in Figure 8-2. Some producers located along the axis joining the two markets will shift from market A to market B; hence, the new boundary point will be located to the left of the old boundary. This point is not, however, the final solution because the reduced supply marketed in A will tend to raise the price there. Ultimately, the higher prices received by farmers will influence the quantity supplied. As discussed in earlier chapters, the initial change in demand results in a price adjustment process in both markets. The preceding illustration shows how *one point* on a boundary line between markets may be established. Prices, of course, will vary along the boundary line, depending on the distances from each market. Boundary lines between markets may be linear under some conditions and curved under others. When more than two markets are involved, the locations of boundaries become more complex.

A market area boundary line can be identified by noting the points at which prices paid to producers, net of transfer costs, are the same whether they ship to one market or another.[2] The manner in which theoretical boundary lines are established between adjacent markets is illustrated in Figure 8-3. In this illustration, the price at market A is \$6 per unit and the price at market B is \$5. Prices paid to farmers shipping to these two markets are assumed to decline uniformly in direct relation to the distances from each market; that is, the transfer cost is a linear function of distance, increasing 50 cents each unit of distance. Under these assumed conditions, the boundary is represented by a curved line passing through the points *m*, *n*, *o*, and *p*. These are the points at which producers are indifferent as to which market they supply.

The net price from shipping to alternative markets is the same at each point on the boundary, but the net price varies along the boundary. Farmers located at points *m* and *p* receive \$3.50 per unit regardless of where they send their product, while producers at points *n* and *o* receive \$4. In this example, the boundary is a curved line because the price is higher in one market center than in the other.[3] If the shipping costs are uniform, the boundary line will inevitably lie closer to the lower-priced market than to the higher-priced market.

2. This statement can be formalized algebraically. Let $P_A$ be the price at market center A, $P_B$ the price at market center B, $T_A$ the transfer cost from farm to market A, and $T_B$ the transfer cost from farm to market B. Then the boundary is defined by the points where $P_A - T_A = P_B - T_B$.

3. Given the assumption that transfer costs are a linear function of distance, the market boundary line is a hyperbola. Bressler and King (1970: 126–129, 189–195) provide a much more complete discussion of the theory and practice of establishing boundaries between markets under various conditions, including those associated with differences in product forms.

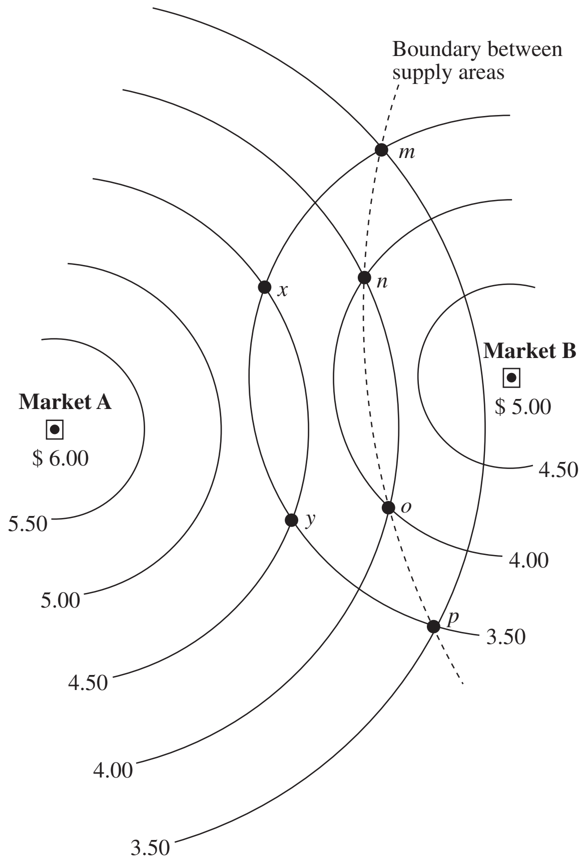



**Figure 8-3.** Location of the boundary between areas supplying alternative markets

If prices in the two market centers were the same, the price boundary would lie halfway between the two markets and would be a straight line perpendicular to one connecting the two centers. Such a boundary line would go through points *x* and *y* in Figure 8-3. Since the same cost-distance relation is assumed to hold for shipments in either direction, the straight-line boundary joins points that are equidistant from the two markets.

The foregoing discussion is about economic boundaries. Research problems are perhaps more often about trade among political units: counties, states, provinces, and countries. In this case, the data relate to production, consumption, and prices within these units as well as trade among the units. This is the topic of the next section.

## Spatial Equilibrium Models

Geographical price relationships can be analyzed by using spatial price equilibrium models. These models, using specific assumptions about the nature of the markets, make it possible to estimate the net price that will prevail in each region and the quantity of a given commodity that any one region will export or import from every other region. Such models determine the optimum (often defined as the least-cost) trading pattern, given the supply and demand conditions within each region. From this optimum trading pattern, an appropriate set of prices is obtained, based on the principles outlined earlier: the difference in price between any two regions that trade with each other will just equal the transfer cost between the two regions, while the price differences between regions not engaging in trade will be less than or equal to the transfer costs. That is, the law of one price prevails in competitive markets when there are no trade barriers.

Spatial equilibrium models are most useful in analyzing inter-regional price relationships and trading patterns where numerous consuming and producing regions exist. If each region produces as well as consumes the product, one cannot always determine by inspection just which areas will have excess supplies available for sale to deficit regions and which will require imports. Nor is it always obvious which surplus regions will supply a particular deficit area.

Although realistic models need to be formulated mathematically, the general principles involved in developing inter-regional trade models can be illustrated with the aid of supply and demand diagrams for two regions (e.g., Houck 1992). Consider two regions with the following demand and supply functions:

| Region A | | Region B | |
|---|---|---|---|
| $Q_{DA} = 100 - P_A$ | (demand in A) | $Q_{DB} = 150 - P_B$ | (demand in B) |
| $Q_{SA} = 40 + P_A$ | (supply in A) | $Q_{SB} = 50 + P_B$ | (supply in B) |

In the absence of trade (referred to as *autarky*), demand and supply are equated at a price of \$30 in region A and \$50 in region B. If we allow for the possibility of trade between regions, then it should be clear that region A will be the exporter and region B will be the importer, since the autarky price is lower in region A. To solve for the spatial equilibrium, we need to construct an *excess supply* curve for the exporting region and an *excess demand* curve for the importing region. Notice that a price above \$30 would create excess supply (supply greater than demand) in region A, which could become available for shipment to another area. For region B, a price below \$50 would create excess demand (demand greater than supply), which could be satisfied by imports to region B. Algebraically, region A's excess supply function ($ES_A$) is simply its regional supply minus demand, while region B's excess demand function ($ED_B$) is its regional demand minus supply:

| Region A | Region B |
|---|---|
| $ES_A = 40 + P_A - (100 + P_A)$ | $ED_B = 150\ P_B - (50 + P_B)$ |
| $= -60 + 2\ P_A$ | $= 100 - 2P_B$ |

The excess supply curve for the surplus region is displayed in Figure 8-4 and is equal to the horizontal distance between the supply and demand curves in region A at prices above the point of equilibrium (e.g., point *b* minus point *a* in the Region A panel). Excess supply is zero at the equilibrium price of \$30. The excess supply curve is positively sloped like conventional supply schedules since the gap between supply and demand in region A widens as the price

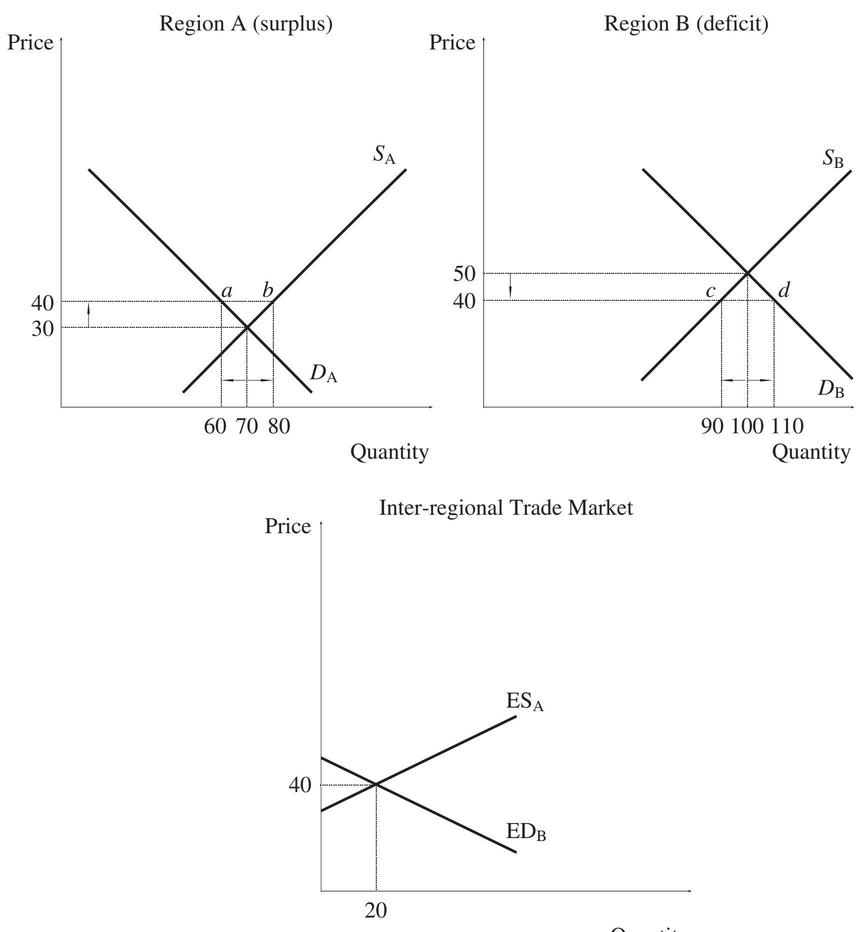



**Figure 8-4.** Two-region spatial equilibrium model with no transfer costs

increases. The horizontal distance between the demand and supply curves below the point of equilibrium in deficit region B (e.g., point *d* minus point *c* in the Region B panel) provides the information needed to construct the excess demand curve. The excess demand curve is negatively sloped since the gap between the quantities demanded and supplied widens as the price declines. The excess demand function intersects the vertical axis at the equilibrium price of $50 per unit, since there will be no unfilled demand at this price.

There are two important properties of the excess supply and demand curves. First, the slopes of excess supply and excess demand curves are functions of the slopes of the underlying regional supply and demand curves for each region. The steeper (flatter) the regional supply and demand, the steeper (flatter) the excess supply and excess demand curves. Second, since both supply and demand of a region can react to a price change, the excess supply or excess demand curve will always be more price elastic than the underlying regional supply or regional demand curve. This result implies that price volatility caused by supply or demand shocks will generally be larger in the domestic than in the traded market.

The excess demand and supply schedules shown in Figure 8-4 intersect at an equilibrium price ($P^*$) of $40 per unit (Inter-regional panel). If no transfer costs exist between the two regions, a total of 20 units of the commodity will be shipped from region A to region B ($ab = cd = 20$ units). The price in both regions will then be the same, $40 per unit. Notice that the introduction of trade has two effects in each region, a demand effect and a supply effect. In region A, the introduction of trade raises the equilibrium price from $30 to $40, thus causing the quantity demanded to fall from 70 to 60 units and the quantity supplied to rise from 70 to 80 units. In region B, the effect is opposite, i.e., the lower price caused by trade increases the quantity demanded from 100 to 110 units and decreases the quantity supplied from 100 to 90 units. Therefore, the possibility of trade hurts consumers in the surplus market and producers in the deficit market, while it helps producers in the surplus market and consumers in the deficit market. Hence, there are gainers and losers from trade.

This analysis is based on the unrealistic assumption that transfer costs are zero. When considering positive transfer costs, notice that no trade would occur if it cost more than $20 to move a unit of product from region A to B, which is the difference in the price intercepts for the excess supply and demand curves and is also equal to the difference in autarky prices for the two regions. In that case, demand and supply would be equated within each region and the price difference would be less than the transfer cost. The effect of introducing (or changing the amount of) transfer costs on the amount shipped between regions can be illustrated by superimposing a wedge between the excess supply and demand curves equal to the per unit transfer costs. To illustrate, consider a transfer cost of $10 per unit shipped, which is shown as the bold vertical line in the Inter-regional panel of Figure 8-5. In the example, since the slopes of the demand and supply schedules in both regions are equal in absolute value, intro-

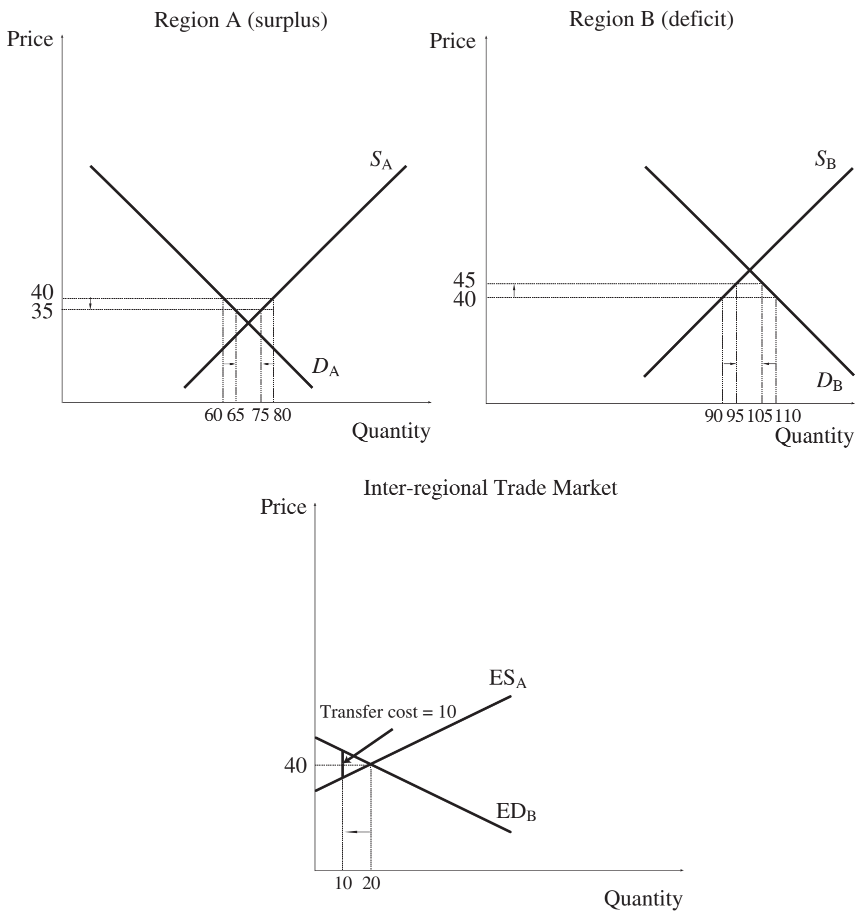



**Figure 8-5.** Two-region spatial equilibrium model with \$10 transfer costs

ducing a transfer cost of \$10 per unit will reduce the price in region A (the surplus region) by \$5 per unit and raise the price in region B by an equivalent amount. That is, with a \$10 transfer cost, the price in region B rises from \$40 to \$45 and the price in region A falls from \$40 to \$35.[4]

4. This can be determined algebraically by using the law of one price, which means that either (1) the price in the surplus region equals the price in the deficit region minus \$10 (the transfer costs) or (2) the price in the deficit region equals the price in the surplus region plus \$10. Using the latter, we can write: $P_B = P_A + 10$. Substituting this relationship into the excess demand curve for region B yields:

$$\mathrm{ED_B} = 100 - 2P_B = 100 - 2(P_A + 10) = 80 - 2P_A.$$

Equating $\mathrm{ED_B}$ with $\mathrm{ES_A}$, and solving for $P_A$ results in $P^*_A = 35$. From the law of one price, $P^*_B = 35 + 10 = 45$. Finally, $Q^*$ can be determined by substituting $P^*_B$ in the ED function or $P^*_A$ into the ES function: $Q^* = -60 + 2\ (35) = 10$.

We can use this model to answer the question, how much of the transfer cost is paid by the exporter versus the importer? For the exporting region, the effect of introducing a \$10 transfer cost is to decrease the price received from \$40 to \$35, and therefore \$5 of the transfer cost is borne by the exporters. The importing region's price increases from \$40 to \$45 due to the \$10 transfer cost and consequently \$5 of the transfer cost is paid by importers. Does this mean that transfer costs are always shared equally by exporters and importers? No. This will only be the case if the relative price elasticities of excess demand and supply are equal. Alternatively, if the price elasticity of excess demand is more elastic than that for excess supply, exporters will pay a higher percentage of the transfer cost. This occurs because the importers are more price sensitive than the exporters and shift more of the incidence of the transfer cost to the exporters. Since the relative price elasticities of excess supply and demand depend on the region's supply and demand curves, the incidence of the effects of changes in transfer costs between surplus and deficit regions is dependent on the nature of supply and demand in each region.

The pattern of trade and the quantity shipped from one region to another may be altered by shifts in regional demand and supply functions, but such changes will not necessarily alter the structure of prices. Price differences between regions (assuming they continue to trade) will remain the same as long as the transfer costs do not change. The structure of prices is a function of trading patterns and transfer costs. Regional shifts in demand or supply are likely to alter the quantity imported (exported) and may result in a higher or lower average level of prices, but unless one or more regions shift from a net deficit to a net surplus position (or vice versa), price differences between regions will not change.

## Impact of Trade Barriers

Spatial equilibrium models are also useful in examining the impact of trade barriers on prices and trade when governments impose such policies. Governments use trade barriers to raise revenue, to protect producers in certain markets, and to subsidize consumers. There are four general types of trade distortions that impact price:

1. import curtailments (import tax and quota),
2. export expansion (export subsidy),
3. import expansion (import subsidy), and
4. export curtailment (export tax and quota).

Governments use 1 and 2 for supporting domestic producer prices and use 3 and 4 for lowering domestic prices to assist consumers. In the following discussion, three types of pro-producer trade barriers are examined: import tariffs, import quotas, and export subsidies. These three policies are chosen because they reflect U.S. international policies, which aim primarily to protect domestic producers at the expense of domestic consumers. Although not discussed here, three trade barriers aimed at helping domestic consumers are import subsidies, export quotas, and export taxes. The impacts of these policies are the exact opposite of those intended to help producers.

In the analysis that follows, we assume that there are two regions in the world: the United States and the rest of the world (ROW). Also we assume the United States is a large country, in the sense that its quantity levels (imports or exports) in the world market impact the world price. In other words, if the United States is an importer, it faces an upward-sloping excess supply curve from the ROW, and if the United States is an exporter, it faces a downward-sloping excess demand curve from the ROW. Finally, for clarity of graphical illustration, we assume that the transfer costs are zero.

### Import Tariff

An import tariff is a tax that is levied on all imports entering the United States. An import tariff can be *ad valorem*, which means it is a percentage of value, e.g., 20 percent of the world price, or fixed rate, which means it is a constant amount per unit of quantity, e.g., $2 per pound. We will consider the effects of a fixed-rate tariff here, but the effects of the *ad valorem* tariff are similar. From Figure 8-6, the equilibrium price and import level in the world market under free trade is $P_W$ and $Q_W$, respectively, and the domestic equilibrium price, quantity demanded, and quantity supplied in the United States is $P_D$, $q_D$, and $q_S$, respectively.

Similar to a domestic tax, a fixed-rate import tariff has the impact of shifting the demand for the imported product downward, reflected in Figure 8-6 by a decrease in the excess demand (ED falls to ED′) curve of the United States in the international market. Note, ED is the only curve in the spatial equilibrium model that shifts because the import tariff affects excess demand only for U.S.

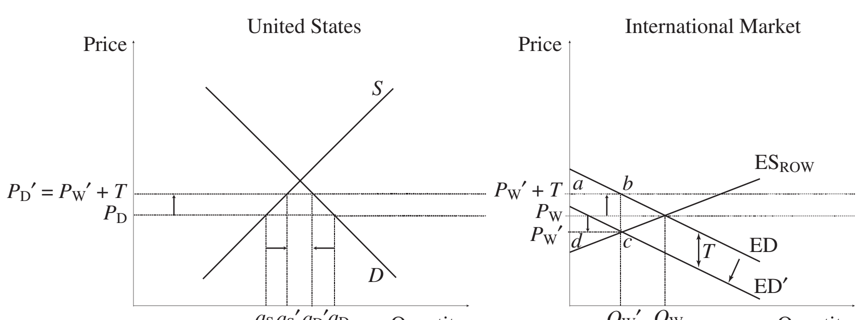



**Figure 8-6.** Fixed-rate import tariff, large importing country

importers, not the domestic demand curve. Since this is a fixed-rate tariff, the difference between ED′ and ED is the constant level of the tariff reflected by a parallel decrease in the excess demand curve. Put differently, for any given level of imports into the United States, the effective price paid by importers is now the old price plus the fixed-rate import tariff. The new equilibrium in the world market is determined by the intersection of the excess supply for the ROW, $ES_{ROW}$, and the ED curve. Therefore, the impact of the fixed-rate import tariff for the world market is to decrease the world price, $P_W$, to $P_W'$ and equilibrium imports into the United States falls to $Q_W'$. The impact in the domestic market is to increase the domestic price to $P_D' = P_W' + T$, where $T$ is the tariff level. Since the United States is a large country, the world price falls due to the import tariff because the United States is buying fewer imports in the world market.

There are several impacts of the fixed-rate import tariff on the U.S. and world markets. First, the increase in the equilibrium domestic price to $P_D'$ has a positive effect on production, reflected by quantity supplied increasing to $q_S'$. Second, the higher domestic price has a negative effect on consumption since the quantity demanded decreases to $q_D'$. Consequently, it should be clear that domestic producers are helped and domestic consumers are hurt by import tariffs. The flatter the slopes of the supply and demand curves, the more production will rise and consumption will fall if a tariff is imposed. Even a modest increase in the degree of protection can lead to a substantial percentage decline in the volume of imports. The percentage effect will depend on the elasticities of the demand and supply schedules in the importing country and the proportion of the total consumption that is imported. Third, import tariffs are revenue sources for the government, and in this case, the government receives the rectangular area *abcd* in revenue, which is equal to the tariff rate times the new equilibrium level of imports. Fourth, this pro-domestic producer trade barrier helps foreign consumers of the commodity since the world price is lower now that the United States is buying fewer imports. On the other hand, foreign suppliers are hurt by the lower world price, which is the reason that less-developed countries, which rely on their own agricultural exports as a source of revenue, disdain pro-producer trade barriers in developed countries. The United States and other countries use import tariffs for many agricultural commodities to try to protect their domestic producers.

## Import Quota

An import quota is a quantitative restriction on the amount of imports allowed to enter domestic market. For example, the United States used to have a very strict import quota for cheese imports that was roughly equivalent to 1 percent of U.S. milk production. Import quotas are often used in conjunction with domestic price-support programs to allow the domestic government to support domestic prices without supporting world prices, which would be far more costly to the government. Supporters of freer trade dislike import quotas even

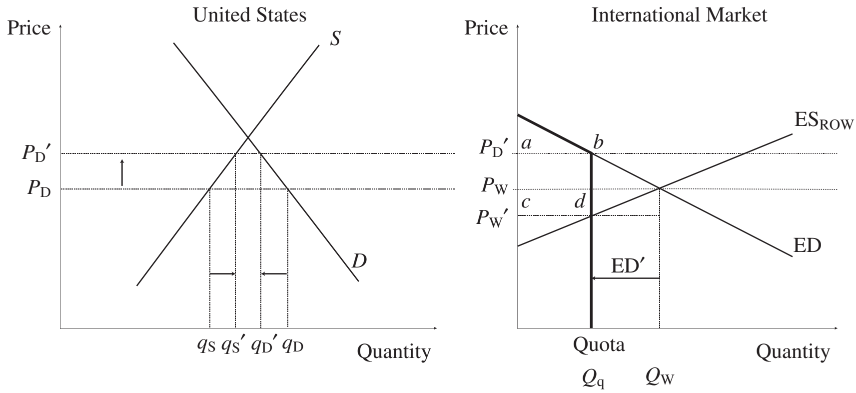



**Figure 8-7.** Import quota, large importing country

more than import tariffs because they are generally the most restrictive barrier to accessing the domestic market. The impacts of import quotas are similar to import tariffs.

The world and domestic markets with an import quota are illustrated in Figure 8-7. Again, let the free-market equilibrium be depicted in the graph by $P_W$, $Q_W$, $P_D$, $q_S$, and $q_D$. For the import quota to have any effect on either the domestic or world markets, it has to restrict imports into the United States so they are less than $Q_W$. Suppose the import quota is set at $Q_q$. This has the impact of changing the excess demand curve for the United States to ED′, which is the same as ED between zero and $Q_q$ imports but becomes perfectly inelastic (fixed) at $Q_q$. The new equilibrium is determined where ED′ equals $ES_{ROW}$. Since the United States is a large country, the world price falls due to the import quota because the United States is now buying fewer imports in the world market due to this absolute restriction. The impact in the domestic market is an increase in the domestic price to $P_D'$ since imports are being restricted relative to free-trade levels.

There are several impacts of the import quota on the U.S. and world markets. First, the increase in the equilibrium domestic price to $P_D'$ has a positive effect on production reflected by quantity supplied increasing to $q_S'$. Second, the higher domestic price has a negative effect on consumption since the quantity demanded decreases to $q_D'$. Hence, it should be clear that domestic producers are helped and domestic consumers are hurt by import quotas. Third, import quotas are a source of windfall gains to whoever holds the quota rights to import; i.e., whoever holds the rights to import is allowed to pay the lower world price for the imports and then sell the commodity in the domestic market at a higher price. Because of this, the quota itself has value, which is reflected by area *abcd* in Figure 8-7. Who gets this revenue depends on how the government allocates the import quotas. If the government sells the import quotas, the government captures these windfall profits and the effects are identical to a tariff. Alternatively, if the government gives the quotas away, then the lucky grantees receive these gains. Fourth, import quotas help foreign consumers of the commodity since the world price is lower since the United States is buying less in the world market. Foreign suppliers are hurt due to the lower world price. The United States and other countries used import quotas for some commodities historically to try to protect their domestic producers and contain price-support program costs, e.g., cheese and casein import quotas, but have abandoned them since the World Trade Organization has mainly eliminated such restrictive trade barriers.

## Export Subsidy

An export subsidy provides an economic incentive to exporters to export more than they would in the case of free trade. These subsidies can be either monetary payments or in-kind incentives, such as providing extra grain to exporters for each ton that is exported. Also, a devaluation of a country's currency is a de facto export subsidy since it lowers the effective price of that country's exports in the world market. In the case of monetary payments, export subsidies can be either *ad valorem* (a percentage of value) or a fixed rate per unit sold. Here we consider the effects of a fixed-rate export subsidy for the United States.

The impact of the United States implementing a fixed-rate export subsidy is illustrated graphically in Figure 8-8, where the excess supply curve for U.S. exports increases from ES to ES′ in the world market. The new curve is parallel to the old one, reflecting the fact that producers now receive a constant per unit subsidy on each unit exported. The equilibrium world market price, determined where ES′ is equal to $ED_{ROW}$, falls to $P_W'$ because the United States is now exporting more in the world market, $Q_W'$. In the United States, the domestic price rises to $P_D' = P_W' + S$, where $S$ is the fixed-rate subsidy. The higher

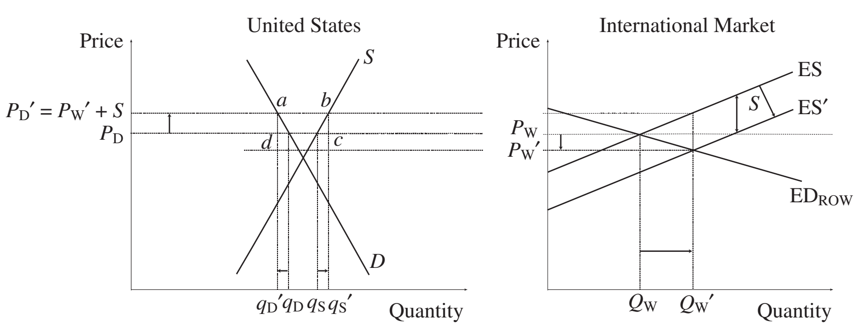



**Figure 8-8.** Fixed-rate export subsidy, large importing country

domestic price causes the quantity supplied to increase to $q_S{}'$ and the quantity demanded to fall to $q_D{}'$. Hence, U.S. producers are better off and U.S. consumers are worse off from export subsidies. Unlike the import tariff, which has similar price and quantity results, export subsidies cost the government money. In this example, the rectangular area *abcd* represents the total monetary costs to the U.S. government that are paid to exporters. This policy has similar effects to the previous two on domestic and foreign producers and consumers. While the United States has extensively relied on export subsidies to enhance agricultural exports, the European Union has relied even more on these subsidies as a means to support their domestic agricultural sector.

## Empirical Applications

When more than two regions are involved, it is difficult to determine the pattern of trade or the structure of prices graphically, and therefore we must rely on mathematical models. To obtain a solution where many regions have been defined, one must know the supply and demand relationships in each region (expressed in algebraic form) and the transfer costs from every region to every other region. Then, the regional supply and demand schedules can be summed and the resulting aggregate supply and demand functions can be solved for the equilibrium price; an early reference is Judge and Wallace 1958, and for a comprehensive treatment, see Takayama and Judge 1971. Once the aggregate equilibrium price is known, the model solves for regional prices, and these prices are inserted into the regional supply and demand functions to determine production, consumption, and net imports and exports by region. In a simple model, the sum of excess supply is assumed to equal the sum of excess demand, but models can be modified to take account of movements into and out of storage, as well as government policies.

Indeed, many models of spatial and trade relationships now exist in the literature. Empirical results from spatial equilibrium models are used mainly to assess the potential impact on trade and prices of changes in government policies that affect production, consumption, and trade, including changes in quotas and tariffs, the introduction of embargoes or barriers to trade, etc. (e.g., Abbott and Paarlberg 1986). These models have also been widely used to examine the economic effects of climate change on agricultural markets (e.g., Adams et al. 1990).

Varying specifications of regional supplies and demand are possible, and models may involve a single or multiple products. A simple specification, for example, may involve a single product with its supply treated as homogenous and exogenously given. A possible generalization is to assume that importers differentiate between sources of supply (Armington 1969), and a more general analysis may require that the model takes account of interrelationships among commodities. Computable general equilibrium (CGE) models attempt to account for interrelationships in the economy but require substantial data and parameter inputs. This, in turn, raises issues about the quality of the inputs used in models of trade and space (e.g., Peterson, Hertel and Stout 1994; Neary 2001).

It is beyond the scope of this book to discuss this vast literature, but it is important, when one reads it, to identify and understand the assumptions and limitations of various models. For example, is it necessary for the analysis to differentiate between sources of supply? Products from different regions may not be identical, at least in the minds of buyers (e.g., Soshnin, Tomek, and de Gorter 1999). Are substitution relationships important? If so, then own- and cross-price elasticities of demand and supply are needed for every relevant product at each origin and destination.

Solutions to large spatial models can be obtained at modest cost because of developments in computer hardware and software, but this does not mean that the structure of prices derived from such models will necessarily correspond closely to observed prices. A model itself cannot tell why deviations occur. The observed discrepancies may be attributable to errors in data, the assumptions underlying the model, or errors in the details of specifying the model. International trade models may fail to reflect observed patterns of trade because countries seek to diversify their sources of supply to avoid the risk associated with relying on one supplier or elect to purchase from certain suppliers for political reasons. In general, solutions to international trade models suggest a higher level of specialization in trade between nations than is observed in practice.[5] Formal models also may fail to reflect the strategies that individual firms or countries adopt in an attempt to preserve market shares or to attract new buyers. For example, exporters may absorb freight costs, subsidize exports, or offer other financial incentives such as inexpensive credit, and it may be difficult to model these policies.

The foregoing discussion raises a number of issues in modeling and appraising regional price relationships. One set of issues is the specification of supply and demand relationships and the measurement of transfer costs. The determination of transfer costs is discussed briefly in the next section. Selected issues in modeling supply and demand relationships are covered in the final chapters of

5. Technically, the number of regions that trade with each other, based on the solution of a spatial equilibrium model, will be much smaller than the total number of possible trading relationships (Grennes, Johnson, and Thursby 1978: 25–26). This is because each deficit region will tend to purchase from only one or possibly two surplus regions, based purely on cost relationships. Japan, for example, would purchase all or at least most of its wheat from Australia because it is the least-cost solution given the model specification, but in practice, Japan also buys from other suppliers. The relationship between the total number of potential trades among regions and the number of trades occurring in the final solution of a spatial equilibrium model can be expressed mathematically. If there are $m$ supply and $n$ deficit regions, the number of potential trades equals $mn$, but the number of actual trades in the model will never exceed $m + n - 1$. Thus, for example, given 5 exporting nations and 20 importing nations, the number of trading combinations is $5 \times 20 = 100$; the number of trades appearing in the final solution will not exceed $5 + 20 - 1 = 24$.

the book. Another set of issues relates to the behavior of actual regional prices relative to the behavior expected under competitive market conditions. Do market prices conform to those expected under a competitive markets structure? That is, are markets efficient? We address this topic in the final section of this chapter.

## Determining Transfer Costs

Transfer costs, including trade barrier distortions in international trade, are the most important single factor in determining spatial price relationships. Simple models assume that transfer costs between regions are constant per unit of product, regardless of the volume shipped. This assumption can be relaxed, but doing so requires additional data that allow the research to examine volume-cost relationships.

Average transportation rates are not a good measure of actual transfer costs between any two points. The components of arbitrage costs include not only transportation rates but also loading and unloading charges; transactions costs such as entrepreneurial expertise and time, contracting, insurance, and financing; and possible fees associated with testing, grading, and/or meeting phytosanitary standards. Also, a variety of risks may be associated with moving goods over space, and these risks are a type of cost. Time lapses exist between the start of the transaction and the delivery of the commodity in the new location. Consequently, prices can vary, and there is a risk of contract non-performance. For international trade, exchange rate risk exists and needs to be factored into international transfer costs or ameliorated, say, by using foreign exchange futures contracts.

Clearly, obtaining a comprehensive, high-quality measure of transfer costs is difficult. Nevertheless, they can be viewed as having a fixed component that is independent of the distance traveled and a variable component that depends on the distance over which the commodity is moved. Transportation costs *per unit of distance* often decline as the distance traveled increases; thus, the cost of moving commodities between two points is not necessarily a linear function of distance. In most cases, however, the transfer costs *per unit of product* rises as distance increases but less than proportionately.[6] As noted earlier, it is the cost per unit of product that is relevant in modeling price relationships.

Inter-regional price differences are typically analyzed based on the least-cost method of moving commodities between points, since these markets tend to be highly competitive. But it may not be possible for every shipper to use the least-cost system all the time. Insufficient capacity may exist on the lowest-cost



6. The total transfer cost increases with distance, and consequently the cost per unit of product moved increases. Since the transfer cost contains a fixed component regardless of the volume and distance traveled, the cost per unit increases but at a decreasing rate.

transportation method at some times, and as a consequence, market prices necessarily reflect the costs of less efficient but available transportation methods. McNew (1996) points out that the 1993 flood on the Mississippi River disrupted barge transportation, which was the least-cost transportation method for grains from river locations to New Orleans. More recently, Hurricane Katrina shut down the Mississippi River near New Orleans for a period of time, and grain that would normally be shipped by barge was shipped by rail, which is more expensive for longer distances. Hence, price relationships for corn at different locations on the river changed temporarily.

In sum, it is difficult to obtain accurate estimates of the relevant costs of arbitrage between regions, including how these costs may vary with the passage of time. It certainly is insufficient to use published transportation rates. The question is, what are the overall costs of transfer faced by potential arbitragers? Unfortunately, this question is difficult to answer, and we turn to some issues related to appraising market efficiency.

## Observed versus Theoretical Price Differences: Are Spatial Markets Efficient?

If markets are competitive with no trade barriers, then at any point in time regional prices should differ only by transfer costs. Put differently, markets are considered inefficient if profitable arbitrage opportunities persist. That is, if regional prices are adjusted for transfer costs, then the law of one price should hold; the law of one price has been the basis for numerous tests of market efficiency and market integration (e.g., Ravallion 1986; Barrett and Li 2002; Negassa and Myers 2007).

Precisely because transfer costs are difficult to measure and vary with the passage of time, the results of many tests of market efficiency are questionable (Barrett 2001; McNew 1996). Also, since empirical analyses use time-series data, the assumption of continuous trade between regions is a key implicit assumption in many empirical models; however, as both Barrett and McNew point out, trade between regions is often discontinuous. Regional prices can have switching regimes; they do not have a consistent relationship with the passage of time.

Within a developed economy like the United States, markets are very likely to be integrated, and one should not be surprised that commodity prices are consistent with the expectation of relatively low prices in surplus production regions and higher prices in deficit production regions.

The complexity of spatial analysis, even in a country like the United States, is demonstrated by the data plotted in Figure 8-9. They are the daily bid prices for number 1 yellow soybeans for a 5-week period in St. Louis and Kansas City, Missouri, respectively. The prices pertain to beans of the same quality. St. Louis

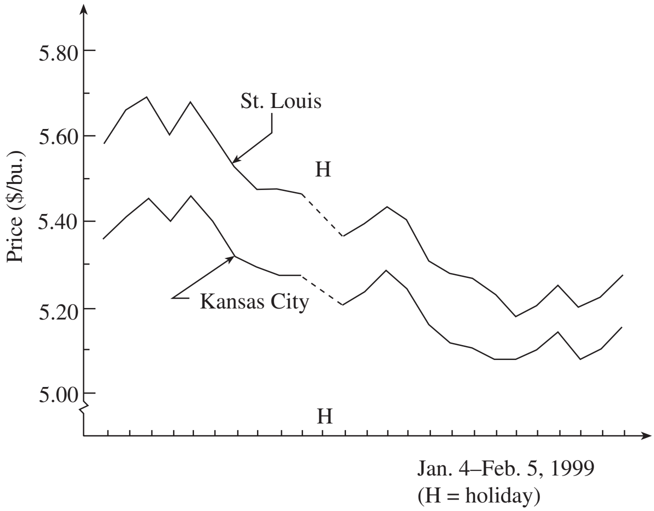



**Figure 8-9.** Number 1 yellow soybean prices, St. Louis and Kansas City, by day. *Source:* Data from USDA (1999)

and Kansas City are located in regions of surplus soybean production, and they are connected by river, rail, and truck transport. Given the high-quality information available to traders, these market locations are undoubtedly integrated; imperfections in pricing are minimal, and inspection of the figure suggests that the two prices are highly correlated. Both prices trend downward and tend to change in the same direction from day to day.

Nonetheless, a closer inspection of Figure 8-9 shows that the price differences between the two markets narrowed during this period—the price in St. Louis declined relative to the price in Kansas City. Thus, the price changes were not moving in a lock-step one-to-one relation, and a statistical analysis that assumed such a relationship might conclude that these two market places are not integrated. Based on the models illustrated in Figures 8-1 and 8-2, the observed behavior in soybeans prices is consistent with at least two possible explanations, which do not depend on market imperfections. The first explanation is that the costs of transfer were associated with the price level and were changing; as the price level declined, so did the transfer costs. As McNew (1996) has noted, the relationships of prices between integrated markets may not be linear. The possible second explanation is that the relative prices changed due to changes in regional demand and so shifted the areas supplying the two markets. These are not mutually exclusive explanations.

The principal point of the soybean example is to remind readers that the analysis of price behavior among locations cannot necessarily be based on simple linear models;[7] and if one has strong reason to believe that markets are integrated, this presumption should not be overthrown lightly. It is true, however, that markets are not always efficient. In some countries, information about prices may be inaccurate, incomplete, and/or not available quickly. Institutional or legal barriers may exist to the movement of commodities between regions. Inspection requirements, sanitary codes, tolerance limits for chemical residues, tariffs, import quotas, and licensing requirements are typical of the devices used to restrict inter-regional or international trade. In extremely poor countries, the impediments to market integration can also be due to extremely poor infrastructure for connecting markets and to lack of access to timely market information (Mutambatsere, Mabaya, and Christy, 2007). If such imperfections are thought to exist, model specifications should explicitly incorporate them rather relying on the indirect evidence provided only by price data.

Further, as noted previously, buyers may have a preference for a commodity produced in a particular region over the commodity produced in other regions; however, this is a reflection of differences in preferences, not an inefficiency. Products that appear to be homogenous may not, in fact, be substitutable in the minds of buyers.

## Concluding Remarks

In the absence of barriers to the free movement of commodities, competitive market prices in various regions respond to changes in supply and demand in different regions and to changes in transfer costs. A change in demand or supply in one region can have far-reaching effects on other regions, including those not directly involved in trade with that region. Changes in transfer costs, likewise, can alter the relative advantage of producers in different areas. In general, a decrease in transfer costs will benefit more-distant more than nearby producing areas (in relationship to a central marketplace). Thus, it is important to know something about the factors that influence spatial price relationships in attempting to explain changes in the competitive positions of different regions.

Prices allocate commodities not only spatially but temporally as well. The next chapter deals with temporal dimensions of prices, including the economics of storage. Markets determine both the quantity stored and where it is stored. Thus, there is a spatial aspect of carrying inventory. A discussion of the combined temporal and regional allocations of inventories is, however, beyond the

7. A linear statistical model may be written $P_{1t} = a + bP_{2t} + e_t$, where $P$ is the price; the subscripts 1 and 2 represent the regions; $t$ is time, where, say, $t = 1, 2, \ldots, T$ days; and $e$ is an equation error term. In this model, $b = 1$ and $a$ is a constant parameter, assuming an efficient, integrated market where transfer costs are constant over the sample period. The point of the discussion in the chapter is that transfer costs are unlikely to be constant and $b$ need not be 1, even in an efficient market. McNew (1996) fit a non-linear (cubic) functional form in relating the price of corn in New Orleans to prices in Memphis and St. Louis.

scope of this book (see Benirschka and Binkley 1995). We do note that conceptual models suggest that optimal storage will occur in producing areas and that time in storage varies directly with distance to market.

## References

Abbott, P. G., and P. L. Paarlberg. 1986. "Modeling the Impact of the 1980 Grain Embargo," in *Embargoes, Surplus Disposal and U.S. Agriculture*. USDA Ag. Econ. Rep. 564. Chapter 11.

Adams, R. M., C. Rosenzweig, R. M. Peart, J. T. Ritchie, B. A. McCarl, J. D. Glyer, R. B. Curry, J. W. Jones, K. J. Boote, and L. H. Allen, Jr. 1990. "Global Climate Change and U.S. Agriculture," *Nature* 345:219–224.

Armington, P. S. 1969. "A Theory of Demand for Products Distinguished by Place of Production," *IMF Staff Papers* 16:159–176.

Barrett, C. B. 2001. "Measuring Integration and Efficiency in International Agricultural Markets," *Rev. of Ag. Econ.* 23:19–32.

Barrett, C. B., and J. R. Li. 2002. "Distinguishing between Equilibrium and Integration in Spatial Price Analysis," *Am. J. Ag. Econ.* 84:292–307.

Benirschka, M., and J. K. Binkley. 1995. "Optimal Storage and Marketing over Space and Time," *Am. J. Ag. Econ.* 77:512–524.

Bressler, R. D., Jr., and R. A. King. 1970. *Markets, Prices, and Interregional Trade*. New York: John Wiley and Sons.

Grennes, T., P. R. Johnson, and M. Thursby. 1978. *The Economics of World Grain Trade*. New York: Praeger.

Houck, J. P. 1992. *Elements of Agricultural Trade Policies*. Prospect, IL: Waveland Press.

Judge, G. G., and T. D. Wallace. 1958. "Estimation of Spatial Price Equilibrium Models," *J. Farm Econ.* 40:801–820.

McNew, K. 1996. "Spatial Market Integration: Definition, Theory, and Evidence," *Ag. & Res. Econ. Rev.* 25:1–11.

Mutambatsere, E., E. Mabaya, and R. D. Christy. 2007. "Integration and Equilibrium in the Maize Markets in Southern Africa." Paper presented at American Agricultural Economics Association 2007 Annual Meeting, Portland, Oregon, July 29–August 1.

Neary, J. P. 2001. "Of Hype and Hyperbolas: Introducing the New Economic Geography," *J. Econ. Lit.* 39:536–561.

Negassa, A., and R. J. Myers. 2007. "Estimating Policy Effects on Spatial Market Efficiency: An Extension of the Parity Bounds Model," *Am. J. Ag. Econ.* 89:338–352.

Peterson, E. G., T. W. Hertel, and J. V. Stout. 1994. "A Critical Assessment of Supply-Demand Models of Agricultural Trade," *Am. J. Ag. Econ.* 76:709–721.

Ravallion, M. 1986. "Testing Market Integration," *Am. J. Ag. Econ.* 68:102–109.

Soshnin, A. L., W. G. Tomek, and H. de Gorter. 1999. *Elasticities of Demand for Imported Meats in Russia*. Dept. of Ag., Res., & Mgr. Econ. (now Dyson School of Applied Econ. & Mgt.) Working Paper 99-19. August.

Takayama, T., and G. G. Judge. 1971. *Spatial and Temporal Allocation Models*. Amsterdam: North Holland.

USDA. Various years. *Grain and Feed Weekly Summary and Statistics*. Ag. Mkting. Ser. Washington, DC.

CHAPTER 9

# Price Variation through Time

Agricultural product prices vary substantially more through time than the prices of most industrial products and have different time-series properties than do the prices of volatile financial assets that are also traded in competitive markets. For example, Hull writes, "[For IBM stock] . . . our predictions for the future should be unaffected by the price one week ago, one month ago, or one year ago. . . . the probability distribution of the price at any particular future time is not dependent on the particular path followed by the price in the past" (2009: 259f). In contrast, the current price for spot delivery of an agricultural commodity is typically related to previous prices. This chapter identifies regularities in the behavior of agricultural prices and provides a basis for understanding why they exist and persist.

The behavior of commodity prices is typically the result of a complex mixture of changes associated with seasonal, cyclical, trend, and random factors. Moreover, the existence of significant transactions costs can influence day-to-day price behavior. The most common regularity observed in an agricultural price series is a seasonal pattern. Some commodities also exhibit cyclical behavior; i.e., a tendency exists for production and prices to vary systematically (cyclically) over a period of years.

Logical reasons exist for these systematic patterns of behavior. Economists have developed conceptual and empirical models that are intended to explain commodity price behavior. This chapter deals primarily with conceptual models; the final chapters of the book discuss empirical analyses. We note now, however, that the ability to forecast prices using empirical analyses based on current information is not synonymous with being able to profit from these forecasts. In efficient competitive markets, systematic behavior is associated with unavoidable costs of arbitrage, such as the costs of carrying inventory.

In this chapter, we start with seasonal behavior and then continue with systematic components that might be observed over a period of years, such as trends and cycles. Finally, we return to the behavior of daily prices. Daily prices can be seen as reflecting, in part, the seasonal and cyclical components, but they also may have some additional features associated with the costs of adjusting to new information.

## Seasonal Variation in Prices

### Sources of Seasonality

Seasonal price behavior is a systematic pattern that occurs within a year. For many agricultural commodities, the main source of seasonality originates from the supply side. Many crops are produced annually and are storable. The market must allocate the available supply (annual production plus the inventory, if any, carried from the previous year) over the year to meet a continuous demand. Livestock and livestock products, like milk, also may have seasonality in production. This is related to seasonality in feed supplies and/or seasonality in the gestation-birth process (e.g., beef cows typically calve in the early spring). For a product like milk, production occurs daily but is seasonally large after calving in the spring. As discussed subsequently, seasonal price patterns can change for a variety of reasons, including changes in seasonal production patterns and in storage costs.

Seasonal variability in demand also exists for some agricultural products and is related to factors like holidays and the season of the year. Thus, in the United States, the demand for turkey meat is largest for the Thanksgiving (November) and Christmas (December) holidays. The demand for ice cream rises in hot weather and declines in cool weather. To meet seasonally large demands, the economics of production and storage sometimes results in relatively continuous production over the year, with storage (e.g., of frozen turkeys) occurring in periods of relatively low demand to meet subsequent peak-demand periods. The least-cost solution to meeting seasonal variation in demand may involve a consideration of both production and storage costs.

Thus, model specifications will differ by commodity, but in general, models of seasonal price behavior need to take account of current and expected future production, current and future demands, and inventory carrying. Inventory carrying, of course, involves costs. In this chapter, we discuss relatively simple models that assume annual production, with inventories either going to zero each year or with inventories being carried over from one year to the next. We will assume, for simplicity, that the demand functions are the same for each time period (say, month) within the marketing year. Seasonal price behavior arises from the seasonality of supply. An annual crop is harvested in period 0, and sellers have the option of selling at harvest or holding inventories for sale in subsequent periods. In the simplest competitive model, storage costs per unit of time are known and fixed, and the entire harvest is used each year. Hence, the ending inventories are zero, and each year, the quantity produced equals quantity demanded.

As elaborated later in the chapter, if a market is perfectly competitive, the differences in prices between two intra-year periods must equal the cost of storage for that intra-year period. Marginal revenue from storage must equal the marginal cost of storage. Thus, assuming the marginal cost is a constant per unit of time, the seasonal price rise is the same for each time period. Appendix I to this chapter provides an algebraic representation of this model. It is important to recognize that the model assumes perfect information about the crop size, current and future demands, storage costs, and length of the storage period. Of course, decision makers do not have perfect information in actual markets.

The late-summer, fall-harvested potato crop in the United States is an example of a commodity that is uneconomical to store from one crop year to the next. Production is used through the winter and early spring, and inventory goes to zero as the spring crop becomes available.

In sum, given knowledge of seasonal demands, available supply, and storage costs, the market determines the initial and subsequent prices that exactly allocate the fixed supply over time. This is an example of arbitrage over time. (In Chapter 8, we discussed arbitrage over space.) Higher prices are required in the successive time periods to induce inventory-holders to carry stocks. We elaborate on this point later (and in Appendix I to this chapter).

## Seasonal Patterns

Assuming an annually produced crop that is storable and a perfectly competitive market, prices will be lowest at harvest time and rise exactly by the marginal cost of storage per unit of time. This is illustrated in Figure 9-1. In the top panel, no inventory is carried over from one year to the next, and there is a discontinuity in prices between years. The lower panel assumes continuous storage, with prices rising over most of the marketing year but then declining rather quickly in anticipation of the new harvest.

The driving force in price behavior throughout the year is inter-temporal arbitrage by merchants and traders. The equilibrium condition is that, for any fixed time interval of storage, the expected future price minus the current price must exactly equal the marginal cost of storage for that time interval. The expected price minus the current price is the marginal return to storage. If the expected price exceeded the current price by more than the cost of storage, an incentive would exist to reduce current consumption and store a larger amount for future use. This would have the effect of raising the current price and reducing the expected price, and this process would continue until the equilibrium condition was met. This is referred as the law of one price in a temporal context.

A mathematical notation is $E[P_{t+i} \mid I_t] - P_t = m_t$, where $E$ is the expectation operator, $P$ is the price, the subscripts denote time periods, $I$ is the information-conditioning price, and $m_t$ is the marginal cost of storage over the interval $t$ to $t + i$. In reality, the expected price is unknown. (Also, $m_t$ may not be fully known at time $t$, but we will mostly ignore this issue for clarity of exposition.)

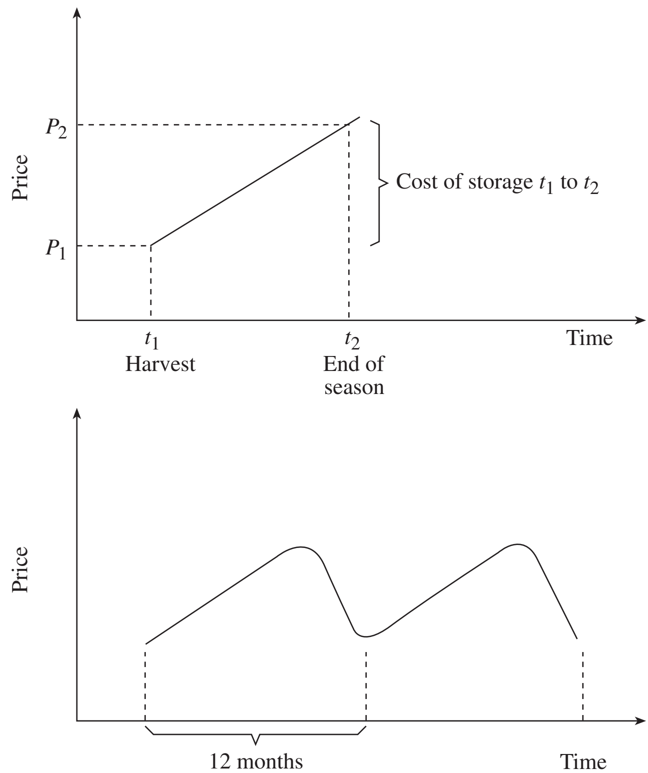



**Figure 9-1.** Theoretical seasonal price behavior

The expectation notation for price anticipates the need to define how expectations are formed and emphasizes that expectations can change as new information arrives in the market. Alternative hypotheses about how expectations are formed are discussed in the following section on cycles, and in Chapter 12, we point out that, when a market exists for contracts for future delivery of the commodity, the current price quote for a futures contract is registering the market's current expectation about a particular future (expected) price.

Given the assumption of a perfectly competitive market with known supply and demand conditions each season, the seasonal price pattern is depicted in the lower panel of Figure 9-1, as a smooth, repeating seasonal pattern. The storage costs implicit in this diagram can be divided into four components. The first category is the conventional costs of inputs, such as return on investment in the warehouse space, wage rates, insurance premiums, etc. The second category is the opportunity cost associated with the present value of the inventory, which in turn depends on the price of the commodity and interest rates. While costs of inputs like insurance premiums are relatively stable from year to year, opportunity costs based on the commodity's price and interest rates are not.

The third category of costs is a type of benefit, namely the convenience yield of holding stocks. This helps explain the "backwardation" observable in the lower panel of Figure 9-1. That is, as the new harvest approaches, the expected price for the new crop is below the current price (the expected return to storage is seemingly negative); nonetheless, some stocks are carried from one year to the next. Presumably this occurs because merchants and processors obtain a benefit from avoiding stock-outs, and this benefit provides an incentive to carry some inventories into the new marketing year notwithstanding the otherwise negative return to holding stocks. We will have more to say about this in Chapter 12.

The fourth category of costs, from the perspective of risk-averse inventory holders, is related to uncertainty associated with the expected future price. In a perfect world, $E[P_{t+i}]$ is known, but in the real world, it is not known with certainty. The errors in estimating returns to storage create price risk; i.e., the realized change in prices over the storage interval may not cover storage costs. This price risk is a type of cost. (This argument is analogous to the one made for producers in Chapter 4.)

In practice, the "normal" seasonal price pattern does not prevail each year. Prices may rise by more than the cost of storage, or prices may even decline over the season. Those storing the commodity necessarily act on information that is subject to change. At harvest, they have estimates of production, inventories, and expected demands, but as time passes, this information changes. The initial estimate of supply may turn out to be smaller or larger than the actual supply. Anticipated demands may not materialize, or unanticipated demands may arise.

The timing of the harvest of the new crop also influences seasonal price patterns, and end-of-season supplies of semi-perishable commodities, such as cabbage, potatoes, and onions, are often uncertain, partly because of variability in storage quality. The market must attempt to match the supply of, say, potatoes in storage with the timing of the new-crop potatoes. This means allocating the fall-harvested potatoes through the winter while trying to anticipate the size and timing of the harvest of new-crop potatoes in the spring. If supplies toward the end of the season are short, prices rise dramatically; if supplies are large, season-end prices decline.

The foregoing points are illustrated by the monthly prices of onions produced in New York State (Figure 9-2). For an 8-year period, 2002–03 to 2009–10, prices rose on average \$6.06 per cwt. from the monthly low to high. In 2008–09, however, prices declined \$9.10 per cwt. from September to May. In contrast, prices rose from \$12.70 to \$38.60 per cwt. in 2006–07 (not shown). These large changes, of course, have a large influence on the 8-year average. Eight years is a short sample, but it does illustrate the potential for large variability in seasonal

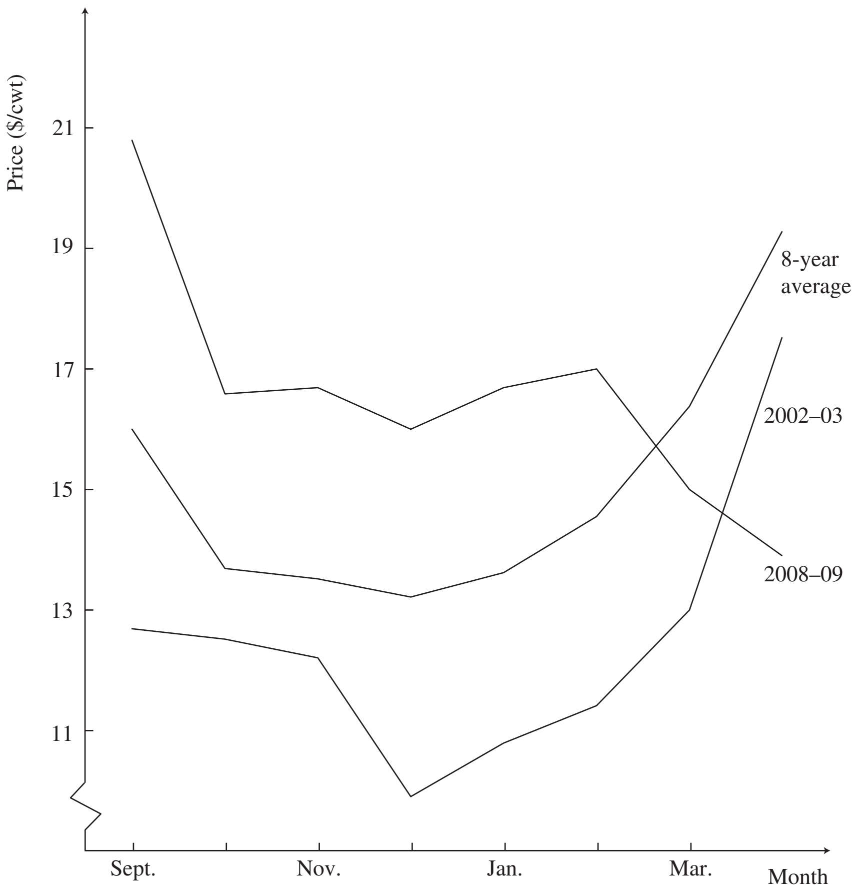



**Figure 9-2.** Fresh market onions: monthly average prices received by farmers in New York State for two crop years and an eight-season (2002–03 to 2009–10) average

behavior and hence in returns to storage. When a futures market exists for the commodity (none exists for onions), hedging helps assure a positive return to storage, or alternatively, in some years, futures prices signal that hedged storage is not profitable (see Chapter 13).

Although intra-year price changes deviate from the smooth pattern depicted in Figure 9-1, that diagram serves to emphasize the logic behind the seasonal component of prices. Figure 9-2 helps illustrate the reality of seasonal price behavior. But both figures imply that prices in adjacent days, weeks, and months are associated with each other. The prices are, using a statistical term, autocorrelated.

Prices also become more variable as the storage season progresses; i.e., the variance of prices immediately after harvest is smaller than the variance of prices in months later in the marketing year. Within a marketing year, inventories are declining, and when inventories are small late in the season, changes in expectations can have a large price effects. In particular, unanticipated increases in demand combined with a small inventory can cause spikes in prices, and consequently, the distribution of prices is likely to become increasingly skewed to the right as the marketing season progresses. Thus, if one visualizes probability distributions of prices by months for annually produced, storable crops, the shapes (parameters) of these distributions change as the months after harvest pass. The mode gets larger, as does the variance, and the distributions can become increasingly skewed (Peterson and Tomek 2005: 297).

### Systematic Changes in Seasonal Patterns

A particular price pattern, as just discussed, depends on the economic conditions of that year; price differences between months depend on these conditions. It follows that changing fundamental conditions from one year to the next should change the seasonal pattern. If the costs of storage increase, then the seasonal increase in price will become larger.

Seasonality in prices will also shift with changes in the pattern of production. For example, the number of sows farrowing (female hogs giving birth) used to be seasonally large in the spring; hence, the marketing of pork was seasonally large in the fall. With changing production technology, the seasonal pattern of output has lessened over time, and consequently, one should expect a decline in the amplitude of the seasonal prices of hogs. The soybean market is somewhat analogous in that soybean output has increased in the southern hemisphere, and thus the world market now experiences two major harvests per year. The upward trend in southern hemisphere output influences the seasonal price pattern of northern hemisphere soybeans, partly by moving the timing (phase) of peak prices to an earlier date.

Put in more technical terms, a seasonal pattern has a fixed period of 12 months, and with fixed seasonal conditions, the amplitude of the prices would be constant from year to year. But, in reality, reasons exist for the magnitude and date of the peak of seasonal patterns to shift. *Ceteris paribus*, the amplitude could increase or decrease as storage costs rise or fall, and amplitudes can flatten out as the production of a commodity like pork has reduced seasonality. This potential for systematic changes plus the usual irregularities in individual years make the use of historical seasonal price patterns for forecasting difficult. Although good reasons exist for systematic seasonal behavior in commodity prices, this does not mean that they are easy to forecast or that the forecasts can be used for profitable arbitrage.

## Annual Price Behavior

Year-to-year variability in commodity prices can be explained using models of price determination under competition (Chapter 5). The demand and supply functions may be viewed as representing annual averages with annual price changes arising from shifts in these functions.

For agricultural products, a major factor in year-to-year price variability is changes in supply. In the United States, the supply of most commodities available in any year is based mainly on current production and to a lesser degree on carryover from the previous year and on imports. Annual fluctuations in the supply of farm commodities, as we have seen, are sensitive to many economic and non-economic factors. Demand also may change because of fluctuations in various demand shifters, including those influencing international trade. Thus, in the simplest terms, farm prices are the outcome of shifts in (derived) demand relative to supply.

Crops have swings in annual production because planned production varies with changing expectations about returns and because yields are sensitive to weather conditions, diseases, and pests. Since the derived demands for many crops at the farm level are price inelastic, year-to-year shifts in production result in price fluctuations that are larger in percentage terms than the change in production. Inventories and imports may help ameliorate the effects of a small crop in a particular location, but occasionally a small crop and short inventories worldwide combine to result in exceptionally high prices, i.e., large spikes in prices.

A good illustration of annual price variability is the fresh onion market at the farm level in New York State, which is depicted in Figure 9-3. Inventories are not carried from one crop year to the next for these late-summer harvested onions, and hence prices are dependent mainly on current production relative to current demand. Over the marketing years, 2000–01 to 2011–12, the season average price varied from $9.70 to $19.70 per cwt. (New York Agricultural Statistics Service 2012). The mean price over this period was $14.78, with a standard deviation of $3.32 per cwt. The standard deviation was 22 percent of the mean. Clearly onion prices are highly variable.

Animal and animal products output are also influenced by weather conditions and diseases, but usually their effects on supply are less dramatic than for crops. But both crop and livestock products can be affected by demand shocks, such as unexpected contamination of the product by a pathogen. The prices of dairy, livestock, and livestock products are also quite variable. For example, the average annual all-milk price for 2000–12 was $15.80 per cwt., with a standard deviation of $2.99, or 19 percent of its mean value.

In contrast to year-to-year changes, there are occasions when prices shift to, and remain at, a new level. In this case, annual variability is around a new level. One such occasion was the entry of the then-existing Soviet Union into the world market for grains in 1973. This created a new source of demand that persisted, and nominal prices shifted to a substantially higher level. More recently, the increased demand for crops (especially corn) for use in manufacturing ethanol, combined with a larger demand for grain and oil seed imports by China and India, shifted the prices of grains and oil seeds to a new higher level

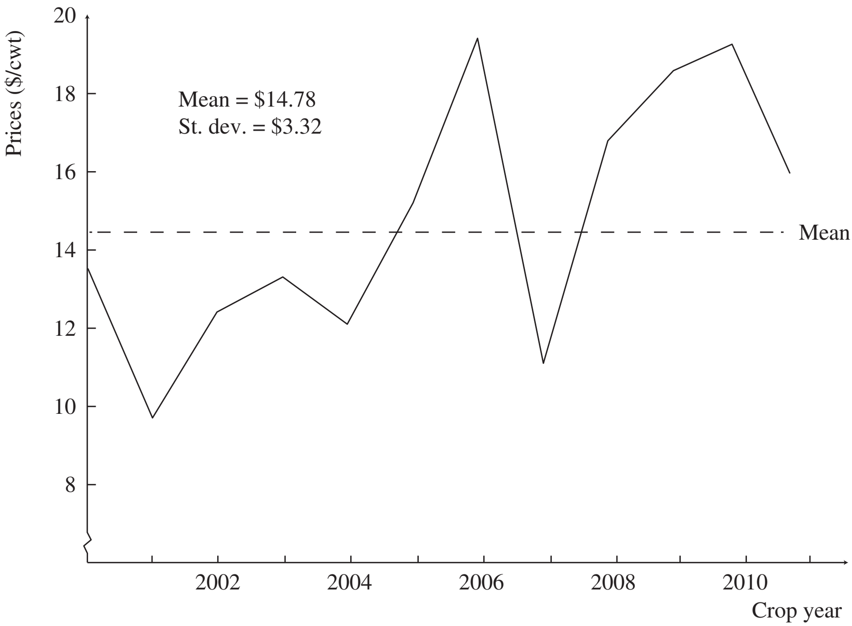



**Figure 9-3.** Fresh market onions, season-average price received by farmers in New York State, 2000–01 to 20011–12 marketing years

in 2006 that has persisted well beyond 1 year. Higher prices for grains and oil seeds also translate to higher prices for other commodities, especially livestock, that consume the higher-priced grains.

The behavior of wheat prices in the United States for the marketing years 1960–61 to 2011–12 is illustrated in Figure 9-4. Three price levels are apparent in this sample period. The average price in 1960–61 to 1972–73 was $1.55 per bushel, and in 1973–74 it more than doubled to $3.30. This level persisted through 2005–06, when the average wheat price increased to $5.89 (2006–07 to 2011–12) per bushel. The mean prices in the early period and the recent period are significantly different (by a statistical test) than the mean in the middle period. Indeed, statistical analysis found that 80 percent of the price variability was associated with the two shifts in the mean.

Figure 9-4 also suggests that, as the mean price increased, so did price variability; thus, although prices since 2005–06 are significantly above those in previous years, the price in any individual year can be well below (or above) $5.89. It is important, therefore, to distinguish between year-to-year variability around a constant mean and variability associated with a shift in the mean.

Shifts in price levels (and increased variability) should be viewed as structural changes, i.e., changes in the true parameters (the mean, etc.) of the underlying probability distributions. Although it is difficult to ascertain whether a change

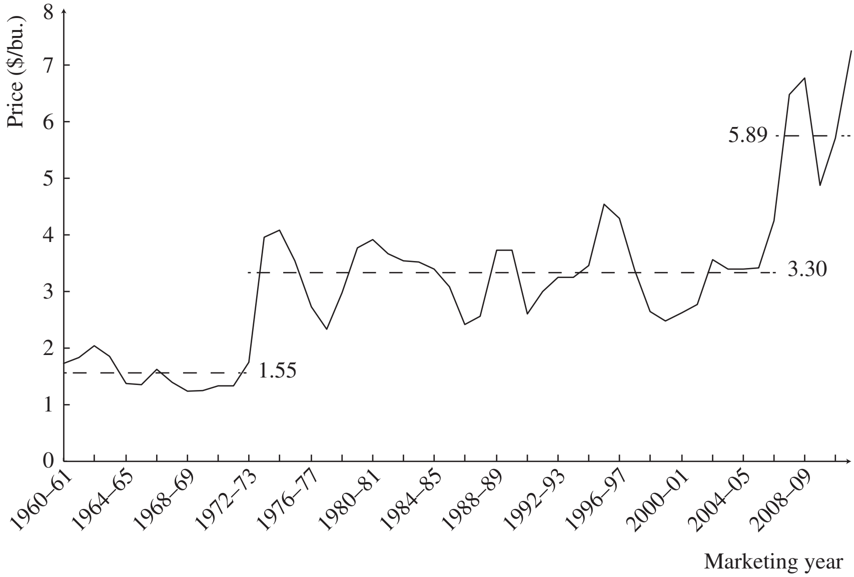



**Figure 9-4.** Farm price of wheat, United States, marketing years 1960–61 to 2011–12

in a price level is transitory or persistent, shifts in price levels that persist through time should be distinguished from continuous trends or from 1-year price changes (for an analysis using advanced econometric tools, see Enders and Holt 2012). In contrast to one-time shifts, trends in prices are incremental changes in the mean each year; we discuss this in a subsequent section.

## Cyclical Behavior

A cycle is a price pattern that repeats itself over a time period that is longer than 1 year. A true cycle is self-energizing and not the result of chance factors. The simplest price cycle has a fixed period (length of time from peak to peak, or from low to low) and a fixed amplitude (net of the other factors influencing prices). Cyclical or cycle-like price behavior for agricultural commodities is more complex than this. Cycle-like behavior can be initiated by an external event. For example, a drought reduces supply and raises prices.[1] The higher price leads producers to increase production in the next year, which results in a lower price. This, in turn, reduces production and raises prices. And so on. Such a pattern implies that the annual behavior, discussed in the previous section, could have a systematic component. Such a cycle, however, is likely to die out—dampen—unless started again by another external event.

1. Such an external event could also be on the demand side, e.g., the U.S. government enacting ethanol subsidies, which significantly increased the demand for corn.

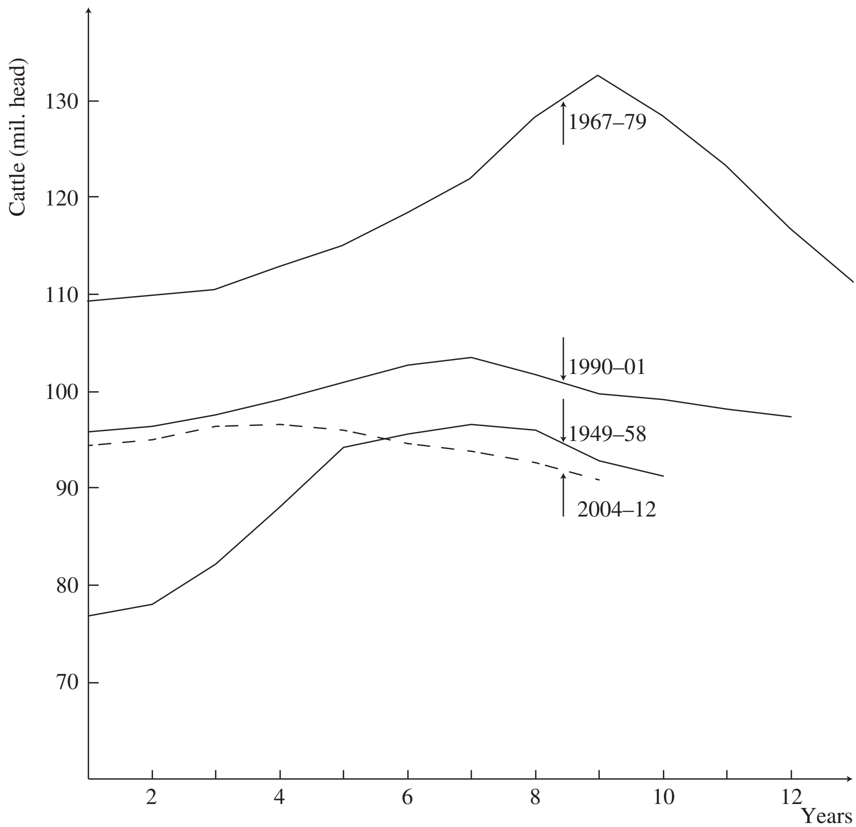



**Figure 9-5.** Cattle and calves on farms, United States, January 1 in selected years, 1949–2012

If economic cycles are broadly defined to include systematic, relatively repetitive behavior over a period of years, for whatever reason, then cyclical behavior does appear to exist for some agricultural commodities, especially as measured by the number of cattle or of hogs on farms. But the lengths and amplitudes of livestock production cycles are by no means uniform. Producers have some flexibility in how fast they expand or contract production, although (as discussed in Chapter 4) the expansion phase of animal production is constrained at any point in time by the size of breeding herd or flock. Thus, the expansion phase will take longer than a contraction phase.

The total number of cattle and calves on farms in the United States on January 1 is plotted in Figure 9-5 for four of the eight cycles occurring within the period 1928–2012.[2] Measuring from low to low, the cycles range from 9 to 12 years

2. All six cycles through 1989 are shown in the third edition of this book; in this edition, we omit four for clarity of presentation.

in length. Cattle numbers peaked in 1975 at about 132 million head. Thus, the plots from 1928 through 1975 reflect both an upward trend and cycles. Subsequently, cattle numbers trended downward.

In Figure 9-5, we have not attempted to disentangle the trend from the cycle, but it is clear that, for the last period, a downward trend in cattle numbers has dominated the cyclical effect. It is likely that the decline in numbers would have been more pronounced without the upward pressure of the cyclical effect. Net of trend, the number of cattle typically has grown for 6 to 8 years and then declined for 3 or 4 years. In sum, systematic cycle-like behavior is occurring, but the various periods are not identical.

Cycles in prices are typically not observable from a simple data plot. Nonetheless, there is much empirical evidence that the annual prices of agricultural commodities are autocorrelated (i.e., the price this year is correlated with prices in previous years). Deaton and Laroque (1992) found evidence of autocorrelations in the annual prices for many commodities, and Mundlak and Huang (1996) provided evidence of autocorrelations in herd numbers and prices for beef cattle for four countries. But, since the empirical decomposition of a time series into random and systematic components is difficult, the debate about the nature of price cycles is likely to continue.[3]

The length (period) of a cycle for an agricultural commodity is presumably related to the time required to produce a new crop or generation of animals. The cycle in hog numbers, however, is approximately 4 years long, although a new generation of hogs can be produced in less than 12 months (implying a 24-month cycle). Thus, it is not surprising that the exact mechanism (or mechanisms) that is creating cycles is unclear. Nonetheless, two factors are thought to create cyclical behavior in commodity prices: the way expectations are formed and the costs associated with responding to changed expectations.

The usual conceptual model, underlying cyclical behavior in prices of crops and livestock products, is based on hypotheses about the way price expectations are formed. Recall (Chapter 4) that the quantity of a commodity supplied in time $t$ is determined by expectations formed in a prior time period. A notation for a supply equation is $A_t = A(P^*_{t-1})$, where $A_t$ is the quantity supplied in time $t$ and $P^*_{t-1} = E[P_t|I_{t-1}]$ is expected price in time $t - 1$. This simply says that the current quantity supplied is based on expectations formed in a prior time period. The lag arises because of the biological and reproductive nature of the production process, and of course, the length of the lag between a decision about output and the realization of the output varies from one commodity to another. The one-period lag specified in the notation is merely a convenience.

3. Holt and Craig (2006) provide evidence for systematic cyclical price behavior using advanced econometric tools. We also note that a question in measuring systematic behavior in prices is, have the nominal prices been deflated by a general index of prices? The deflated series may have systematic behavior that does not exist in the nominal prices.

For animal products with long time lags in the production process, like beef, the reader should also recall the identity introduced in Chapter 4, in which current births of calves is constrained by the size of the cow herd. Given the current stock of mature and young animals at a point in time, the economic problem is to allocate this stock between current consumption and future use. The rate of culling of mature animals in the herd can be varied, and likewise, the allocation of young female animals between the breeding herd and slaughter can be altered. The decision about altering these rates (of slaughter, additions, etc.) depends on the current price relative to the (discounted) expected price in the future while also taking account of the expected costs of maintaining the herd (Rosen 1987).

A key variable in this model is expected price, and four hypotheses exist in the economics literature about how expectations are formed. The simplest hypothesis is *naive expectations*, namely $P^*_{t-1} = P_{t-1}$. The only information ($I$) used to form expectations in period $t - 1$ is the observed price. This is the basis for the simple cobweb model, which is discussed more fully in a subsequent section. The outcome of this assumption in a cobweb model is prices that alternate up and down from year to year; i.e., for a crop, the annual prices will have a negative autocorrelation (see Appendix II to this chapter).

Two of the hypotheses about the formation of expectations make the expected price dependent on current and past prices (i.e., if expectations are formed at time t – 1, then the relevant information set is prices in $t - 1$, $t - 2$, $t - 3$, etc.). The *adaptive expectations* model hypothesizes that expected price is a geometric-weighted average of current and past prices; the geometric weight can be estimated statistically (Nerlove 1956). In the *quasi-rational expectations* model, the expected price is hypothesized to equal the forecast from an autoregressive model. Namely, the forecast of the price expected in time $t$ is based on a statistical model that makes this price a function of previous prices (Feige and Pierce 1976); the number of lags is determined empirically.

For the *rational expectations* model, the information set determining the expected price is hypothesized to be the current information on all the variables thought to determine prices in the forthcoming period (Muth 1961). In other words, the decision maker is assumed to have expectations about the levels of supply and demand, and hence about the price that will prevail, at the time output is realized. The expected price is equivalent to a forecast from a structural econometric model that includes all the variables thought to explain supply and demand, and the forecast will necessarily be based on information about these variables that is known at time $t - 1$, when the forecast is made.

The *futures price* model is applicable to commodities that have a futures market (Gardner 1976). In this case, the current quote of the price of a contract for future delivery is hypothesized to be the rationally expected price. For example, the price of December corn futures on April 15 can be viewed as a forecast of the price of corn at contract maturity in December. This assumes that the futures market efficiently aggregates all the information that influences prices; i.e., the futures quote is as good as, or better than, other forecasts of the subsequently realized price.

In practice, we are uncertain about how farmers and other decision makers form expectations. Indeed, firm managers may form expectations in different ways. However, the important point is that, since time lags exist in the production process and since expectations must be based on existing information, it is possible that cycles are introduced into price behavior. These cycles are likely to be more observable for livestock and livestock products that require relatively long time periods for supplies to change. An increase in the supply of beef first requires an expansion of the basic breeding herd. Young female animals must be withheld from slaughter and reach sexual maturity, and after breeding, a gestation period exists. Then, after birth, calves must grow to slaughter weight. Similar arguments can be made for tree crops.

Since information about demand and supply changes with the passage of time, actual price outcomes will differ from those expected. Production decisions involve risk, and as noted previously, the riskiness of returns is a type of cost (for risk-averse producers), and one would not expect cycles to have exactly constant lengths and amplitudes. Furthermore, the empirical identification of cycles is complicated by the existence of other systematic components, such as seasonal and trend effects, as well as by random factors. Consequently, it is difficult to isolate the cyclical component and use this information for forecasting. Analogous to seasonality, prices can have cyclical behavior that is difficult to forecast precisely, especially many years into the future.

The costs of arbitrage over time include those associated with maintaining inputs, like a breeding herd, and the risks associated with errors in forecasting returns. (These costs can be viewed as analogous to the costs of storage.) Consequently, cycle-like, persistent price behavior is not arbitraged away.

## Cobweb Model[4]

### Description

Formal models have been developed to help explain cyclical behavior in quantities and prices. The simplest of these is the cobweb model. Like other simple models of price determination, it is not fully realistic, but because it is an important point of departure for building more realistic models, we provide here an in-depth discussion. An important aspect of the cobweb model is its recursive nature. In recursive models, prices and quantities are determined sequentially; i.e., they are linked in a causal chain. This is in contrast to many of the models of price determination (see Chapter 5) where prices and quantities

4. Ezekiel (1938) wrote one of the basic papers on the cobweb model. Subsequently, Waugh (1964) provided an updated discussion.

are assumed to be determined simultaneously. Because of the time lags in the production process for agricultural commodities, recursive models are sometimes the relevant foundation for modeling.

As already suggested, a high price can lead to large production; the large supply results in low prices, which in turn result in smaller production, and so forth. More explicitly, the cobweb model arises from three factors that, if present, will result in cyclical behavior of price and quantity. First, a time lag must exist between the decision to produce and the actual realization of output. Second, producers must base production plans on current (and perhaps past) prices. Hence, realized production, because of the time lag, is a function of one or more past prices. Third, current price must be a function of this predetermined production and the current level of demand.[5]

Consequently, the basic model can be sketched as follows. The current quantity supplied is a function of expected prices, and producers use naive expectations.

$$A_t = A(P^*_{t-1}), \text{ and}$$

$$P^*_{t-1} = P_{t-1}.$$

$$\text{Thus, } A_t = A(P_{t-1}).$$

Quantity ($A$) supplied in time $t$ is a function of the observed price ($P$) in $t-1$. The quantity supplied in $t$ is equal to the quantity consumed ($Q$) in $t$; $A_t = Q_t$. Thus, the market clearing price is determined by the inverse demand function:

$$P_t = D^{-1}(Q_t).$$

The causal chain follows. Given an initial price, $P_0$, quantity $A_1$ is supplied and consumed, thereby determining the price $P_1$. This price in turn determines $A_2$, and so forth.

Notice, if the supply equation is substituted into the inverse demand equation, we obtain

$$P_t = f(P_{t-1}).$$

If the cobweb model is a good depiction of reality, then the current price must be correlated with the price in the previous time period. Given that a negative relationship exists between price and quantity in the demand function and a positive relationship in the supply function, then logically the two prices should be negatively correlated.

5. The model can be modified to allow for the effects of carrying inventory. For simplicity, we ignore the effects of storage.

Suppose farmers based their supply decisions at time $t - 1$ on both $P_{t-1}$ and $P_{t-2}$; then $A_t = A(P_{t-1}, P_{t-2})$, and this implies a price function $P_t = f(P_{t-1}, P_{t-2})$. The way price expectations are formed can influence the autocorrelations in prices. In the simple cobweb model, current price is a function only of the past price with a one-period lag, but prices with lags of two or more periods are plausible. We also note that the existence of price risks, based on expectations formed in $t - 1$, can create an additional source of autocorrelations in prices; risk is a function of past errors in expectations.

Two supply relations are implicit in the cobweb model: short-run and very short-run supplies. The conventional short-run supply function makes the quantity supplied a function of expected price, and consequently, current production is a function of past prices. Once production is realized, the model assumes that this quantity must be sold. The market price adjusts to the available supply. In other words, price is determined as illustrated in Figure 5-1.

The name *cobweb* arises from the pattern formed by joining the successive price-quantity observations on a conventional supply-demand diagram (Figure 9-6). For convenience, we assume that poor weather resulted in a small crop and hence a relatively high price ($P_0$) at time period 0. The static short-run supply curve for "normal" weather, however, is shown as *S*. Hence, on the basis of $P_0$, producers plan to produce $A_1 = Q_1$, which is realized in time 1. Once

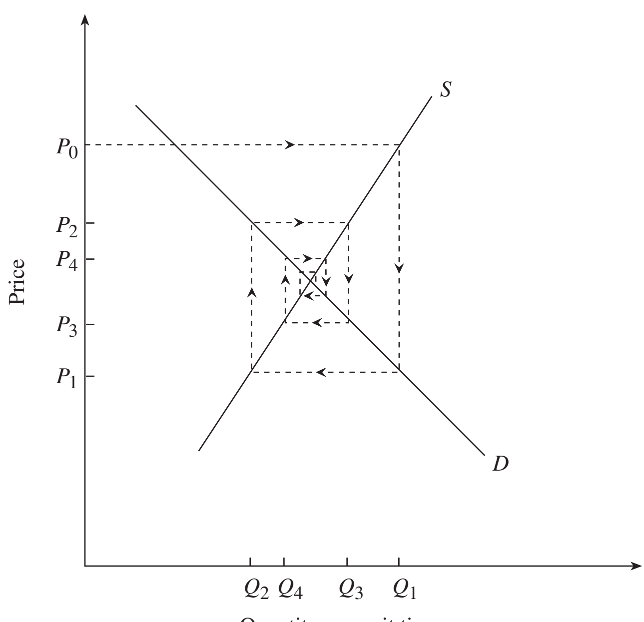



**Figure 9-6.** A cobweb model with a convergent cycle

produced, the quantity $Q_1$ is sold at time 1, and the market clearing price $P_1$ is determined by market demand relation $D$. Then, $P_1$ is the basis for the quantity supplied at time 2, $A_2 = Q_2$. As the process continues, the cobweb pattern develops.

The assumptions of the model are as follows.

1. Price is determined in a competitive market structure; producers are price takers.
2. An observable lag of at least one time period is required for a production response to a price change, and production plans are based on the current price; i.e., producers use naive expectations.
3. Once output is realized, prices adjust to clear the market of this quantity. Thus, price is determined by the shifts in very short-run supply.
4. A cycle depends on actual production equaling planned production. More generally, for a clear cobweb pattern to occur, demand and supply relations must be static.

The cobweb model may lead, in principle, to divergent, convergent, or constant-amplitude cycles in price and quantity. But, it is believed that the only realistic alternative is convergence. Prices are not expected to diverge (i.e., to "explode") through time, and a constant-amplitude cycle requires special mathematical conditions. For linear functions, the absolute values of the slopes must be equal for the cycle to be continuous (for non-linear functions, see Waugh 1964). A convergent cycle is illustrated in Figure 9-6, and Appendix II to this chapter provides an algebraic treatment.

A 1-unit time lag in the production process is assumed in much of the foregoing discussion. This lag, combined with naive expectations, results in a cycle that is 2 units in length. If the lag had covered two time periods, then the cycle would have been 4 units in length. Specifically, if 12 months elapse between harvests, then the cobweb model implies a cycle 24 months long. If, however, farmers used adaptive or quasi-rational expectations, the cycle could be longer.

## Limitations and Modifications

Basically, the limitation of the cobweb model described in the preceding section is that it is too elementary; it needs to be modified to make it more realistic. The assumption of naive expectations may be wrong,[6] and of course supply depends on more factors than expected price. Likewise, the derived

6. As explained in Chapter 12, the spot price for grains is highly correlated with prices of contracts for future delivery. Thus, naive expectations (current cash price) may differ little from rational expectations (the futures quote). Nonetheless, if farmers are thought to form expectations rationally, the futures price series is probably the preferred measure of expectations. In livestock markets, spot prices have a smaller correlation (than the grains) with prices for contracts for delivery at distant points in time, and one would expect the choice of alternative measures of expectations to make a difference in empirical results.

demand function must account for the complex set of factors influencing the level of demand. Actual models of commodity price behavior can, in principle, take account of these criticisms. Additional equations and explanatory variables can be added to the model, and alternative measures of price expectations and risk can be used.

Building "correct" models is not easy, however, and many specifications exist in the literature. For annual crops, for example, modeling the carrying of inventories helps explain autocorrelation in prices both from month to month and year to year (Peterson and Tomek 2005). Likewise, models of livestock markets have attempted to account for the systematic behavior of supplies and prices with varying levels of success (e.g., see Sarmiento and Allen 2000 and the references therein).

A related criticism is the seeming inconsistency between the persistence of cycles (at least for cattle and hog numbers) and empirical econometric models, which imply that the cycles should dampen. That is, if econometric models are fitted to historical data and then used to simulate future prices, these simulated prices dampen. Indeed, as noted earlier, economists believe that prices and quantities should dampen; explosive cycles are illogical.

If cycles are indeed regular and, hence, have a predictable component, then rational farmers could profit by taking counter-cyclical actions (arbitrage), and such actions would moderate or eliminate the cycles in prices and production. But, as noted earlier, the most likely explanation for the persistence of cycles relates to how expectations are formed and the costs of inter-temporal arbitrage. In particular, any method of forming expectations is subject to errors that can be large, especially for commodities that have long lags between the decisions to produce and the realization of the output. Forecasts of prices beyond a few months into the future are very imprecise. The paradox is, then, that historical prices and quantities can have measurable systematic behavior but that this information is insufficiently precise to provide profitable arbitrage in the future.

In sum, the cobweb model depicted here is too elementary for application to real-world situations. More complex models are necessary for useful applied analyses, but simplicity has the pedagogical virtue of isolating key features of dynamic behavior. Clearly, features of commodity markets, such as the lags between the decisions to produce and the realization of output, inventory adjustments, constraints inherent in breeding herds and so forth, can produce systematic patterns of price behavior. Because of the complexities of the underlying generating mechanisms of "cycles," however, forecasts of prices are not sufficiently precise to be used to make profitable decisions that would eliminate cycles.

## Trends

This section is about persistent and systematic upward (or downward) movements in economic variables. These are sometimes called *deterministic* trends, and an example of such a trend is $Y_t = a + bt$, where $Y$ is an observable variable and $t$ represents time (where, say $t = 1, 2, \ldots, T$). If the coefficient $b$ is positive, then $Y_t$ is trending upward by $b$ amount each time period. Typically, an error term is appended to the equation, $Y_t = a + bt + e_t$, where the mean of $e_t$ is zero. In this case, the equation implies that random variability exists around the systematic shift in the mean of $Y_t$.

We are not directly concerned here with statistical estimates of trends but, rather, are addressing whether or not prices of individual agricultural products are trending, and if so, why. Clearly, many economic variables have trends. For example, the gross domestic product (GDP) of a country like the United States has increased over time. The rate of growth varies, indeed is sometimes negative, but the general direction of change has been positive. Indexes of aggregate price behavior also tend to trend upward, as discussed in Chapter 10; inflation is defined in terms of positive growth rates in indexes that measure the aggregate behavior of the many prices in an economy.

The foregoing need not imply, however, that the nominal prices of individual commodities are trending. For an observable trend in a commodity price series, continuous relative shifts in supply and demand must occur. It is possible, for example, that a steady stream of new production technology causes the supply function to steadily shift to the right. If supply increases faster than demand, then the commodity's price will trend downward.

A simple model of farm-level prices illustrates the point. The supply equation is $A_t = A(P_{t-1}\ X_t)$, where $X$ is a variable that shifts the level of the supply equation (e.g., technology) and the other variables are as defined before. Let supply equal demand, $A_t = Q_t$, and price be determined by the inverse demand function $P_t = D^{-1}(Q_t\ Y_t)$, where $Y$ is a variable that shifts the level of demand. A trend in $P_t$ can occur if, say, $X_t$ grows consistently relative to $Y_t$.

As noted in a previous section, the *number* of cattle and calves on farms in the United States trended upward until 1975. A plausible explanation is that the demand for beef was growing steadily as the population and per capita income of U.S. residents grew, and so cattle numbers expanded to meet this demand. Starting in the 1950s, however, chicken became increasingly important as a substitute for beef, and also by the 1970s, concerns were growing about the health effects of "excess" red meat consumption. Ultimately, the demand for beef started to decrease (as shown in Figure 2-2), and the upward trend in cattle numbers reversed.

The foregoing discussion, however, relates to the number of cattle, not the nominal prices of cattle. Although cattle prices had to be high enough, on average, to encourage the expansion of cattle numbers from the 1930s through the 1970s, the increases in supply were roughly related to the increases in demand, and trends in prices of cattle are much less obvious than the trend in output.

If one examines plots of individual commodity prices in the United States, they tend to be trend stationary over considerable lengths of time (e.g., Wang and Tomek 2007). That is, changes in supply and demand for many commodities appear to remain roughly in balance for a considerable number of years, at least in the United States. On occasion, a shift in demand relative to supply can persist and shift the level of prices; as illustrated for wheat prices in Figure 9-4, this appears to have occurred on two occasions between 1960 and 2012. If an analyst is not aware of such jumps in level, an estimated trend equation may appear to be statistically important, even though the price behavior would be better modeled by allowing for discrete shifts at points in time.

A potential source of confusion in discussing trends is the distinction between nominal and deflated prices. When the general price level is increasing, dividing a nominal price (with a constant mean) by an index (with an increasing mean) results in a deflated ("real") price series that is trending downward. Thus, statistical analyses of the deflated price series may find trends. Relative prices and their possible trends are discussed further in Chapter 10.

We emphasize, however, that no inherent theoretical reason exists for spot prices of an individual commodity to be trending.[7] Although good reasons exist for seasonality in prices and perhaps for cyclical behavior as well, it need not be the case that nominal commodity prices have trends. Of course, a commodity's price may trend, at least when viewed ex post, but when a historical trend has been identified, it is dangerous to assume that it will continue in the future. The underlying fundamental cause of the trend may, indeed probably will, change. In terms of the simple model introduced here, the ability to forecast a trend in price, $P_t$, would require the analyst to forecast the shifts in supply and demand, $X_t$ and $Y_t$, and this is difficult to do.

## Short-Time Price Variation

Unlike many products sold at retail, commodity prices at the farm and wholesale levels can change weekly, daily, and even within a day, and this section addresses the behavior of prices for short time units. For convenience, we will think in terms of daily prices and use daily prices for soybeans in central Illinois as an example; the data are plotted in Figure 9-7.

Given the assumption of a perfectly competitive market with no changes in supply and demand, the daily adjustments will merely reflect the systematic components discussed in the previous sections. For example, if today's price of a storable grain is \$6.055 per unit, if the only systematic component is seasonality, and if the storage cost is 0.2 cent (\$0.002) per unit per day, then tomorrow's

7. Likewise, no fundamental reason exists for nominal prices of commodities to have a *stochastic* trend. The notion of a stochastic trend is that the time series is following a *random walk* (essentially, it has no systematic behavior), but as this chapter shows, commodity prices are likely to have systematic components that cannot be eliminated by arbitrage. Hence, one should not expect cash prices to follow a random walk, and attempts in the literature to find stochastic trends in nominal commodity prices seem misplaced. Futures prices—as distinct from spot prices—may follow approximately a random walk (Chapter 12).

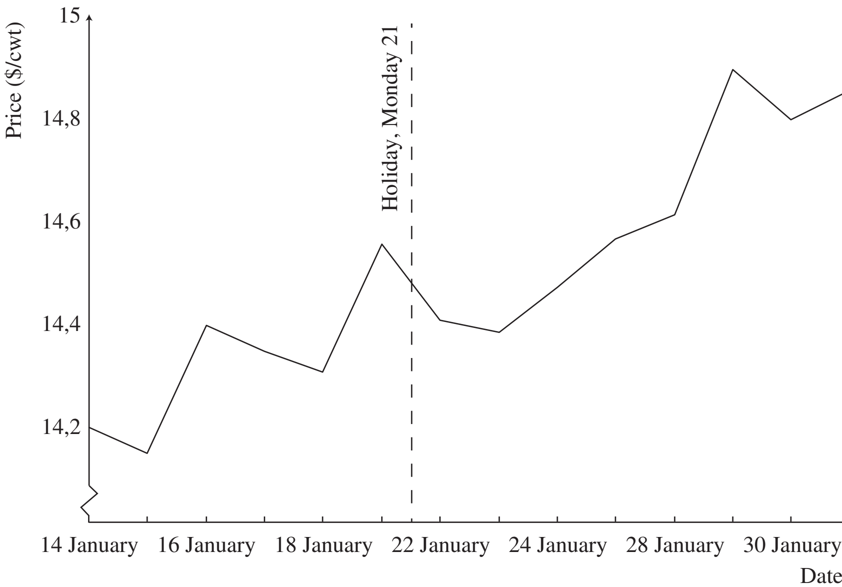



**Figure 9-7.** Price of number 1 yellow soybeans, central Illinois, days, January, 14–February, 1, 2013

price will be \$6.057, the next day's price \$6.059, and so on. Over a 30-day storage period, prices will rise 6 cents (0.2 cent × 30 days), and if we observe average prices in consecutive months, they will differ by 6 cents.

Of course, in practice, new information that influences prices arrives frequently. A more or less continuous process of evaluating the information takes place. Today's price is conditional on today's information; in notation, this is $P_t|I_t$. As new information arrives in the market, prices change, say, to $P_{t+1}|I_{t+1}$. Estimates of the size of the inventories may be revised, or say, prices may respond to a change in the number of hogs marketed. On the demand side, an unexpected order may enter the market or news may arrive about a possible pesticide residue in a grain, which reduces demand. And so forth.

The new information also may pertain to expected (future) changes in the factors that influence price (Working 1958). For example, the market's estimate of the next year's crop size is changed by new information about weather conditions and/or by news arriving about expected changes in domestic or international government policies. The precise mechanism for the linkage of current cash prices to prices for future delivery will be discussed further in Chapter 12, but the general idea is that decisions made today influence prices in subsequent time periods. For instance, to provide inventories to meet an increase in future demands, current consumption must be reduced, and this requires a higher current price.

Truly new information is a surprise (i.e., is not expected) and is analogous to a random variable. It is a revision of past information. The effect appears as random variability in prices around the systematic components. Markets are assumed to adjust appropriately to new information, but with the passage of time, information changes and hence prices adjust.

The magnitude of the effect of new information depends on the economic context in which it arrives. For a storable commodity, news about an increase in demand can have a small effect on price if stocks are large but can have a large effect on price if stocks are small. The prices illustrated in Figure 9-7 are in the context of soybean inventories in January 2013, which were small relative to expected demand. Thus, new information resulted in some relatively large positive price changes from one day to the next. Conceptually, this behavior can be explained in terms of a non-linear supply of storage equation (introduced in Chapter 5 and discussed in Chapter 12).

Since the production of many farm commodities has a seasonal component, the frequency of the flow of information can be seasonal. In particular, news about the size of next year's crop and the timing of its harvest is associated with the growing season. There is a steady flow of information about the North American soybean crop from May through September. By mid-September, the size of crop and the timing of its harvest is relatively well established, and there is relatively little new news about this crop's supply at harvest time and immediately thereafter. The consequence of this seasonality in news frequency is seasonality in the variance of daily (as well as weekly and monthly) prices. Moreover, inventories for these commodities are declining over this same time period. Thus, conceptually, one can visualize probability distributions of daily prices, in which the mean, variance, kurtosis, and skewness parameters are changing systematically with the passage of time (e.g., Yang and Brorsen 1992).

Still another factor influencing the behavior of daily prices is the structure of the market and transaction costs. Even if markets have large numbers of buyers and sellers—are relatively efficient—the cost of acquiring and acting on new information may result in persistence in price behavior from one day to the next (Grossman and Stiglitz 1980). Also, spot markets may have varying degrees of pricing efficiency; they may differ in the quantity and quality of information available to traders, in the cost of obtaining and using the information, and hence in the degree to which traders are well informed. In principle, a large proportion of traders could be poorly informed. In this case, price adjustments would be slower than in an efficient market, and it is likely that some portion of the today's price change would be the result of old news. Thus, daily prices could have an autocorrelation component that is related to transactions costs and possibly to market imperfections. This transaction effect is thought to be small in markets in the United States and other high-income countries, but it may be important in markets in less-developed economies.

The prices for some agricultural products in countries like the United States are established within an institutional framework that may reduce or eliminate daily price changes, particularly the transaction price received by farmers. Contracts can be negotiated between sellers and buyers that fix the price received by the seller or that base the price received on a formula. An increasing proportion of transactions prices at the farm level in the United States is determined by contract provisions. There are many types of contracts. One type is a forward contract that offers the producer a fixed price, which is received when the crop is delivered. The contracted price is typically based on a current futures market quote. Thus, the contracted price depends on the time and date that the contract is consummated. Such contracts assure a certain price to the producer for a contracted quantity, but market prices remain variable. Buyers are able to offer a forward contract, with a known price, because they can either hedge the risk of price changes in a relevant futures market or they expect to profit from the contract. The effects of alternative price discovery mechanisms are discussed in Chapter 11; hedging is discussed in Chapter 13.

## Models of Time Series

Based on the discussion in this chapter, a time series of commodity prices can be viewed as consisting of systematic and random components. A notation is $P_t = S_t + e_t$, where $P$ is the price, $S$ is the systematic component, $e$ is the error or random component, and the $t$ subscript stands for the time dimension. Depending on the frequency of data observations, $t$ can stand for days, weeks, months, quarters, or years.

Observations on prices and other related variables can be used to study price behavior, but the analyst must specify a model of the time-series process generating the observations, although of course the exact process is unknown. The general, conceptual models do not provide sufficient information to identify exact data-generating mechanisms for specific commodities.

Thus, numerous alternative models exist in the literature. One general category of alternative models is *structural* econometric models. Such models attempt to capture the underlying supply and demand characteristics of markets. A second category of statistical models is *associational* models, which attempt to capture the empirical regularities in the data. For example, current price can be made a function of lagged prices and an error term; this is called an autoregressive model. In addition to statistical models with an error term, some empirical analyses assume purely deterministic processes. For example, simple or weighted averages of current and past prices are used to forecast forthcoming prices and take positions in futures markets.

We discuss a few of these models in the last three chapters of the book, but the literature is vast and can be introduced only briefly. Moreover, a deep understanding of many of these methods requires a background in mathematics and statistics beyond the level of this book.

## Concluding Remarks

Market prices for agricultural commodities are variable. This can be attributed broadly to three factors. First, commodity markets are relatively competitive, and farm-level (derived) demand and supply are typically price inelastic. Thus, shifts in one relationship relative to the other can result in large price changes. Second, the biological nature of the production process means that output is partly dependent on random events and is often seasonal in nature. Hence, the size of inventories influences prices. Third, and related, farmers and inventory-holders must use expected prices in making decisions, and the way expectations are formed and used introduces systematic price and quantity variability.

Since expectations are typically not realized, yield and price risk exist. Although prices have systematic behavior, precise forecasts are difficult to make. The consequence is price risk, which can be viewed as a cost to farmers as well as to buyers of commodities. For many years, governments have sought ways to help farmers reduce or shift price risks. This has involved programs to support prices, i.e., to put a floor under prices, or to supplement incomes. Recently, in the United States, emphasis has been placed on managing risk through revenue insurance programs and/or by using futures and options markets to hedge. We have a bit more to say about these ideas in Chapters 11, 12, and 13.

## Appendix I: Simple Seasonal Model

The following model allocates a fixed quantity, $A$, over $T$ intra-year time periods. It is adapted from Bressler and King (1970: Chapter 11). The model assumes

$A$ is the fixed supply for the marketing year;

$Q_t = a + bP_t$, identical demand equation in each time period (say "months"), and $b < 0$;

$t = 0, 1, 2, \ldots, T-1$ time periods and zero carryover after $T-1$; hence,

$\sum_{t=0}^{T-1} Q_t = A$; and

$C_t = mt$, storage cost per unit of $Q_t$, which is a linear function of time.

In sum, $A$, $a$, $b$, $m$, and $T$ are assumed to be known at the initial time period, 0. The model thus determines the initial price, $P_0$, and subsequent prices that exactly allocate $A$ over the $T$ periods.

In perfect competition, the marginal cost (MC) of storage equals the marginal revenue (MR). By definition, MR = $P_1 - P_0$, and

$$MC = \frac{\partial C_t}{\partial t} = m.$$

Hence,

$$\begin{aligned} P_1 - P_0 &= m \\ P_2 - P_1 &= m \\ P_2 - P_0 &= 2m, \\ &\vdots \end{aligned}$$

Prices increase by the amount $m$ each time period.

The solution for the $P_t$'s and $Q_t$'s is as follows:

$$\begin{aligned} A &= \sum_{t=0}^{T-1} Q_t \\ &= \sum (a + bP_t) \\ &= Ta + b\sum P_t = Ta + b\sum (P_0 + tm) \\ &= Ta + TbP_0 + bm\sum t. \end{aligned}$$

Since

$$\sum_{t=0}^{T-1} t = \sum_{t=1}^{T-1} t = \frac{T(T-1)}{2},$$

$$A = Ta + TbP_0 + \left[\frac{T(T-1)}{2}\right]bm.$$

$$\begin{aligned} P_0 &= \frac{1}{Tb}A - \frac{Ta}{Tb} - \frac{T(T-1)}{Tb2}bm \\ &= \frac{A}{Tb} - \frac{a}{b} - \frac{(T-1)m}{2}. \end{aligned}$$

The seasonal prices and quantities are

$$\begin{aligned} &Q_0 = a + bP_0 \\ &Q_1 = a + bP_1, \quad \text{where } P_1 = P_0 + m \\ &\vdots \\ &Q_{T-1} = a + bP_{T-1}, \quad \text{where } P_{T-1} = P_0 + m(T-1). \end{aligned}$$

And by definition $\sum Q_t = A$

Note, the initial price depends on $A$, $a$, $b$, $T$, and $m$, where $b < 0$. If the market has erroneous information, then the initial price will be wrong. Subsequently, prices will need to adjust as better information enters the market.

Also, the model assumes a particular form of storage costs. In financial economics, it is common to specify

$$C_t = e^{m(T-t)},$$

where $e = 2.718 \ldots$;
$t = 1, 2, \ldots, T$;
$m = i + d - c$;
$i$ is the interest rate (annualized if $T - t$ is in proportion of years);
$d$ is the storage costs as percentage of time $t$ price; and
$c$ is the convenience yield (sometimes omitted).

It follows that seasonal prices will have a non-linear form. In other words, the nature of seasonal prices depends on the underlying storage costs per unit of time.

## Appendix II: An Elementary Cobweb Model

The simplest cobweb model can be written as three equations:

$$A_t = \delta + \gamma P_{t-1} \text{ (supply assuming naive expectations).}$$

$$A_t = Q_t \text{ (market clearing).}$$

$$P_t = \alpha - \beta Q_t \text{ (inverse demand).}$$

With price on the vertical axis, the slopes are

$$\frac{dP}{dQ} = -\beta \text{ for the demand relation, and}$$

$$\frac{dP}{dQ} = \frac{1}{\gamma} = \gamma^{-1} \text{ for the supply relation.}$$

Thus, the slope conditions for the three types of cycles are

$$|-\beta| > |\gamma^{-1}| \text{ a divergent cycle,}$$

$$|-\beta| < |\gamma^{-1}| \text{ a convergent cycle, and}$$

$$|-\beta| = |\gamma^{-1}| \text{ a continuous cycle.}$$

One method of computing the *paths of* $Q$ and $P$ through time is by substitution. Although not recommended for actual computations, it may help the reader to understand the conditions leading to various types of cycles. Since

$$P_t = \alpha - \beta Q_t, Q_t = A_t,$$

and, hence,

$$Q_{t+1} = \delta + \gamma P_t,$$

it follows that

$$Q_{t+1} = \delta + \gamma(\alpha - \beta Q_t) = (\delta + \gamma\alpha) - \gamma\beta Q_t.$$

Also,

$$\begin{aligned} Q_{t+2} &= (\delta + \gamma\alpha) - \gamma\beta Q_{t+1} \\ &= (\delta + \gamma\alpha) + \gamma\beta[(\delta + \gamma\alpha) - \gamma\beta Q_t] \\ &= (\delta + \gamma\alpha)(1 - \gamma\beta) + (\gamma\beta)^2 Q_t. \end{aligned}$$

Let $t = 0, 1, 2, 3 \ldots$, and the equations for computing $Q$ in each period may be summarized as follows:

$$\begin{aligned} Q_1 &= (\delta + \gamma\alpha) - \gamma\beta Q_0. \\ Q_2 &= (\delta + \gamma\alpha) + (1 - \gamma\beta) + (\gamma\beta)^2 Q_0. \\ Q_3 &= (\delta + \gamma\alpha)\left[(1 - \gamma\beta) + (\gamma\beta)^2\right] - (\gamma\beta)^3 Q_0. \\ &\vdots \end{aligned}$$

Of course, to compute the sequence of $Q$'s, we must have estimates of $\alpha$, $\beta$, $\gamma$, and $\delta$ and know the initial value of $Q_0$.

Since $\beta$ is negative, the levels of $Q$ are oscillating from period to period. The conditions for the three types of cycles also may be stated as

$(\gamma\beta)^2 > 1$ for a divergent cycle,

$(\gamma\beta)^2 < 1$ for a convergent cycle, and

$(\gamma\beta)^2 = 1$ for a continuous cycle.

For instance, if $(\gamma\beta)^2 = 1$, that is, $|-\beta| = |\gamma^{-1}|$, then our sequence of $Q$'s becomes

$$Q_1 = (\delta + \gamma\alpha) - (1)Q_0;$$
$$Q_2 = (\delta + \gamma\alpha)(0) + Q_0 = Q_0;$$
$$Q_3 = Q_1;$$
$$Q_4 = Q_0.$$

In other words, the $Q$'s oscillate between

$$[(\delta + \gamma\alpha) - Q_0] = Q_1 = Q_2 = Q_3 = Q_5 = \text{etc.}$$

and

$$Q_0 = Q_2 = Q_4 = \text{etc.}$$

Readers with some mathematical background may have observed that $Q_{t+1} = \gamma\beta Q_t$ and $P_{t+1} = \gamma\beta P_t$ are first-order difference equations. General solutions are available for such equations at various points $t$ through time, but this topic is beyond the scope of this chapter.

The time path for the $P_t$'s is found analogously to the path for $Q_t$. These paths will be modified if an alternative definition of price expectations is used. For example,

$$A_t = \delta + \gamma P^*_{t-1}$$

and

$$P^*_{t-1} = \pi_1 P_{t-1} + \pi_2 P_{t-2},$$

then

$$A_t = \delta + \gamma(\pi_1 P_{t-1} + \pi_2 P_{t-2})$$
$$= \delta + \gamma\pi_1 P_{t-1} + \gamma\pi_2 P_{t-2}.$$

Given

$$A_t = Q_t$$
$$P_t = \alpha - \beta[\delta + \gamma\pi_1 P_{t-1} + \gamma\pi_2 P_{t-2}].$$
$$= (\alpha - \beta\delta) - \beta\gamma\pi_1 P_{t-1} - \beta\gamma\pi_2 P_{t-2}.$$

$P_t$ is now autocorrelated with $P_{t-1}$ and $P_{t-2}$.

## References

Bressler, R. G., Jr., and R. A. King. 1970. *Markets, Prices, and Interregional Trade*. New York: John Wiley and Sons.

Deaton, A., and G. Laroque. 1992. "On the Behavior of Commodity Prices," *Rev. of Econ. Studies* 59:1–23.

Enders, W., and M. T. Holt. 2012. "Sharp Breaks or Smooth Shifts? An Investigation of the Evolution of Primary Commodity Prices," *Am. J. Ag. Econ.* 94:659–673.

Ezekiel, M. 1938. "The Cobweb Theorem," *Quart. J. Econ.* 53:255–280.

Feige, E. L., and D. K. Pierce. 1976. "Economically Rational Expectations: Are Innovations in the Rate of Inflation Independent of Innovations in Monetary and Fiscal Policy?" *J. Pol. Econ.* 84:499–522.

Gardner, B. L. 1976. "Futures Prices in Supply Analysis," *Am J. Ag. Econ.* 58:81–84.

Grossman, S. J., and J. E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets," *Am. Econ. Rev.* 70:393–408.

Holt, M. T., and L. A. Craig. 2006. "Nonlinear Dynamics and Structural Change in the U.S. Hog-Corn Cycle: A Time-Varying STAR Approach," *Am. J. Ag. Econ.* 81:215–233.

Hull, J. C. 2009. *Options, Futures, & Other Derivatives*, 7th ed. Upper Saddle River, NJ: Prentice Hall.

Mundlak, Y., and H. Huang. 1996. "International Comparisons of Cattle Cycles," *Am. J. Ag. Econ.* 78:855–868.

Muth, J. F. 1961. "Rational Expectations and the Theory of Price Movements," *Econometrica* 29:315–335.

Nerlove, M. 1956. "Estimates of the Elasticities of Supply of Selected Agricultural Commodities," *J. Farm Econ.* 38:496–509.

New York Agricultural Statistics Service. 2012. *New York Agricultural Statistics, 2011–2012*. Albany, NY.

Peterson, H. H., and W. G. Tomek. 2005. "How Much of Commodity Price Behavior Can a Rational Expectations Storage Model Explain?" *Ag. Econ.* 33:289–303.

Rosen, S. 1987. "Dynamic Animal Agriculture," *Am. J. Ag. Econ.* 69:547–557.

Sarmiento, C., and P. G. Allen. 2000. "Estimating Intertemporal Supply Response in the Fed Beef Market," *Am. J. Ag. Econ.* 82:630–634.

Wang, D., and W. G. Tomek. 2007. "Commodity Prices and Unit Root Tests," *Am. J. Ag. Econ.* 89:875–889.

Waugh, F. V. 1964. "Cobweb Models," *J. Farm Econ.* 46:732–750.

Working, H. 1958. "A Theory of Anticipatory Prices," *Am. Econ. Rev.* 48:188–199.

Yang, S.-R., and B. W. Brorsen. 1992. "Nonlinear Dynamics of Daily Cash Prices," *Am. J. Ag. Econ.* 74:706–715.

CHAPTER **10**

# General Farm–Non-farm Price Relationships

This chapter focuses on the general level of farm prices, not on the prices of individual commodities. Measures of the general, or average, level of prices attempt to aggregate over the diverse behavior of individual price changes. Price movements can be visualized as a swarm of insects, with the individual insects changing positions relative to each other but with the group as a whole either rising or descending. Averages, represented by price indexes, help identify the underlying general movements in prices.

Changes in the general level of farm prices have political as well as economic consequences. For example, a rising price level makes it easier for farmers (and others) to pay off debts that are fixed in nominal terms. Deflation, in contrast, tends to transfer real income from debtors to creditors, and farmers are usually debtors, especially those who have recently entered farming or have expanded their business. Economic hard times often influence farm and other government policies.

Given the potential importance of changes in the general level of farm prices to farmers and other agribusiness firms, the causes of major price movements are first reviewed. These factors originate both from the farm and the non-farm sectors of the economy. Thus, macroeconomic variables, like monetary, fiscal, and trade policies, can be important influences on prices. The second section discusses issues of measuring changes in the price level and interpreting the resulting indexes; the third section examines changes in the prices of farm products relative to other prices. We also provide a detailed appendix to the chapter on the construction and use of index numbers.



## Variables Influencing the General Level of Farm Prices

Changes in the general level of farm product prices may be viewed as resulting from shifts in aggregate demand relative to aggregate supply.[1] To expand on

1. Cochrane (1958) used aggregate supply and demand schedules in his influential book to elucidate the reasons for major swings in farm prices.

this idea, a model of farm and retail prices (introduced in Chapter 6) is used as a framework. This model helps show the linkages among the retail, marketing, and farm sectors in the economy. It consists of six equations: retail-level supply and demand, supply and demand for the marketing sector's inputs, and supply and demand for the farm sector's inputs (commodities). Rather than considering an individual commodity and product, like wheat and bread, however, we now are thinking in aggregate, sector terms, e.g., in terms of the agricultural/food sector.

Three of the six equations can be used to help give a sense of the interrelationships among prices. The aggregate retail-level demand for food is written in inverse form as

$$P = D^{-1}(Q, Y),$$

where P is a measure of the retail price of food in the aggregate, say, the total food component of the Consumer Price Index (CPI) in the United States; $Q$ is a measure of aggregate food consumption; and $Y$ is a set of demand-shift variables, including income, population, and the prices of all other goods and services in the economy.

The farm-level derived demand function may be specified as

$$P_{\mathrm{A}} = D_{\mathrm{A}}^{-1}(A, P, P_{\mathrm{B}}),$$

where $P_{\mathrm{A}}$ is an index of prices received by farmers, $A$ is a measure of aggregate farm output that is used to manufacture $Q$, $P_{\mathrm{B}}$ is an index of prices of inputs used in manufacturing and marketing foods (transforming $A$ to $Q$), and $P$ is the index of retail food prices. Note that $P_{\mathrm{A}}$ depends, among other things, on the retail-level demand function. The derived demand for commodities is also influenced by international markets, both imports and exports of commodities.

The farm-level supply of $A$ is specified as

$$A = A(P_{\mathrm{A}t-1}, X),$$

where $P_{\mathrm{A}}$ and $A$ are as defined above, and $X$ is a set of supply shift variables. Two key variables contained in $X$ are an index of lagged prices of inputs used in farming (an index of prices paid by farmers for their inputs) and a measure of technological change.

These three equations help make clear that the prices received by farmers, $P_{\mathrm{A}}$, are influenced by many factors, some that originate within the agricultural sector but others that originate in other sectors of the national and international economy. The demand for farm products depends on national incomes and populations, while the supply of farm products depends on the prices of inputs used by producers. Clearly, many of these inputs come from the non-farm sectors of the economy. One such price is the cost of capital, namely interest rates.

Improved production technologies are an important driver of supply increases in agriculture, and new technologies depend, in turn, on research. Output from research depends on the dollars invested by private firms and by the public sector (both economic and political decisions) in research.

Thus, fundamental questions include, why do incomes grow (or not grow) and what influences the stream of new technologies entering agriculture? Deep answers to these questions are beyond the scope of this book, but we do note that growth in incomes and technology is influenced by the economic policies of nations. Changes in monetary and fiscal policies, in environmental policies, in trade policies, and in exchange rates are among the variables that can affect the general price level, including the prices of farm products (e.g., Antle 1999; Johnson 2000). For example, increases in the growth rate of the national economies like China and India increase their demand for grain imports. Changes in aggregate demand result, *ceteris paribus*, in changes in commodity prices.

A given change in income can have differential effects on different sectors of the economy, which depend on the differences in income elasticities of demand. When we generalize this idea to think in terms of a matrix of own-price, cross-price, and income elasticities for various sectors of the economy, it is clearly possible that a given exogenous change in a national economy (e.g., a policy change) can have differential effects on the farm and non-farm sectors. Farm and non-farm price relationships are discussed further in a later section.

Price levels have tended to increase in most countries since the end of World War II in 1945, although the rate of inflation has varied greatly from country to country and from one time period to another. In the United States, annual rates of inflation since 1975, as measured by the CPI for all items, have varied from over 11 percent (1978–79) to −0.4 percent (2008–09). For the 10 years ending in 2012, the average annual rate of inflation was 2.2 percent. Other countries, notably some in Latin America, have experienced inflation rates of as high as 50–100 percent per year. The inflation rate in Zimbabwe in 2008 is estimated to have been 2,600 percent *per month*, a hyperinflation (for even higher rates, see Koech 2012: table 1).

Internal macroeconomic policies are responsible for the varying rates of inflation among countries. Rapid and sustained inflation cannot occur unless a country's government creates more money or otherwise pursues policies that accommodate inflation. Inflationary pressures on prices may develop as a result of increases in either government or private demand that are not matched by increases in output. A government, for example, may create money to meet its increased demands without resorting to increasing taxes. Hyperinflation can develop only if government authorities behave irresponsibly. Massive amounts of money are required to sustain high rates of inflation. (In Zimbabwe, the government issued 100 trillion-Zimbabwean-dollar banknotes!) Governments may be reluctant to curb the growth in the money supply (including bank deposits) because they fear the political consequences, since an anti-inflation policy may result in a recession and an increase in unemployment. Ultimately, however, policy changes are required because a country's currency becomes worthless and cannot be used for transactions.

Farm and food prices are influenced by general rates of inflation, but are far from perfectly correlated with the general price level. Consider Figures 10-1 and 10-2, which plot the index of prices received and paid (Figure 10-1) and the index of prices received and the CPI (Figure 10-2) from 1970 to 2012. Some of the most dramatic increases in farm prices in the twentieth century were associated with wars or the threat of wars. Since 1970, there have been two periods with large jumps in farm prices. The first (starting in 1973) was associated with a change in grain-import policies of the (former) Soviet Union. Prior to 1973, the Soviet Union was largely a closed economy that traded primarily with its eastern bloc neighbors. But a change in policy resulted in a large increase in the Soviet demand for grains, which in turn was intended to help increase the production of meat. This resulted in a notable increase in commodity prices, and prices did not return to their former level (Figure 10-1). The second (starting in 2006) was also primarily on the demand side, this time because of the huge increase in demand for grain for biofuels (especially corn for ethanol in the United States) and for exports to China and India, but also on the supply side with increases in production costs, mainly due to increasing energy prices.

In the mid- to late 1970s, both prices received and paid by farmers tended to rise with general inflation, but from 1980 to about 2006, the index of prices received by farmers tended to vary around a constant mean. The ratios of prices received to prices paid by farmers, therefore, declined over this period (Figures 10-1 and 10-2). From 2006 to the end of 2012, both the index of prices received and paid rose at significantly higher rates than general inflation rate (the index of prices received annually increased by an average of 10.9 percent and the index of prices paid annually increased by 7.2 percent from 2006 to 2012, while the CPI only grew by an average of 2.3 percent each year during this period). The ratio of prices received to prices paid rose slightly from 2006 but was unchanged in 2012. The experience of the United States during the 1970s demonstrated that an upward movement in commodity prices, whether caused by an increase in world demand or a decrease in aggregate output, can be reinforced by an easy money policy. A reversal of such policies, such as occurred in the early 1980s, can lead to lower commodity prices (e.g., from 1983 to 1985 in Figure 10-1), especially if the more restrictive monetary policy coincides with a period of increasing production and declining world demand. Since 2006, both prices received and paid by farmers have dramatically increased due to the factors discussed.

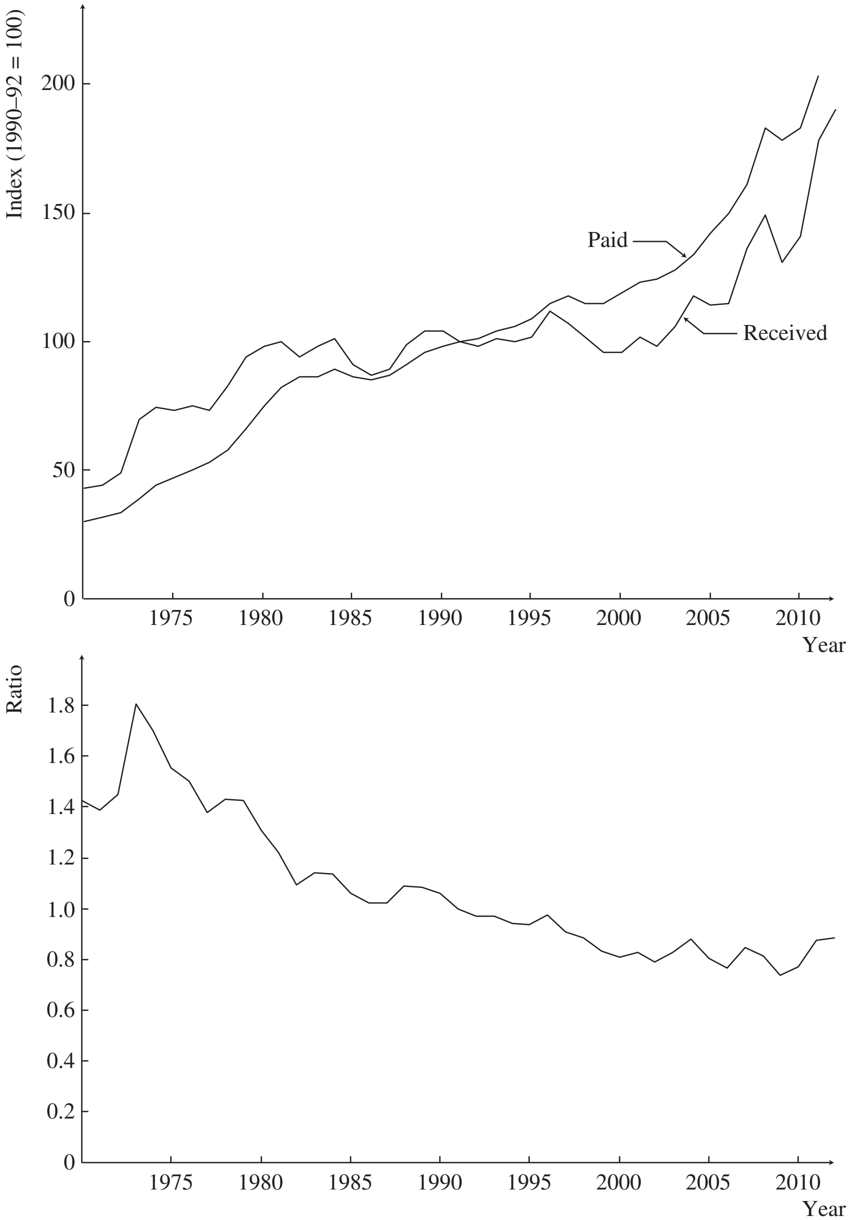



**Figure 10-1.** Indexes of prices received and prices paid by farmers, and their ratio, United States, 1970–2012

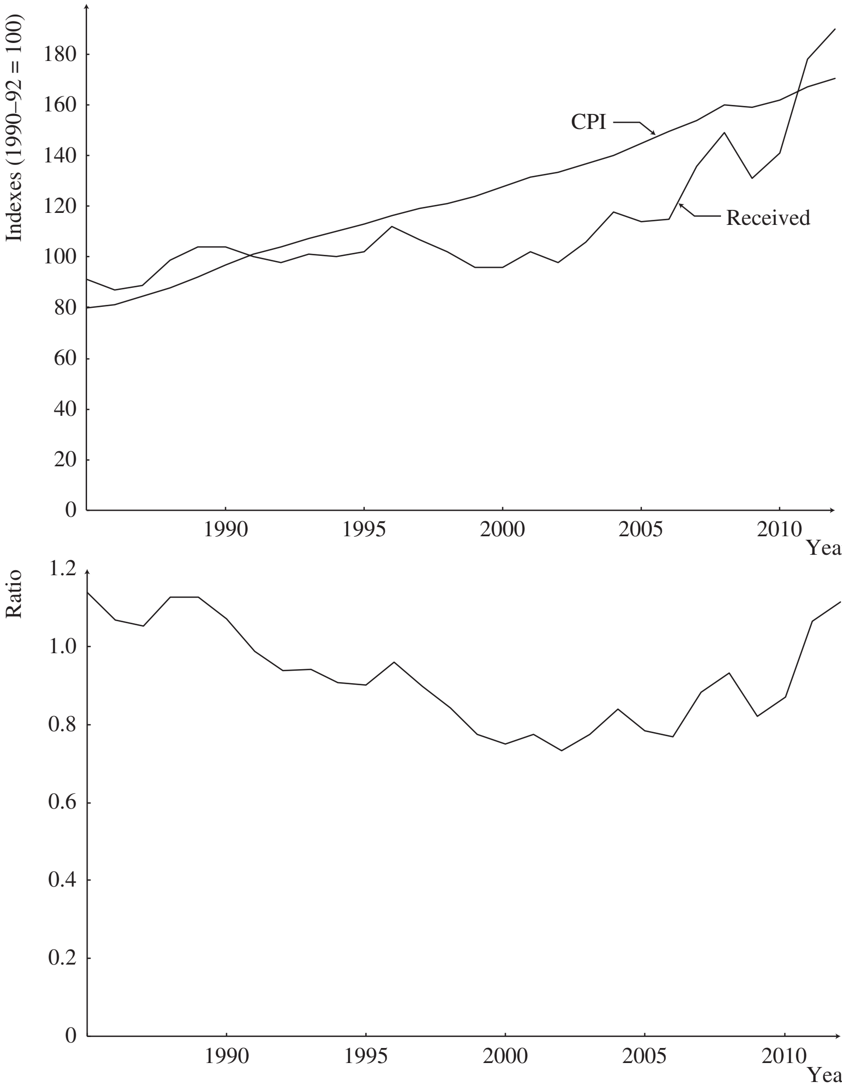



**Figure 10-2.** Indexes of prices received by farmers and the CPI, and their ratio, United States, 1985–2012

For the world as a whole, the evidence suggests that the supply of food matched or exceeded the growth in the demand for food in the last half of the twentieth century. Commodity prices tended to decline relative to other prices, as reflected by the ratio of price received to the CPI in Figure 10-2. As discussed later in the chapter, this trend reversed after 2005.

Within the global economy, relative prices in individual countries depend on agricultural, trade, and macroeconomic policies of individual countries. Agricultural policies influence production incentives, and trade policies influence the demand for and supply of exports and imports. Macroeconomic policies influence prices through their effect on interest rates and incomes. Policies that influence the supply and demand for credit can also affect exchange rates, assuming they are free to respond to market forces. Farmers in the United States learned during the 1970s and early 1980s what producers of export commodities in other countries have known for years—namely that changes in exchange rates can have a major impact on the prices of exports and/or the volume of exports.

In general, one expects prices to rise in a country that devalues its currency (or where market forces result in a devaluation) because imported items become more expensive. Devaluation encourages exports, thereby adding to the demand for goods sold abroad. But unilateral devaluation by one country does not always produce these effects because of countervailing actions by importing or competing exporting nations. Indeed, where countries allow their currency prices to move freely, market forces determine the exchange rates.

The following example illustrates the effects of changes in relative currency movements. If the U.S. dollar declined by 20 percent relative to the Japanese yen (Y), say from Y150 to Y120 per dollar, then in theory a ton of U.S. grain priced at $100 per ton, which formerly cost the Japanese Y15,000 ((150Y/$1) × $100), now costs Y12,000 per ton. Other factors remaining constant, the decline in the value of the dollar (the dollar buys fewer yen, i.e., the yen buys more dollars) makes U.S. exports less expensive for Japanese to purchase but makes Japanese imports more expensive for Americans to buy. If, however, Japan were to tax imports by an equivalent amount to the devaluation (or to use trade quotas), then there would be little increase in the Japanese demand for grains.

Foreign aid and international lending policies also can influence the demand for imports and, therefore, commodity prices. An increase in aid to less-developed countries, for example, helps to ease one of the constraints limiting their capacity to import, namely a shortage of foreign exchange (say, U.S. dollars). Considerable debate exists about foreign trade policy because of the trade-offs between the benefits and costs of reducing barriers to trade. Policies that help less-developed countries increase exports will help them increase foreign exchange earnings. But concerns exist about the effects of these larger exports on the competing industries in other countries. A discussion of these trade-offs is beyond the scope of this book. The general point of the foregoing discussion is that a large variety of economic policies around the world can influence the supply and demand for commodities and hence commodity prices.

## Measuring Changes in the General Level of Prices

Because index numbers are used to measure rates of inflation and relative price movements, it is important for price analysts to understand the composition and potential sources of bias in the index numbers that they are using. The index numbers of major interest to agriculture and those used to measure overall rates of inflation (or deflation) are described in this section.[2]

From the standpoint of agricultural product prices, the most important index numbers are those that measure the average level of prices *received* and *paid* by farmers. For the United States, these index numbers are reported in the USDA's *Agricultural Prices* (e.g., USDA 2000; also available in secondary sources). The index of prices received by farmers reflects the average level of all major commodities sold by farmers, with the relative importance of weights attached to each commodity based on the contribution of that commodity to the total revenue of farmers. Sub-indexes are computed for groups of commodities such as food grains (wheat, rice, and rye), feed grains (corn, oats, sorghum, and barley), fruits, vegetables, and livestock and livestock products.

The index of prices paid by farmers includes a component for items used in production (energy, fertilizer, seed, feed, etc.), farm wages, interest, and items purchased for home consumption. The USDA uses the CPI to reflect the cost of living in farm homes rather than attempting to duplicate the work of the Bureau of Labor Statistics in collecting consumer prices.[3]

Overall rates of inflation are commonly measured using one of three index numbers: the CPI, the Producer Price Indexes (PPIs), or the GDP (gross domestic product) Implicit Price Deflator. The CPI attempts to measure the cost of living of households, but it does not cover changes in governmental or environmental factors that influence consumers' well-being. Thus, it focuses on prices of food, apparel, housing, transportation, medical care, entertainment and recreation, education, and other goods and services used by consumers. It excludes, therefore, the "cost of living" of the government, i.e., the prices paid by governments for the goods and services they use. Likewise, it is not a measure of the costs of inputs used by manufacturers or farmers.

It is not really appropriate to speak about "the PPI" in the United States. Rather, three general indexes are reported for three stages of processing: for crude materials used for further processing; for intermediate materials, supplies, and components; and for finished goods. In other words, three general wholesale price indexes are available for the differing stages of the manufacturing and marketing processes. The PPIs collectively cover many prices at wholesale levels in the economy, such as consumer durables and non-durables, capital equipment, materials and components, and processed fuels and lubricants.

2. The appendix to this chapter provides a brief explanation of index number construction, sources of bias, and uses. One of the uses of price indexes is to deflate a nominal price series to obtain "real" prices, i.e., to adjust for changes in the overall price level. As this section points out, price indexes differ. Hence, the choice of a deflator can have a profound effect on empirical analyses.

3. Indexes similar to those discussed for the United States are computed in many other countries. High-quality price indexes, however, depend on high-quality inputs; hence, the quality of indexes does vary from country to country.

These indexes, however, do not directly cover services, which are an increasingly important part of the economy.

The GDP deflator index is the most comprehensive of all of the index numbers because it is based on the total value of goods and services produced in the domestic economy each year. The major components of GDP are personal consumption expenditures, gross private domestic investment, government consumption expenditures and gross investment, and exports and imports of goods and services. The GDP deflator is used to estimate the real (deflated) value of the GDP of the economy. In other words, the total nominal value of national output is estimated and then adjusted by the GDP deflator to obtain real economic growth.

Estimates of the rate of inflation will vary depending on which index is used to measure changes in the average level of prices. This is illustrated in Table 10-1. The three price indexes for the United States were transformed to a common base period (2005 = 100) for the period 1970–2010. These indexes were then used to obtain the percentage changes over five-year intervals (Table 10-1). Two broad points are demonstrated by these data. First, the rates of inflation were significantly higher from 1970 to 1990 than from 1990 to 2010. Second, the rates vary depending on the index used. The highest rate of growth in prices occurred for the PPIs in the 1970s. This was related to the increased export demand for commodities, mentioned earlier, and to the inflation in energy prices related to OPEC's policies during that decade. Over the entire 40 years, the CPI had the largest growth.

The different behavior of these general measures of inflation must be kept in mind in interpreting real ( i.e., deflated) prices or incomes. The results depend on which index is used to adjust the actual (nominal) data. For example, it may be logical to divide consumers' income by the CPI to obtain the real purchasing power of their income, but it probably is not logical to divide a commodity's

**Table 10-1.** Price indexes, United States, 2005 = 100

| Year | CPI, all items | Change (%) | GDP implicit deflator index | Change (%) | PPI, intermediate goods | Change (%) |
|---|---|---|---|---|---|---|
| 1970 | 20 | — | 24 | 22 | 23 | 13 |
| 1975 | 28 | 39 | 34 | 38 | 38 | 64 |
| 1980 | 42 | 53 | 48 | 42 | 59 | 56 |
| 1985 | 55 | 31 | 62 | 29 | 67 | 14 |
| 1990 | 67 | 21 | 72 | 17 | 74 | 11 |
| 1995 | 78 | 17 | 82 | 13 | 81 | 9 |
| 2000 | 88 | 13 | 89 | 9 | 84 | 3 |
| 2005 | 100 | 13 | 100 | 13 | 100 | 19 |
| 2010 | 112 | 12 | 111 | 11 | 119 | 19 |

*Source:* Data from the *Economic Report of the President* (2013).
*Note:* CPI, Consumer Price Index; GDP, gross domestic product; PPI, Producer Price Indexes.

price by the CPI to obtain the relevant relative price faced by farmers in making production decisions.

Many price indexes in the United States and other countries, such as indexes of prices paid and received by farmers and retail price indexes, are based on a "fixed mix" of goods and services. That is, fixed weights are applied to a given set of items priced in the indexes; however, with the passage of time, the items included in the index and their weights are revised. Also, the time period used for estimating the weights need not be the same as the base period for the prices. (These topics are discussed more fully in the appendix to this chapter.) Consequently, an index such as the CPI reflects the changes in prices that have occurred for the same items (and to the degree possible, the same-quality items) from one time period to the next. In contrast, the weights in the GDP deflator change every year; they are based on the composition of the GDP of the economy. The weights for the farm price indexes in the United States are based on 5-year moving averages. Maintaining up-to-date indexes is difficult, especially if governments do not provide adequate resources for their construction, but it should be noted that the U.S. Bureau of Labor Statistics has and is working to address issues associated with weighting methods.

Measures of rates of inflation affect economic policy, and as a consequence, there has been considerable debate about possible upward biases in the CPI (e.g., Boskin et al. 1998; Diewert 1998). There are two broad problems with price indexes that use fixed weights. First, the weights reflect consumer spending patterns in an earlier period. The weights are updated periodically, but inevitably they are based on historical experience. In contrast, changes in relative individual prices cause consumers to shift their spending patterns. For example, if the price of chicken declines relative to price of beef, consumers tend to substitute chicken for beef, but the weights for these items are fixed. Thus, the index tends to overestimate the cost of living.[4] A related potential problem is sampling the increased diversity of outlets; a TV set may be available for $500 in a local appliance store but may be only $450 in a large discount store or from Internet sources. As alternative outlets have become increasingly available to consumers through time, the proportion of purchases from them has increased; thus, there is the question of substitution among outlets for the same product.

A second problem area—and one that received much attention in the 1990s—is quality change and new-product bias. With research and development, new products become available, and product quality improves. When a new product is introduced, such as videocassette recorders (VCRs) and microwaves in past years, it may be an entire decade before the product is introduced into the CPI

4. O'Brien-Place and Tomek (1983) estimated that least-cost diets, which permit the substitution of cheaper for more expensive foods, grew 6.6–7.4 percent from 1970 to 1980, and this contrasted with a growth of 8.3 percent in the food-eaten-at-home component of the CPI. In other words, the growth in least-cost solutions were 10–20 percent less than the CPI rate. Boskin et al. (1998) reported a substitution bias for the overall CPI of 0.4 percentage point per year.

(Boskin et al. 1998: 8). It is also true that the quality of products change; the personal computer today is far more powerful than the one in use 25 years ago, or even last year. And automobiles are more durable now than they were in the 1970s. It is very difficult to make adjustments for the changing quality of a product in price indexes.[5] Even if precisely the same product definition is used (e.g., a four-door sedan, with a specific set of attributes), the intrinsic quality, as measured by durability, may still have improved. Hence, it is likely that the CPI, as well as price indexes in other countries, is biased upward, i.e., exaggerates the rate of inflation.

When the weights used to construct price indexes are changed only periodically, users must keep in mind that, the more remote the period used to establish weights, the less reliable the index is likely to be as an indicator of the current rate of inflation. Of course, the indexes with outdated weights can still be accurate measures of the average change in prices for the particular market basket of goods and services that is being covered by the index.

## The Terms of Trade of Farm Products

### Definition and Measurement

Relative prices may be referred to as the *terms of trade*. In this book, we are primarily concerned with changes in prices received by farmers for the products they sell relative to the prices paid by farmers for farm inputs. An improvement in the terms of trade occurs when the prices received rise relative to those paid, and the terms of trade deteriorate when the reverse occurs.[6]

Indexes of prices received by farmers and those paid by farmers are plotted in the top portion of Figure 10-1, and their ratio is provided in the lower part of the figure; the plots are for the years 1970 through 2012. If both index numbers had risen (or declined) at the same rate, the ratio would have remained constant. It did not do so because the two series changed at different rates. The ratio rose from 1970 through 1973 and then tended to decline thereafter. Both indexes use 1990–92 = 100 as the base, and the same pattern of change would be observed regardless of the years selected for the base period (= 100).

5. The United States devotes considerable resources to obtaining an accurate CPI. The Bureau of Labor Statistics collects between 70,000 and 80,000 price quotations for over 400 goods and services at about 22,000 retail outlets, either monthly or bimonthly, and additional information is collected about rent and owners' equivalent rent. About once a decade, the weights are changed based on consumer expenditure surveys (Boskin et al. 1998). But given the complexity of the U.S. economy, it is not surprising that it is difficult to provide quick accurate adjustments for the entry of new products and changes in quality.

6. The phrase *terms of trade* traditionally refers to the price ratio of items exported to those imported. When the terms of trade improve, this usually is taken to mean that export prices have risen relative to import prices. It also should be noted that the ratio of farm to non-farm prices is technically not the same as the ratio of price received to prices paid by farmers. Prices paid by farmers include some inputs of farm origin, such as animal feed and heifers entering beef cow herds. For convenience, we use the index of prices paid by farmers.

The magnitude of the ratio is, however, a function of the years selected as the base period. This point is illustrated by the data in Table 10-2, which shows the ratios (prices received/prices paid) for the year 2005 using three different base periods: 1980, 1990–92, and 2010. The ratio increases as the base period is shifted from 1980 to 2010. This is a function of the movement of relative prices during this particular time period; i.e., U.S. farmers faced relatively less favorable prices over most of the period in our example. If one examines the ratio of prices received to prices paid as depicted in Figure 10-1, it is clear that the ratio is lower in 1990–92 than in 1980 and that it continued generally downward through 2005. It moved up slightly thereafter but still was less favorable in 2010 than in 1990–92 or 1980. The less favorable the relationship in the base year, the larger the ratio will be in the current year. Thus, the ratio for the year 2005 increases as the base is moved from 1980 to 2010. If we had searched for a base period in which the ratio of prices received to those paid was higher, then the 2005 ratio would have been smaller. The logic is simply that the magnitude of the ratio for any individual year depends on how it compares with the base year (in which the ratio is always 1).

One message from Table 10-2 is that a user must be careful in interpreting the magnitude of a particular ratio of indexes of prices as a measure of the welfare of farmers. One may imply that farmers are faring relatively well or, alternatively, are doing poorly simply by the choice of the base period. Likewise, the values of a deflated price or income series are dependent on the base period used in the index; the deflated series can be made arbitrarily larger or smaller by the choice of the base period. Changing base years, however, does not affect the relative changes (percentage increases or decreases) through time; it only affects the magnitude of the index. In other words, the ratio plotted in Figure 10-1 would have the same shape whatever base period was chosen, but the base period would influence the absolute magnitudes of the ratio.

In any case, changes in relative prices are not necessarily an accurate indicator of changes in real income. A 10 percent decline in the ratio, for example, will approximate the actual percentage decline in farmers' real income only if efficiency of production has not changed (i.e., if output per unit of input remains unchanged). If productivity increases by 10 percent between the base year and

**Table 10-2.** Ratio of prices received to prices paid by farmers in 2005, United States, calculated for different base periods

| Base period | 2005 ratio |
|---|---|
| 1980 | 0.614 |
| 1990–92 | 0.803 |
| 2010 | 1.042 |

*Source:* Calculated from data in the *Economic Report of the President* (2013).

the current year, farmers as a group will be as well off despite the 10 percent decline in the ratio of prices received to prices paid. As discussed in the next subsection, one possible reason for a decline in prices received is that improved productivity has shifted the supply function relative to the demand function. It is nonetheless true that the terms of trade for farmers are used as a measure of well-being of producers, which is why it is important to understand the limitations in estimating these indexes.

### Causes of Changes in Relative Prices

There is much debate and a considerable literature (and many hypotheses) about the reasons for changes in the prices of farm products relative to changes in the prices of non-farm goods and services. In this literature, which goes back many years, one should distinguish between reasons for differences in the short-run variability of the two types of prices and reasons for differences in the trends in these prices.

Those who have read the previous chapters of this book should not be surprised that agricultural product prices tend to be more variable than many other prices, although the averages of a group of prices (indexes) are less variable than the individual prices. This is sometimes called the fixed-flex paradigm (Andrews and Rausser 1986: 414–415). Prices of industrial goods and services tend to be sticky or inflexible, at least in the short run. Many non-farm prices are based on administrative decisions made by individual firms. A decline in demand is often met by a reduction in output with prices remaining relative stable. In contrast, the prices of many farm products are determined by (or are indirectly influenced by) competitive bidding in auction markets (Chapter 11).

Considerable empirical evidence supports the hypothesis that price flexibility (or inflexibility) is related to the type of market structure (competitive versus monopolistic or oligopolistic). In a classic study, Gardiner Means (1935) found that prices fell less between 1929 and 1931 (the Great Depression period) in relatively concentrated industries than in less-concentrated industries. Another example is that between 1979 and 1982 the agricultural implement industry reduced its output 41 percent in response to a sharp decline in demand (Andrews and Rausser 1986: 416) but implement prices continued to rise, even after they were adjusted for the general effects of inflation. The contrast with agricultural prices is striking. During the same 3-year period, farm prices in real terms fell 27 percent. Despite the decline in real prices, total farm output continued to rise, although at a slower rate than earlier.

The existence of fixed and flex price market sectors may also result in short-run "overshooting" of prices in flex price sectors in response to changes in monetary variables. (Financial markets, as well as agricultural markets, have flexible prices.) When some prices respond slowly to a monetary change, one hypothesis is that flexible prices bear the burden of short-run adjustments. Using monthly prices for the United States from 1975.1 (1975, first quarter) to 1999.3, Saghaian, Reed, and Marchant (2002) use a measure of the money supply to represent the monetary effect and find empirical support for the overshooting hypothesis. The variability of prices received by farmers apparently depends not only on real supply and demand factors (which are, of course, influenced by economic policy) but also on the effects of the nominal money supply.

Looking for trends (rather than month-to-month changes), one can find periods where prices received by farmers in the United States have declined relative to the prices they paid for inputs. This can be viewed as a consequence of supply shifting to the right faster than demand. A stream of new technology (that is profitable to use) has entered agriculture especially since 1945. Demand has also increased, but it is possible for the growth in supply to exceed the growth in demand for a considerable period of time. Since 2005, however, the prices received by farmers have increased relative to prices paid as a consequence of sharply higher demands for commodities.

If longer time periods are examined (say 100 or more years), then it is difficult to ascertain whether relative prices are trending. Aside from the problems of measuring price levels over such long periods, it is clear that the trends—changes in the mean of the ratio—are small relative to the variance. In other words, the trend, if any, that is found depends on the time period selected to measure the trend. Indeed, if the reader were to examine Figure 10-1 as it appeared in the first edition of this book (showing relative prices from 1900 through 1970), no obvious trend would exist, even though relative prices both rose and fell over that 70-year period. One can find sub-periods with upward trends, downward trends, or no trend.

Prebisch (1959, 1964; cited in Deaton 1999) argued that commodity prices will fall relative to the prices of manufacturers, and his argument is based partly on differences in the income elasticities of demand for the two sectors. That is, the income elasticity for farm products is thought to be lower than for manufactured products. Thus, growth in aggregate incomes will increase the demand for manufactured goods faster than the demand for farm products. Implicit in this argument is the assumption that supply increases in the two sectors are similar.

According to Deaton (1999), Arthur Lewis (1954) provided a better theoretical account of why agricultural product prices declined relative to other prices in the last part of the twentieth century. Lewis's argument, in contrast to Prebisch's, is based on supply-side reasoning. The farm sector in most parts of the world has continuously had an "excess" supply of labor because, as new technologies have entered agriculture, labor needs to shift from farming to other employment. Consequently, returns to labor and management in farming tend to be persistently low. Put another way, persistently low product prices, resulting from new technologies, provide incentives for labor to move out of farming; this process is likely to continue as long as this resource reallocation is required by market forces. Recent empirical evidence, however, does not support this hypothesis of deteriorating terms of trade for farm commodities (Balagtas and Holt 2009).

A general conclusion is that changes in aggregate price levels will depend on shifts in aggregate demand and supply, and on how these shifts relate to the general economy. Hence, any forecast of relative price changes requires knowledge about relative changes in aggregate supply and demand. It is clear that the world population is going to continue to grow during the first half of the twenty-first century, although probably at a slower rate (about half) than in the year 2000. Also, real incomes will probably grow, but the income elasticity of demand for agricultural products in the aggregate is low, perhaps 0.2. Nonetheless, growing population and income do suggest an upward trend in demand. Thus, the key unanswered, and probably unanswerable, question is, what will be the rate of growth of agricultural supply? This, in turn, depends in important ways on the availability of new, profitable-to-use technologies, and it is extremely difficult to forecast this rate of change. In 1998, Tweeten suggested that supply growth appeared to be slowing relative to demand and, if so, commodity prices would strengthen relative to other prices; and indeed many commodity prices have risen sharply since 2005.

## Concluding Remarks

Whether one is considering year-to-year changes or longer-run trends in the average level of agricultural product prices, they are difficult to forecast. Price forecasts require ancillary forecasts of changes in supply and demand, and prices are influenced by shocks, like unexpected political decisions and weather events. These can modify, at least in the short run, the effects of systematic trends in income, population, and technology. Moreover, these trends themselves cannot be projected with precision (although this should not stop analysts from trying to improve their estimates).

While it may not be possible to forecast the timing, magnitude, and direction of changes in the average level of farm prices, one can be reasonably certain that instability in the terms of trade will persist. Fluctuating prices are the norm rather than the exception for most of the agricultural sector. Consequently, farmers and other agribusiness firms are likely to maintain a strong interest in strategies and public policies that are designed to manage price risks or to compensate for unstable terms of trade.

## Appendix: Price Indexes

Price-level movements are measured by index numbers. That is, index numbers are used to measure average changes in prices (or other variables) at some point in time relative to a base period. A specific index number, such as the CPI in the United States, provides an empirical measure of a general concept, such as the level of retail prices paid by consumers (i.e., their cost of living). Clearly, any empirical measure is not going to be a perfect measure of the true concept. This appendix reviews selected methods of constructing index numbers, some problems encountered in computing such numbers, and the interpretation and use of price indexes. Readers interested in a relatively non-technical discussion of the underlying conceptual issues may consult Diewert (1998).

## Constructing an Index

In its simplest form, an index number may be thought of as a ratio. The denominator of the ratio contains the base period observation, and the numerator contains the current observation. The resulting ratio measures current observations as a percentage of the base period. The ratio of the current price of a particular good (or service) to a base period is the *price relative*.

$$R_{ji} = P_{ji} / P_{0i},$$

where $P_{ji}$ is the price in the $j$th (say, current) time period for the $i$th product and $P_{0i}$ is the price in the base time period (0) for the $i$th product. As time passes ($j$ changes), the price in the numerator changes, and a series of price relatives is obtained for the given base-period price.

An index number can be constructed as an average of price relatives. The arithmetic mean for a given date, $j$, simply sums the individual price relatives, $R_{ji}$, over the $i = 1, 2, \ldots, n$ items to be included in the index and divides the sum by $n$. For example, in 2010 the average price received by U.S. farmers for all beef cattle was \$92.00 per cwt.; for hogs, \$55.10 per cwt.; for broilers, \$0.493 per lb.; and for turkeys \$0.612 per lb. In 2012, the comparable prices were all higher: \$122.00, \$64.20, \$0.523, and \$0.719, respectively. To compute an index for 2012, with 2010 as the base period, the price relatives can be averaged.

$$\begin{aligned} I_{2012} &= \frac{\left(\frac{122.00}{92.00} + \frac{64.20}{55.10} + \frac{0.523}{0.493} + \frac{0.719}{0.612}\right)}{4} \\ &= \frac{1.326 + 1.165 + 1.061 + 1.175}{4} = 1.182. \end{aligned}$$

It is common practice to multiply the index by 100 and define the index, in this example, as 118.2. That is, from the price relatives, we note that beef cattle prices rose 32.6 percent, hog prices increased 16.5 percent, broiler prices increased 6.1 percent, and turkey prices increased 17.5 percent from 2010 to 2012, and a simple average of the changes indicates that meat prices increased 18.2 percent relative to the 2010 base period (index = 100).

Other averages such as the geometric or harmonic mean could be used. These generally give different numerical results. A more important problem, however, is that simple averages give each item in the index equal weight. In the meat animal example, turkeys received equal weight with beef cattle, but in the United States, turkey marketings are smaller than cattle marketings. The solution to this problem is to develop a weighting scheme that measures the relative importance of the various items included in the index.

For a prices-received-by-farmers index, a potential weighting method is to use percentages based on the value of the marketings of each commodity in the index relative to the total value of marketings. For a prices-paid index, weights can be based on the importance of the individual inputs relative to the total cost. A common approach is to base the weights on a particular historical time period. The time period used for computing the weights can be the same as the base period for prices, but it need not be.

Fixed weights for a farm-prices-received index may be defined as follows. The weight for one commodity (the *i*th) in an index based on *n* commodities is

$$W_{0i} = \frac{V_{0i}}{V_{01} + V_{02} + \cdots + V_{0i} + \cdots + V_{0n}},$$

where $V_{0i} = P_{0i}Q_{0i}$ is the value of marketings of the *i*th commodity in the base period 0, and the denominator is the sum of marketings for all *n* commodities in the index. Weights are obtained for each of the *n* products in the index, and by definition the sum of the weights, $i = 1, 2, \ldots, n$, is 1.

An index is then computed for a particular time period (the *j*th), say, in months or years, as a weighted arithmetic mean of price relatives:

$$I_j = W_{01}R_{j1} + W_{02}R_{j2} + \cdots + W_{0n}R_{jn},$$

recalling $R_{j1} = P_{j1}/P_{j0}$, and so on. For example, the simple average of price relatives for the four livestock price relatives, computed above, could be modified using weights:

$$I_{2012} = (0.4)(1.326) + (0.3)(1.165) + (0.2)(1.061) + (0.1)(1.175) = 1.2096.$$

The weights used in the example are hypothetical, but they illustrate that, if the weights are changed from being implicitly equal to giving cattle a larger weight (0.4) and turkeys a smaller weight (0.1), the resulting index number changes from 118.17 to 120.96.

The foregoing equation is one way of defining the *Laspeyres index*, which uses quantities ($Q_{0i}$) in a base period as the weights. To see this, recall that

$$W_{0i} = \frac{V_{0i}}{\sum V_{0i}} = \frac{P_{0i}Q_{0i}}{\sum P_{0i}Q_{0i}}, \quad \text{where } \sum_{i=1}^{n} = \text{sum over } i = 1 \text{ to } n.$$

Then substitute for $W_{0i}$ in the definition of $I_j$ as follows:

$$I_j = \sum_{i=1}^{n} \frac{P_{0i}Q_{0i}}{\sum P_{0i}Q_{0i}} \frac{P_{ji}}{P_{0i}} = \frac{\sum Q_{0i}P_{ji}}{\sum Q_{0i}P_{0i}}.$$

In this equation, only the $P_{ji}$ changes with the passage of time, $j = 0, 1, 2, \ldots, J$. The preceding equation is also defined as a *weighted aggregative price index*.

There are other, alternative weighting systems for index numbers. The *Paasche index* replaces the base-period weights ($Q_{0i}$) with current-year weights ($Q_{ji}$). It is difficult, however, to obtain and maintain current-quantity weights, especially on a monthly basis. As noted elsewhere, the prices-paid and prices-received indexes for farmers in the United States now use a moving 5-year average of weights. The weights change gradually and are based on historical data. In the past, these indexes used fixed weights, but the weights were updated from time to time and often were not based on the same years as the base period for prices.

*Fisher*'s *ideal index* is the square root of the product of the Laspeyres and Paasche indexes. Many indexes in the United States and in other countries use a Laspeyres or a modified Laspeyres formula. The GDP deflator is now computed as a Fisher's ideal index.

## Problems in Constructing Indexes

The problems of index number construction can be classified under three headings: (1) selecting the components of the index, (2) choosing the base period, and (3) choosing the weights.

*Index Components.* As mentioned previously, an index is constructed to represent a particular concept, but the cost of collecting data prohibits exhaustive coverage. The data must be obtained through some type of sample. A CPI is intended to measure the cost of living. In the United States, the Bureau of Labor Statistics estimates that its CPI for all Urban Consumers (CPI-U) covers prices for 87 percent of the total population. Prices are collected for a market basket of goods and services involving thousands of individual observations on prices (see footnote 5). But, as noted earlier in the chapter, a dynamic economy involves new products entering the market, quality changes in existing products, and products becoming less important and perhaps disappearing from the market.

Thus, the analyst has the problem of how to select items to include in the index and how to introduce new products and remove old ones. In the CPI of the United States, prices are obtained for precisely defined items (in terms of attributes) for which observations are readily available. A procedure also exists to introduce new items, but this can take a considerable length of time. The sampling problem is complicated by the numerous outlets selling the items and the changing character of these outlets. In addition, the posted or listed prices may not be the actual purchase price paid by the consumer. Auto dealers routinely give discounts. Grocery stores have "membership clubs" in which members receive discounts from the posted prices.

To compute the Index of Prices Received by Farmers, the USDA obtains the average price received by farmers for each commodity each month. No quality specifications are contained in the instructions. Since quality may vary, the index can change because the average quality of the product sold changes. This emphasizes that indexes, by the selection of the data, can measure quite different concepts (either intentionally or unintentionally). The intent of the prices-received index is to use prices that, when multiplied by quantities sold, give the total revenue of farmers. Thus, the prices used in this index are consistent with the concept being estimated.

*Base Period.* The selection of a base period is rather arbitrary. The main criterion is to have an up-to-date base that is not too far removed from the current period. This allows the base-period prices, the $P_{0i}$, to represent a recent product mix and quality. The base period is also sometimes thought of as a time of "normal" prices, but the definition of *normal* is a matter of judgment. Although the farm price indexes in the United States use updated base periods, they still also include an index with a 1910–14 base because this is a period of relatively favorable prices for farmers. Note that the $P_{0i}$ need not be a single year's price but can be an average of prices over a period of years. At the time of writing this book, the Index of Prices Received and Paid by Farmers in the United States used a base period for prices of 1990–92. The previous base for prices was 1977.

*Weights.* The quantity (value) weights need not come from the same time period as the price base. When the base periods for the prices-received and prices-paid indexes in the United States were shifted to 1990–92, the weighting scheme was changed and is based on 5-year moving averages.[7] Thus, these indexes are modified Laspeyres formulas; the weights change slightly each year. The CPI and PPIs are also Laspeyres-type indexes, with fixed weights but with procedures to update the weights. In contrast, since 1995 the GDP deflator has been computed as a Fisher's ideal index. It is called a chain-type annually weighted index.

In computing a farm-prices-received index, the weights could be based on the quantities produced or quantities marketed. Quantities marketed are typically used. In a country like the United States, the quantity produced and the

7. The previous version of the index used weights from 1971–73 with a price base of 1977. Thus, with the passage of time, the years used for the price base and for the weights can change.

quantity marketed are nearly identical. In a developing country, a large proportion of the food produced is often consumed on farms and not sold. In this situation, quite different weights would be obtained using the quantity produced than using the quantity marketed.

The quality (accuracy) of an index in representing a concept like the cost of living depends in important ways on the quality of the data going into the index and hence on the amount of resources devoted to obtaining accurate price and quantity information. Personal judgments inevitably enter into decision making about index number construction, but this does not necessarily mean that index numbers are biased. Government agencies, like the USDA, try to make index numbers as representative of the price or production series they are trying to measure as is possible within the limits of their financial resources.

## Possible Biases in Index Numbers

Index numbers are necessarily imperfect measures of the underlying concept. Any broad-gauged index must use a sample of the population of prices, and thus a potential problem is the quality of the sample. Assuming a representative sample of prices is obtained at any point in time, the principal sources of bias are the weights used to measure the relative importance and possible changes in the quality of the items included in the index.

Alternative weighting systems will give different estimates of the percentage changes in prices. The use of constant weights (Laspeyres-type index) generally leads to an upward bias in an index of consumer prices (see footnote 4). This occurs because consumers tend to purchase less of items that have risen most in price since the base year. Consumers are substituting relatively less expensive goods for the relatively more expensive goods. Consequently, the prices of items that have risen most in price tend to be over-weighted by the historical base-period weights.

The use of current-period weights (Paasche-type index) does reflect the current consumption pattern (or for a producer index, the current marketing pattern), and for a consumer index, it reflects the current cost of living relative to having consumed the same combination of goods in the base period. A problem, as already noted, is the cost of revising the weights every year. Also, if both prices and weights differ in years other than the base year, then this raises questions about comparisons with years that are not the base year.

Perhaps the major source of upward bias in consumer or prices-paid indexes is related to the issue of including quality changes in the items in the index as well as the problem of adding new products and deleting outdated products. A new-model tractor or automobile, for example, may cost more than an earlier model, but it also may incorporate technological improvements that are not easily specified in the product's definition. The product definition may include a variety of attributes (horsepower of the engine, model type, accessories, etc.) but still miss an improved computer module, improved materials, or other quality-enhancing changes. Or a new product is introduced that, say, improves the productivity of the farmer, but it may be years before it enters the index. The typical effect of these omissions is that the price index is biased upward. It shows a larger increase in prices than the true rate of inflation.

## Uses of Price Indexes

The major use of a price index is simply to describe the average price movement of a combination of goods and services with the passage of time relative to the base period when $I_0 = 100$ (or 1.0). A current index number of 180.7 means that prices on average are 80.7 percent higher than those in the base period. For a Laspeyres index, it also is valid to compare adjacent years, such as the values 180.7 and 185.0. The period-to-period change in this case is 4.3 points, or 2.4 percent.

For a Paasche index, the current period and the base period have the same weights and are comparable. But since the weights change each year, adjacent time periods will have different weights for the prices. Thus, the index has changed because both the prices and the quantities have changed. Therefore, if one wants to compare how the same quantities are priced in different years, neither of which is the base period, then a problem arises in interpretation.

The current indexes of prices paid and received by farmers in the United States represent an intermediate case. The weights are based on lagged 5-year averages. Thus, the current year is comparable to the base period because the same weights are being used for both sets of prices. The adjacent years, neither of which is the base period, will have similar but not identical weights.

A problem encountered in using index numbers is to interpret them correctly. If carefully constructed, a price index is likely to provide a reasonable approximation of the underlying concept. But the selection of alternative base periods for purposes of comparison can lead to different conclusions about relative price movements and the welfare of different groups. In the 1940s and again in the early 1970s, average farm prices rose relative to average consumer prices. The opposite was true in the 1950s and in most of the 1990s. Thus, selecting 1940 as a base year (just before the relatively favorable change in farm prices) would tend to make farmers look well off in subsequent years. In contrast, if the early 1970s were selected as a base period, then farmers would not look well off in subsequent years. (The same idea is illustrated in Table 10-2.)

Sometimes monthly price indexes are seasonally adjusted. Prices are assumed to have a "normal" seasonal pattern within the year; for this reason alone, a monthly index will move up and down. The seasonal adjustment attempts to remove the effect of the seasonal price pattern. The appropriate comparison in a seasonally adjusted index is between the same month in different years. Comparisons between adjacent months within a year can be confusing (although not necessarily wrong, if properly understood). The *unadjusted* index, for example, can rise from one month to the next, but if the price rise is less than the normal seasonal rise, then the seasonally *adjusted* index will decline. Thus, analysts must ascertain whether they are using adjusted or unadjusted indexes and interpret them accordingly.

In addition to their descriptive functions, price indexes are used to “deflate” various price and income series. The deflated series is obtained simply by dividing a price or income series by appropriate index numbers. The resulting price series indicates how the individual price has changed relative to the denominator, the price index. For instance, if the retail price of pork remains constant while the CPI rises, the “real,” or deflated, price of pork falls. The same idea applies to calculating incomes in “constant dollars.” The real per capita income of consumers is obtained by dividing the current nominal income by the CPI. The nominal and real incomes will be identical for the year in which the index equals 100 (1.0). Since general price indexes tend to trend upward, shifting the base period to a more recent time period makes the index numbers smaller. Dividing a given nominal series by smaller numbers makes the ratios larger. Thus, the magnitudes of the deflated series will depend on the base period used for the index.

Some of the advantages of and problems with using deflated rather than nominal data in empirical analyses are discussed in Chapters 14 and 15. We note here, however, that the CPI is used rather routinely as a deflator, and this is sometimes inappropriate. The CPI is intended to measure the movement of retail prices; the CPI is far from perfectly correlated with other measures of different groups of prices. In constructing a series of deflated prices, it is important to use a deflator that is internally consistent with the intended use of the deflated price series.

Finally, perhaps the most important use of price indexes, at least to many people, is their use as cost-of-living adjustments (COLAs) for social security retirement payments and pension plans. Currently, in the United States, the federal government uses the CPI for Urban Wage Earners and Clerical Workers (CPI-W) to make adjustments in social security payments to retired individuals. Specifically, the average percentage change in the CPI-W from the third calendar quarter of one year to the third quarter of the next year is used to adjust social security benefits (starting in the subsequent January). Many other pension plans make a similar use of price indexes for adjusting pension benefits.

As of 2013, debate was continuing about the use of the CPI-W as a basis for increasing social security benefits; it allegedly overstates the true price increases faced by retirees because it ignores the substitution effect discussed above and perhaps because the weights on items purchased are not consistent with retiree’s purchase patterns. The CPI-U is the most commonly reported index of prices because it has broader coverage for “all” urban consumers; still another variant is the chained CPI-U, which addresses the issue of weights becoming dated. The choice of price index can make a large difference in the long-run cost of pension programs, and therefore the choices generate much debate. For this and other reasons, it is useful to understand the concepts behind price indexes and the details of their construction.

## References


Andrews, M. S., and G. C. Rausser. 1986. "Some Political Economy Aspects of Macroeconomic Linkages with Agriculture," *Am. J. Ag. Econ.* 68:413–417.

Antle, J. M. 1999. "The New Economics of Agriculture," *Am. J. Ag. Econ.* 81:993–1010.

Balagtas, J. V., and M. T. Holt. 2009. "Commodity Terms of Trade, Unit Roots, and Nonlinear Alternatives: A Smooth Transition Approach," *Am. J. Ag. Econ.* 91:87–105.

Boskin, M. J., E. R. Dulberger, R. J. Gordon, Z. Griliches, and D. W. Jorgenson. 1998. "Consumer Prices, the Consumer Price Index, and the Cost of Living," *J. Econ. Persp.* 12(1): 3–26.

Cochrane, W. W. 1958. *Farm Prices: Myth and Reality*. Minneapolis, MN: Univ. of Minnesota Press.

Deaton, A. 1999. "Commodity Prices and Growth in Africa," *J. Econ. Persp.* 13(3):23–40.

Diewert, W. E. 1998. "Index Number Issues in the Consumer Price Index," *J. Econ. Persp.* 12(1):47–58.

*Economic Report of the President*. 2013. http://www.whitehouse.gov/sites/default/files/docs/erp2013/full_2013_economic_report_of_the_president.pdf (accessed March 11, 2013).

Johnson, D. G. 2000. "Population, Food, and Knowledge," *Am. Econ. Rev.* 90:1–14.

Koech, J. 2012. "Hyperinflation in Zimbabwe," *Globalization and Monetary Policy Institute*, 2011 Annual Report, Federal Reserve Bank of Dallas. Pp. 2–12.

Lewis, W. A. 1954. "Economic Development with Unlimited Supplies of Labor," *Manchester School of Econ. & Soc. Studies* 22:139–191.

Means, G. 1935. *Industrial Prices and Their Relative Inflexibility*. Washington, DC: U.S. Govt. Printing Office.

O'Brien-Place, P. M., and W. G. Tomek. 1983. "Inflation in Food Prices as Measured by Least Cost Diets," *Am. J. Ag. Econ.* 65:781–784.

Prebisch, R. 1959. "International Trade and Payments in an Era of Coexistence: Commercial Policy in the Underdeveloped Countries," *Am. Econ. Rev.* 49(2 [proceedings]):251–273.

———. 1964. "The Economic Development of Latin American and Its Principal Problems," in *Leading Issues in Development Economics*, G. M. Meier, ed. New York: Oxford Univ. Press. Pp. 339–343.

Saghaian, S. H., M. R. Reed, and M. A. Marchant. 2002. "Monetary Impacts and Overshooting of Agricultural Prices in an Open Economy," *Am. J. Ag. Econ.* 84:90–103.

Tweeten, L. 1998. "Anticipating a Tighter Global Food Supply-Demand Balance in the Twenty-First Century," *Choices, The Magazine of Food, Farm, and Resource Issues*. Third Quarter: 8–12.

USDA. 2000. *Agricultural Prices 1999 Summary*. Nat. Ag. Stat Ser. Report Pr1-3(00)a. Washington, DC: NASS USDA. Available from http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1003.

PART III

# PRICING INSTITUTIONS

The institutional arrangements for establishing farm prices, sometimes called *price discovery mechanisms,* are described in this part. Chapter 11 provides an overview of the principal mechanisms used to price farm products; price discovery methods are categorized and their economic consequences are discussed. Chapter 11 also addresses issues related to government intervention in pricing farm commodities. The mechanics and models of commodity futures markets as pricing institutions are discussed in detail in Chapters 12 and 13. These chapters link the behavior of prices for futures contracts to the behavior of cash prices and demonstrate how this information can be used to make hedging decisions.

CHAPTER **11**

# Mechanisms for Discovering Prices

The principal mechanisms used to establish, or "discover," prices and their economic effects are discussed in this chapter. The term *price discovery* is used to denote the institutional mechanisms by which buyers and sellers arrive at specific prices and other terms of trade. It involves the mechanics of pricing rather than the theory of price determination (covered in Chapter 5). Price discovery mechanisms evolve in response to changes in the economy and may influence price behavior. Price discovery is not a costless process; pricing mechanisms can influence the quality of information and the way it is transferred; and alternative price discovery methods may influence how price risk is allocated among market participants.

The emphasis of this chapter is on pricing agricultural products at the farmer–first handler level, but the main price-making institution need not be located at the farm level. For example, a buyer of soybeans in Iowa is likely to make a bid based on a price quotation for the nearby futures contract, which is traded on the Chicago Board of Trade (with delivery, if made, on the Illinois waterway system). Thus, our discussion of pricing institutions is necessarily relatively general.

We also remind readers that economic activity is coordinated not only by markets but within firms. If a food processor vertically integrates back to the farm level—owns the resources used to produce the commodity—then the coordination of the supplying of the raw product to the processor occurs within the firm. Since this book is about markets and market prices, intra-firm coordination is not discussed. But some institutional arrangements for establishing prices provide tight coordination between producers and first handlers. In other words, price discovery results in a variety of outcomes, ranging from a simple implicit contract between two individuals to explicit written contracts with detailed coordinating provisions.

The terms *pricing mechanisms* and *pricing arrangements* are used here interchangeably to denote the institutions and methods used to price farm products. These pricing mechanisms are first categorized and described. Then, we discuss the potential economic effects of alternative pricing mechanisms, including different types of contracts. Finally, government intervention in pricing agricultural products is discussed.

## Alternative Pricing Mechanisms

### Price Discovery in Theory

In a perfect market, information about the factors determining prices is free and equally available to all participants. There are no costs of discovering prices, and hence prices reflect the existing economic conditions, and if new information arrives in the market, price changes instantly to the new, correct equilibrium level.

Economists have sometimes described how buyers and sellers could, in theory, arrive at "the correct" equilibrium price using a re-contracting process. This model assumes given supply and demand conditions (information) and a unique equilibrium price. In this context, buyers and sellers attempt to enter into contracts favorable to them, but when a tentative contract is made, both buyers and sellers reserve the right to re-contract if a more favorable price emerges. Information about prices is freely available through an impartial market reporter. Potential buyers may, for example, make bids at a price below the true, but unknown, equilibrium price, but as a consequence, they will find that the supply is not sufficient to meet the quantity demanded at the bid price. Hence, price is bid up, and at the higher price, a larger quantity is supplied and the initial sellers re-contract at the higher price. This search process continues until the quantity supplied equals the quantity demanded. Buyers pay and sellers receive the final equilibrium price.

In reality, re-contracting is uncommon and not available at zero cost. Nonetheless, the model illustrates two plausible points. First, price discovery can be viewed as a dynamic search process, but for any given information set, a certain amount of "noise" exists in the market. Not every trader will have exactly the same information nor necessarily evaluate the available data in the same way. Alternative price discovery mechanisms involve different ways of "searching." Second, the price discovery process can result in transactions prices tending toward equilibrium. Thus, when new information occurs, prices change as part of an adjustment process; prices search for a new equilibrium based on this new information set (as discussed presently in relationship to Figure 11-1).

It also important to recognize that, in practice, discovering the prices of individual lots of a commodity is potentially very complex. The pricing mechanism must discover both the overall price level, or reference price, and the particular attributes and conditions associated with each transaction that add or subtract from the reference price. As described in Chapters 7 and 8, commodities have attributes that vary from one lot to the next, and these attributes affect prices. Other attributes of the transaction, such as the timing of delivery and payments, also influence individual prices. For example, the transaction may convey an option to the seller about the timing of delivery; the seller may have, say, 5 days to make delivery.[1] Such an option in and of itself has value, and if the seller is receiving a benefit, the buyer is unlikely to convey this benefit without a compensating offset in the transaction price. In sum, for commodities, each transaction price need not be identical, even from a theoretical point of view.

## A Taxonomy

Transactions prices for agricultural products often involve a combination of institutional arrangements, and unlike the theory of price determination, economists do not have a fully agreed-on taxonomy of price discovery methods. It is, however, useful to categorize price discovery methods because this permits some generalizations about the benefits and costs of alternative mechanisms. We distinguish between the procedures or processes for arriving at prices and the outcomes of these procedures. A particular mechanism, like negotiation, can result in different outcomes (i.e., different contracts). Pricing arrangements can influence the costs of price discovery and how price risks are managed. So, pricing mechanisms do influence prices, at least indirectly through the outcomes that they generate.

We use a simple, three-category classification system:

1. negotiation, including private treaties negotiated between individuals and/or firms, and formal collective bargaining between, say, a group of farmers and a processor;
2. auctions, which can be of many different types and institutional arrangements, involving bids (buyers) and/or offers (sellers) at public marketplaces, on electronic exchanges, or via sealed bids; and
3. administrative pricing made by private firms or by public entities.

A negotiated price implies a bargaining process, which may be formal or informal. Auctions involve some type of bidding (and/or offering) procedure in a structured market framework. An administered price means that a decision maker has the ability to set or post a price. A variety of rules or procedures

1. This chapter is primarily about spot or cash prices, i.e., prices for current delivery. These prices should be distinguished from those for future delivery. Moreover, we need to differentiate between forward and futures contracts. Nevertheless, cash transactions may provide the seller with some latitude about the timing of delivery. The term *spot price* is sometimes taken to mean the price for immediate delivery, while the term *cash price* implies some latitude in the timing of delivery. But the two terms are often used interchangeably.

might be used in setting the current price, all of which in some way take account of existing economic conditions.

All methods of price discovery, therefore, require information, and we should keep in mind that prior prices are one type of information. Price changes, by definition, depend on a prior price. Alternative pricing arrangements may convey information in different ways. Also, markets for different products may contain different amounts of information and have differing proportions of informed traders at any point in time. This may be a function of the pricing method and the market structure.

The foregoing categories are broad. There are, for example, many different types of auctions, and the conceptual and experimental literature suggests that varying the auction rules can result in different pricing outcomes (Milgrom 1989; Klemperer 2002). Likewise, a negotiation process may involve differing degrees of market power for buyers and sellers and can result in various kinds of contracts. In particular, a contract may specify that the transaction price will be determined by a formula. The price received by a farmer equals, say, a base price defined in the contract plus or minus amounts that reflect quality, quantity, and location differentials; timing of delivery; and other attributes. Here the formula is established by negotiation, but the actual price received by the producer varies with the "movers" specified in the contract. *Formula pricing* could be treated as a separate (fourth) category in the foregoing taxonomy, but we have chosen not to do so.[2]

## Negotiated Prices

Many prices for commodities around the world are still established by informal negotiation between individuals. One farmer may wish to buy alfalfa hay from another farmer, and they informally negotiate a price. Both individuals are likely to be informed about the current prices of hay in their region, and the buyer has perhaps inspected the hay to determine its quality. The negotiation also would cover other terms of trade such as who bears the cost of transporting the hay and the delivery timing. In this context, the time required to agree on a price is likely to be short.

Informal negotiation is potentially the least-cost method of price discovery for commodities that have varying quality from one lot to the next and that are being sold in a local market. In the alfalfa hay example, only one transaction is involved and the bargaining is not time-consuming. A decentralized pricing system can also work for selling more than one lot when the opportunity costs

2. Any formal contract can be viewed as having a formula, which defines the terms of trade, but contracts have highly variable provisions. For example, a farmer could enter into a contract to deliver the entire crop to a dealer at harvest time, with the dealer agreeing to pay the farmer an average market price. An equal number of bushels is priced each trading day for the length of time defined in the contract. This is a simple formula, but the price received by the farmer depends on forthcoming market prices (which are unknown at the time the contract is signed).

of negotiation are relatively low. If the alternative uses for family labor are limited, which may be true for small producers in both less-developed and developed countries, then the opportunity cost of spending half a day in a farmer's market or bazaar to sell produce is small. Also, by going to a central location, traders learn about the asking prices of other sellers. It is a place for transferring and acquiring information.

Negotiation includes formal, structured bargaining processes. Perhaps the best-known example is the organization of labor unions and the negotiation of wages and other benefits with employers. The result is a detailed contract for a specific time period that defines benefits and responsibilities. Dissatisfaction with commodity price levels has led farmers, in some cases, to form bargaining associations (or cooperatives) in an attempt to negotiate prices with buyers. Farmers hope to achieve results similar to those obtained by the more successful labor unions, but this usually has not occurred.

The ability of bargaining associations of producers to negotiate more favorable prices depends first and foremost on their ability to control a substantial proportion of the total supply. Labor unions can withhold their labor, and they also may be able to limit the substitution of new laborers during a strike. It is difficult for a group of farmers to do the same thing. Usually there is a free-rider problem. Basically, farm organizations are unable to control the actions of members and nonmembers alike. It is difficult to get 100 percent of producers to join, especially when they will probably have to agree to limit their sales. Nonmembers can sell all they produce without any restrictions from the underlying organization and may even be offered a price premium by handlers who are seeking to break the association or discourage farmers from joining it. Even those who initially sign up and agree to withhold supplies may weaken and sell when they discover others are doing so.

In the United States, farmer cooperatives have done relatively little formal bargaining, but as noted in Chapter 6, there is some evidence that as the role of cooperatives increases, *ceteris paribus,* marketing margins are reduced (Rogers and Petraglin 1994). This presumably arises from the competition that cooperatives provide in the marketplace rather than from negotiations.

There is, however, a growing use of formal contractual relationships between farmers and buyers, at least in the United States (MacDonald and Korb 2011). The terms of contracts are subject to negotiation, although the amount of time required to negotiate the terms of such contracts is unclear. For example, processors use contracts to procure some or all of the raw product needed for their plants, but these contracts appear to be rather standardized documents developed and offered by individual processors. Individual producers may have relatively little bargaining power in establishing contract terms, at least where they have few choices of buyers. Nonetheless, the contracts must provide sufficient incentives to attract production, and producers may have choices about to whom and how they will sell their output (e.g., slaughtered hogs). Buyers sometimes experiment with contract design to achieve their objectives (Unterschultz et al. 1998).

One of the problems associated with the growing use of private treaties (negotiations) between a producer and a buyer is the potential lack of public information about the associated prices. The increased use of contracts implies, however, that they have potential benefits for individual sellers and buyers. We have more to say about contracts in a later section.

## Auctions

Auctions are pricing institutions that establish prices by bidding and sometimes by simultaneous offers, using clear rules. Numerous kinds of auctions exist. One of the most common types is the English or ascending bid auction. The auctioneer starts with an initial, reservation price, and the price is bid up by the individuals present at the auction; when the auctioneer has ascertained that no one is willing to bid higher, the item is sold. The English auction sometimes leads to over-bidding, called the "winner's curse." In the Dutch or descending bid auction, the auctioneer announces a relatively high initial asking price, and assuming no bidder accepts the price, the auctioneer lowers the ask price. This process continues until a person present at the auction indicates that he or she will buy at the announced price. U.S. Treasury securities use this mechanism.

Many other types of auctions are possible, including sealed bidding. For example, in awarding a construction contract for a new bridge, a government advertises for bids and (typically) accepts the lowest bid. One type of auction that has been of interest to economists is the second-price, sealed bid auction, sometimes called a Vickery auction. Assuming the auction pertains to selling an object (or a right), the process involves submitting a sealed bid; the highest bid wins, but the winner pays a price equal to the second highest bid. This system may under certain circumstances result in the seller receiving a higher price (Milgrom 1989). Variants of this auction are used in eBay proxy trading.

Double-auction markets have buyers and sellers making simultaneous bids and offers, which is a stylized characterization of organized exchanges for futures contracts and securities. A feature of this auction is that prices are established using known rules and a central clearing entity. Sometimes this involves an explicit physical trading place or an electronic exchange. Buyers make bids while sellers make offers, and a trade occurs when a buyer accepts a seller's offer or when a seller accepts a buyer's bid. Auctions are the least-cost way to discover the value of a particular good or right in some situations. It is likely, for example, that auctions minimize transactions costs for trading in securities or futures contracts. Stocks and futures contracts are homogenous, and potential buyers and sellers know precisely what is being traded. Quality variation is not an issue. Moreover, the handling costs of these transactions are small; no physical commodity must be transported to a central location.

In contrast, centralized auction markets have declined for agricultural commodities where transporting, yarding (housing), and handling costs are large. Nonetheless, local auctions exist for cattle, hogs, and sheep because farmers still find it convenient (least costly) to sell a few animals, such as culled cows, at a nearby location. On-farm auctions are typically used to disperse animals and machinery when a farmer goes out of business. The least-cost solution is for potential buyers to come to the farm rather than to move the items to other locations.

Large central auctions also may persist when the benefits to potential buyers of physically inspecting individual lots and large availability of the commodity offset the costs of transporting the commodity to the central location. This can occur when the items have a highly variable quality; hence, some buyers think that it is important to inspect the items to determine their value. For example, auction markets persist for feeder cattle (younger animals being purchased to be placed in feedlots), partly because feeder cattle have varying-quality characteristics. Many feeder cattle are, however, also sold via individual negotiation, and this illustrates that more than one type of price discovery mechanism may be used for a given product. Many types of goods are sold at auction, such as used cars, machinery, equipment, and even rare works of art, but these items are also sold via private negotiation.

Oral auctions, like the English system, make price information instantly available to all participants at the marketplace. Central marketplaces also allow market reporters to gather price information at low cost. The information provided by publically reported prices is valuable, and auctions and central marketplaces make the collection and reporting of data relatively inexpensive.[3] Thus, a potential benefit to society of auction markets is the information that they provide. These prices are, for example, sometimes the reference prices used in pricing formulas. An additional benefit of auctions is that their rules can be designed to minimize cheating and defaults on transactions.

One kind of auction that is used in experimental economics, but not in the real world, is the Becker-DeGroot-Marschak (1964) auction, which has the useful property of being "demand revealing," meaning buyers reveal their true maximum willingness to pay (WTP). Experimental economists find this property to be useful when trying to determine how much buyers truly value a commodity that is not priced in a market, such as the value for open space or the cost of environmental contamination. In this auction, potential buyers are asked what their maximum WTP is within a range of outcomes, e.g., \$0 to \$10,

3. The Agricultural Market Service of the USDA collects and reports daily prices for numerous agricultural products. These data are intended to represent market transactions, but they are typically reported as ranges for selected qualities and locations. Prices of individual transactions are not available. Unlike securities markets, commodity prices have no opening, closing, high, or low prices. The National Agricultural Statistical Service of the USDA collects data that are used to obtain average monthly prices by state.

and then a price within that range is drawn randomly. All buyers receive the good if their bid is at or above the randomly drawn price, but they pay the randomly drawn price, not their bid amount. If their bid is below the random price, then buyers cannot purchase the item. This auction is demand revealing because it is in the best interest of potential buyers to bid their true WTP. This is true because, if the bid is successful, the item will be purchased at a price that is less than or equal to buyers' bids.

## Administered Prices

Administered prices, as the term implies, are prices set by the administrative decisions of firm managers or by government regulation or policy. Consumers typically face prices in retail stores that are set by store management. A grocery store selling many thousands of items needs an administrative system that minimizes what are sometimes termed "menu costs," i.e., the costs of changing prices. Clearly, it would be excessively costly for grocery stores to negotiate prices with their many customers. In setting prices, managers are influenced by economic factors, including the degree of competition, and by the firm's marketing and pricing strategies in its economic context, but the point is that prices are set by administrative decisions that are likely to involve rules developed by managers, not by negotiation or by an auction.

On the other hand, retailers of big-ticket items, like auto dealers, will negotiate prices. They also have posted (list) prices. Thus, the transaction price is likely to be the outcome of a combination of an administrative decision and negotiation.

Individual farmers typically cannot administer prices, i.e., set the price they wish to receive for their output. A necessary condition to do so is that their output must have some characteristics that differentiate it from alternatives. This could happen, for example, in the sale of produce in a farmers' market. But for bulk, relatively homogenous commodities, farmers have little, if any, discretion in price setting. Thus, as noted in previous sections, commodity prices tend to be discovered via negotiation or auction processes.

In some instances, the price received by a farmer is set (administered) by the buyer. There is a posted bid price, say, for number two yellow corn delivered to the particular buyer's location. In this case, the buyer can either be an independent handler or a producer cooperative or bargaining association. The farmer can decide to deliver at this posted price or not. The posted price, however, is likely to be derived from a futures price obtained via an auction process. So, the posted price varies from day to day, and the price received by the farmer is, in this sense, a combination of prices discovered via an auction and set by an administrative decision.

Agricultural product prices also can be influenced by governmental policies. Indeed, a wide variety of government programs exist throughout the world that can influence prices, and in this sense, administrative decisions influence farm prices. These are discussed in the last section of this chapter.

## Economic Consequences of Price Discovery Arrangements

One way to approach the topic of the economic consequences of alternative pricing mechanisms is to ask, what determines the structure of pricing arrangements and why do these arrangements change? The economic context in which markets exist is important in answering this question. What are the economic incentives influencing pricing arrangements?

Economic development is driven in important ways by technological change and improved human capital (i.e., education, experience, knowledge, etc. of the workforce). The results of economic development include larger per capita incomes and greater diversity and specialization in the economy. Such changes create incentives for changes in markets and pricing institutions. For example, firms that produce new, specialized products may require inputs with specialized attributes. Or because consumers demand pork with lower fat content, farmers must be provided with incentives to produce lower-fat pork.

Processors want to obtain inputs with the desired characteristics at low cost. Not only do firms need to know about aggregate supply and demand, but they also require information about the attributes of the individual lots of the inputs they are buying. Large firms also may be exposed to large risks. Again, they want coordinating mechanisms that help manage risk but at a low cost.

Hence, new types of pricing arrangements and coordinating mechanisms arise as a part of economic development and changing demands. It should be noted, however, that the older pricing mechanisms often remain; so, the diversity of pricing arrangements increases. Potatoes are produced and marketed by very large growers and by small organic farmers. These potatoes are marketed and priced in different ways. One size does not "fit all," and a particular commodity may be priced via a variety of arrangements.

Economists hypothesize that price discovery arrangements—indeed, coordination mechanisms in general—evolve to minimize transactions costs.[4] This concept has wide applicability if transaction costs are broadly defined. These costs include not only the direct costs of arriving at prices, such as management time, but also a host of other associated costs and benefits. These include the physical assembly and transportation of the product; the ability of both buyers and sellers to obtain and evaluate information relevant to the transaction; the ability to manage risks associated with procuring inputs, including price risk and assuring the quality (attributes) desired by the buyer; and so forth.

4. Coase (1937) is a basic reference. He was concerned about what determines the coordination decisions made within a firm and the decisions based on market prices.

Consequently, although negotiation, auctions, and administrative decisions remain the broad over-arching methods for discovering prices, the outcomes of these processes have changed and continue to change. For example, contractual arrangements continue to evolve in new and different ways and seem to be limited only by the ingenuity of the human mind. In the next subsections, we summarize some of these changes.

### Contracts

Formal contracts between sellers and buyers can be classified in several different ways. One way is to distinguish between *production* and *marketing* contracts. Production contracts, as the name implies, provide for detailed close coordination between the farmer and the buyer. The contract assigns ownership of the farmer's output to the buyer. The farmer is paid a (contracted) price or fee for providing labor, facilities and equipment, and some management. The contract typically specifies the quality and quantity attributes, and the buyer may provide inputs like a particular seed variety or the young chickens to be placed in growing houses. By providing such inputs, the buyer helps assure an output with the characteristics the buyer wants. The contract also has provisions about sharing price and production risks.

Hueth et al. (1999: 375f) indicate that production contracts play three roles: they introduce predictability into production systems, they allow market participants to share risk, and they help motivate performance. These benefits are increasingly important as greater emphasis is placed on the need to certify the quality attributes of commodities as they move through the marketing chain (Sexton 2013). Contract performance, however, depends on incentives and monitoring. In other words, trade-offs exist between the benefits of contracting and the costs of assuring performance on the contracts.

A marketing contract pertains only to marketing-pricing provisions. For example, the farmer may agree to deliver a specified quantity to a buyer, and the buyer agrees to pay a price that is defined by a formula written into the contract. Many pricing formulas are possible. The price paid to the farmer might equal an average computed over a specified time period. Another contract might permit the farmer to speculate on prices, say, by allowing the farmer to name the date on which price is to be established. For example, the farmer could deliver corn in November but contract to establish the price on any date between November 15 and June 30, where the price is defined to equal the July futures price minus a fixed discount defined in the contract. Or the price may be set at planting time or shortly thereafter, with the price based on a futures contract price quote; again, this is feasible because the buyer can lay off the price risk by hedging using the relevant futures contract.

Marketing contracts also may have detailed formulas to establish price. For example, one type of contract for selling grain-fed cattle uses a grid pricing formula. The details of the formula can vary among contracts. In a contract analyzed by Schroeder and Graff (2000), the base price was obtained from a USDA quote for Choice Yield grade 3 fed cattle, and premiums and discounts were specified for other yield grades and for certain other attributes. The pricing system does, of course, require the slaughter of the animals and the evaluation of the carcass. An objective of a grid pricing system is to provide for the better transfer of information to cattle feeders about the value of quality attributes than does pricing live animals. Ultimately, this should provide cattle feeders with incentives to produce animals with more desirable attributes. Contracts for milk similarly include a base price plus or minus various premiums and discounts based on milk quality (e.g., the somatic cell counts), fat content, protein content, other milk components, and locational considerations.

Contracts have varying degrees of forward pricing. Thus, distinctions are made among spot, forward, and futures prices.[5] Spot transactions involve "immediate" delivery (see footnote 1). Forward contracts imply that the terms of trade are established at the current time, *t*, but delivery occurs at a later time, *T*. Forward contracts are established by negotiation, and delivery is expected to occur. The terms may vary from one contract to another, and the contracts are not fungible (tradable).

Forward contracts assure the buyer of the delivery of a given quantity (subject, however, to yield risk) and perhaps of a specified quality. In addition to the costs of negotiation and the lack of fungibility, forward contracts are subject to default risk and the contract provisions may contribute to default risk (e.g., Lence and Hayenga 2001). As previously noted, buyers extending forward contracts to farmers use futures and options markets to hedge the price risk that they acquire by making the forward contract.

A distinction must be made between forward and futures contracts. Futures contracts are discussed in Chapters 12 and 13, but we note here that they too permit delivery in the future. Futures contracts are standardized instruments; the contract terms are not negotiated. They are traded via a double auction markets, and each trade establishes a price and the quantity of contracts traded. Thus, in contrast to forward contracts, futures contracts are fungible, and initial positions can be offset. If a contract for December delivery is sold the prior May, this legal obligation to deliver can be offset by the subsequent purchase of the December contract. The risk of default on futures contracts is tiny because of the institutional arrangements for trading (see Chapter 12).

5. Options contracts also have a time dimension. Such contracts convey rights or privileges from a seller to a buyer. The buyer can choose to exercise or not exercise the right defined by the contract; this decision is made on or before (depending on the contract) an expiration date. Until 1984, it was illegal to trade options contracts for agricultural commodities in the United States. It first became legal to trade options on futures contracts; i.e., if the option is exercised, the buyer receives a position in a futures contract. These are exchange-traded contracts; the price (premium) of the option is established by an auction system. It is also now possible to write (sell) off-exchange options on commodity positions, provided that the option seller meets certain legal requirements. Prices of off-exchange options are established by negotiation.

A limited amount of empirical evidence suggests that, from a farmer's perspective, forward contracts are more costly to use than futures contracts (Townsend and Brorsen 2000); however, these studies typically do not account for the fact that the default risk is larger for forward contracts than for futures contracts. Nonetheless, farmers tend to prefer forward contracts over using futures contracts (Harwood et al. 1999), and this implies that the overall benefits and costs of forward contracting are not fully understood by economists or perhaps by producers.

## Information, Transactions Costs, and Price Behavior

Price discovery mechanisms can influence price behavior in at least three ways. First, a pricing mechanism potentially influences the efficiency of price discovery through its effects on transactions costs. If a new or revised pricing method is developed that lowers transactions costs and the market is competitive, the lower costs should benefit both buyers and sellers. *Ceteris paribus,* the buyer can offer a farmer a higher price and simultaneously sell the output at a lower price. (The cost savings per transaction may be small, but the collective savings over many transactions can be large.)

Second, pricing arrangements may be developed that improve information transfer. For example, a grid pricing mechanism, which values attributes of beef carcasses, appears to convey more precise information about desired attributes to cattle producers than does a live weight pricing system (Schroeder and Graff 2000). The contract is designed to reflect to producers the value of the carcasses' characteristics. These values are conveyed by premiums or discounts relative to a reference price defined in the contract. The reference price is typically a market price for a specific type of beef carcass that is available from market reports. Thus, the contract is not contributing information about the aggregate supply and demand.

Third, pricing mechanisms can influence the costs of managing risk. Contracts may be designed to share risk between buyers and sellers; the farmer need not bear all the price risk. Or, if a futures market is introduced where it previously had not existed and if this market provides a lower-cost way to hedge than do forward contracts, then the cost of managing risk is reduced. For producers, a decline in the cost of managing price risk has the effect of shifting the supply function to the right, *ceteris paribus*; likewise, the derived demand of buyers may increase due to the risk reduction.

Thus, a generalization is that changes in pricing mechanisms are a response to price risk and transactions costs and, consequently, can influence price behavior. The degree to which all of this is important depends, in turn, on some underlying assumptions about the traders and the information they hold. One important assumption is that traders are generally risk averse; they see risk as a cost. If traders are risk neutral, then the cost of managing risk is not an issue. The fact that pricing institutions are developed to manage price risk suggests that many traders are indeed risk averse.

A second set of assumptions relates to the quantity and quality of information and whether information is held symmetrically (or asymmetrically) by buyers and sellers. As indicated above, pricing commodities requires two types of information: about the factors influencing the overall price level and about the characteristics of the individual trade. The base price depends on information about supply and demand: current and expected production, inventory levels, effects of substitutes, etc. Often, traders in commodity sectors have agreed-on reference prices. For example, if an active futures market exists, the price of the nearby (nearest time to delivery) contract is used as a reference. Such contracts have known, homogenous delivery terms. For the grid pricing system analyzed by Schroeder and Graff (2000), the base price was obtained from USDA market quotations for Choice Yield grade 3 fed cattle. Transactions prices themselves provide information.

For overall supply and demand information, a model developed by Stein (1992) can be modified to show the potential effect of alternative price discovery mechanisms on the quantity of information and, hence, on pricing errors. Recall that, in a perfect market, prices adjust instantly and correctly to new information. In practice, however, time is required for price adjustments. To illustrate this point, the horizontal axis in Figure 11-1 is time, and the vertical axis measures pricing errors. The errors are the difference between current transactions prices and the true (but unknown) equilibrium price for a given information set.

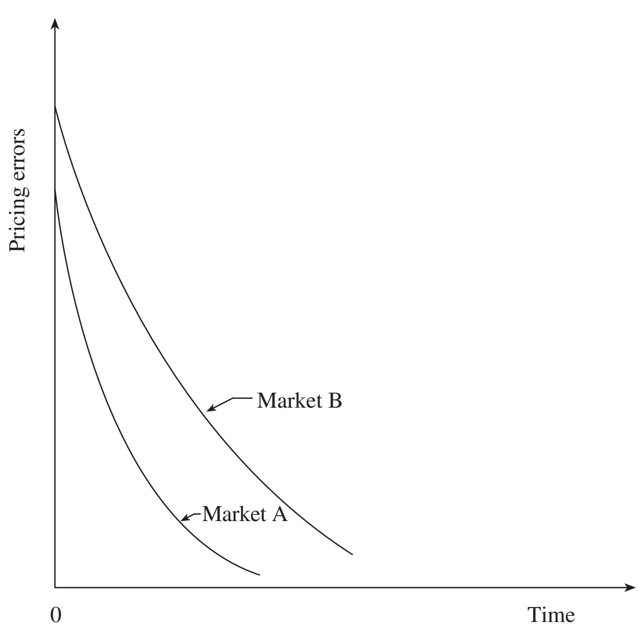



**Figure 11-1.** Pricing errors as a function of time. *Source:* Adapted from Stein (1992: figure 2)

New information is assumed to enter the market at time 0, and the error is hypothesized to be largest then. The errors in price discovery are associated with the costs of learning and adjusting to the new information. As time passes, learning is assumed to occur and the pricing errors decline, as illustrated by the downward-sloping function. (For convenience, the errors are shown as positive; they can be represented by absolute or squared values and hence be plotted in the positive quadrant.)

In active (many traders and a large volume) futures markets, the adjustment to new information may be a matter of minutes or, at most, hours. Price quotations are reported continuously. Of course, potential traders who are not monitoring the market continuously will see the quotations with varying lags after the arrival of the new information. Moreover, varying amounts of time will elapse as traders analyze the news and make decisions. But the elapsed time of adjustment is likely to be small in active markets.

As noted previously, when active futures markets exist, their prices may be used as a basis for establishing cash prices, and these markets provide a continuous stream of price information. For cash markets for commodities where no futures market exists, price adjustments may take days. Prices are likely to be established by negotiation and involve private treaties. Data from spot markets are, at best, reported on a daily basis. In this case, daily prices are potentially autocorrelated because of the length of the adjustment process; i.e., the horizontal axis in Figure 11-1 represents days.

In the United States, market reporters of the Agricultural Marketing Service (AMS) of the USDA collect daily data; cash prices are reported as ranges for selected qualities and locations. For example, prices are reported for various grades of potatoes in Aroostock County, Maine. One such category is units of Russet Norkotah potatoes, baled five 10-pound film bags 2-inch or 4-ounce minimum. Prices on Friday, January 18, 2013, were reported as a minimum of $6.00–7.00 and mostly $6.00–6.50 per unit. In other words, notwithstanding the rather specific quality and location specifications, prices are reported in a considerable range.

Obtaining information about the population of transactions prices is a sampling problem. For any given level of information, the smaller the sample of prices, the less precise the information about transactions will be. Price discovery mechanisms can directly or indirectly influence the sampling rate or the cost of sampling. In addition, markets can have different degrees of complexity and heterogeneity in the product being priced, and the more complex the market, the larger is the sample size needed to achieve a given level of accuracy in pricing. That is, the marginal cost of sampling increases with complexity. Thus, unless the sampling rate is varied, more complex markets are likely to have larger pricing errors than less complex markets. This is represented by the two functions in Figure 11-1, in which the more complex market B has larger errors at any time than does the less complex market A.

Changes in complexity are associated, in part, with changes in the demand for attributes. In an efficient market, prices should reflect the varying characteristics of the individual lots. Thus, as the need to achieve more precise pricing of attributes increases, pricing institutions develop to accommodate this complexity. But, although contracts (negotiated as private treaties) can be designed to price the attributes more accurately, the amount of information that is publically available is reduced. Market reporters have access to less information as the number of private treaties increases as a proportion of total transactions. The sampling rate is, in effect, reduced.

As Tomek (1980) has pointed out, the definition of a "thin market" is, in part, an empirical question. What proportion of the total population of transactions must be sampled to obtain a given level of pricing accuracy? With the shift toward negotiated private treaties in spot markets, it is more costly to sample prices. Those decision makers who are concerned with the public policy about price reporting must ask, what is the desired level of accuracy in pricing, and what is the sampling rate required to obtain the desired accuracy?

Another aspect of pricing accuracy in complex markets is the potential for information to be held asymmetrically. Such asymmetries may favor either buyers or sellers. Large buyers may have more information about the broad economic factors influencing price levels. A large grain buyer in the United States, for example, may have information about growing conditions in other parts of the world earlier than do producers. On the other hand, at the individual transaction level, the farmer may know more about the possible defects of his produce than does the buyer. The production contracts, discussed earlier, help eliminate the asymmetries at the individual transaction level.

In summary, the issues related to the effects of pricing institutions on price behavior are complex. It is clear that alternative pricing mechanisms influence individual transactions prices and perhaps overall price levels as well. Logical hypotheses are that pricing institutions evolve as the economy changes to minimize transactions costs between the individual buyers and sellers, but that these changes can have the side-effect of reducing the amount of publically available information about prices.

The role of futures markets in price discovery and the potential relationship to spot prices are discussed in the next two chapters. Before turning to that topic, however, we discuss some of the ways in which government policy can influence agricultural product prices.

## Government Intervention in Pricing Agricultural Products

Governments in nearly every country have attempted to influence the prices of at least some agricultural products. Sometimes, the objective is to reduce the prices of food paid by urban consumers. Often, especially in developed countries, the objective is to support the prices and incomes of farmers and/or reduce price and income instability. Associated objectives may be to preserve small farms and to achieve self-sufficiency in food and fiber production (or to decrease dependence on imports). Government policies to achieve these objectives fall into two broad categories: (1) domestic policies and (2) international trade policies. International trade policies were discussed in detail in Chapter 8. This section focuses on domestic policies, but it is important to recognize that international trade policies also play a crucial role in influencing prices.

In this section, we review selected types of domestic government intervention that are intended to raise farm prices and perhaps reduce the variability of prices. Such programs require some type of administrative actions and can be broadly viewed as falling into the category of administered prices. Not all government programs, however, directly set prices. There is a large literature on agricultural policy and the effects of these policies (e.g., Knutson, Penn, and Boehm 1996), and our discussion is limited to the consequences for price behavior of such programs.

The economic effects of government programs designed to support or raise farm prices depend, *inter alia,* on the level of support and on the methods employed to raise prices. It is also possible to enhance farm incomes via direct payments or subsidies; such payments may influence prices indirectly via their effects on farmers' incentives to produce. If, however, the program operates by attempting to influence market prices, then it is necessary to restrict supply (e.g., through acreage restrictions or herd buyout programs) or increase demand (e.g., through food subsidies or generic advertising). We turn to a discussion of selected programs.

## Price Supports

Commodity prices can be maintained above the equilibrium level if the government is willing to provide the demand for the surplus production generated by the support price. For storable commodities, the government can buy and hold inventories. The amount that must be acquired and the cost (to taxpayers) of doing so depend on the level of support relative to the equilibrium price and on the slopes of the relevant supply and demand schedules. The level of support is determined by political decisions, but perhaps also the secretary or minister of agriculture is given some discretion in setting prices.

The critical relationships are illustrated in Figure 11-2. If the support price is $P_2$ and the equilibrium price is $P_1$, then, absent any attempt to control supply, farmers will produce the quantity $Q_2$, but buyers will demand only the quantity $Q_1$. The government will have to acquire an amount represented by the shaded area, which is the quantity $Q_2 - Q_1$. Total consumer expenditures are represented by the unshaded rectangle ($P_2 \times Q_1$). Producers receive the sum of the consumer and government expenditures.

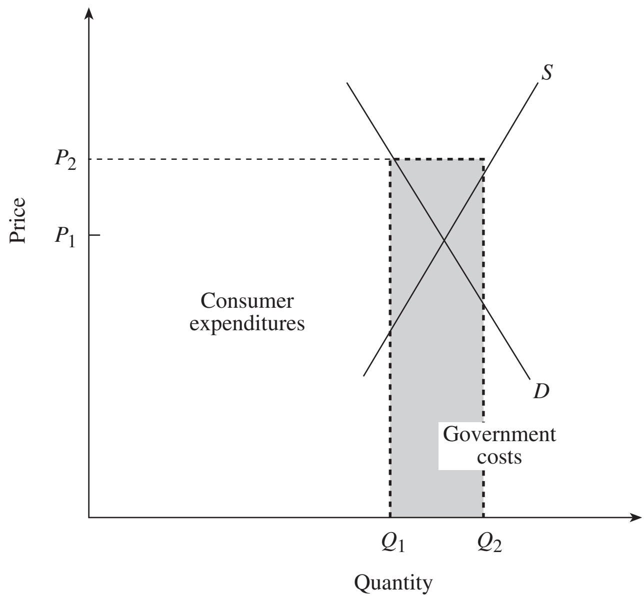



**Figure 11-2.** Consumer expenditures and government costs of supporting prices above equilibrium through a government purchase program

The cost to the government of a given level of support depends on the slopes of the demand and supply functions. If the supply and demand curves shown in Figure 11-2 had been drawn with flatter slopes, it is clear that the government costs would be larger than those shown. The figure is, however, drawn for static supply and demand schedules. The potential to recover government costs (to reduce government-held inventories) will depend on how supply and demand change with the passage of time. For example, if demand grows faster than supply, this will give the government an opportunity to release stocks while maintaining the support level. But if supply grows relative to demand, stocks will accumulate, and at some point, no government, no matter how rich, will be able to maintain a status quo that involves continued stock accumulation.

High support prices also have the potential to stimulate the development of lower-cost substitutes. A plausible hypothesis is that the price support program for cotton in the United States provided an umbrella for the development of synthetic fibers. Likewise, the support of butter prices made butter relatively more expensive than its principal substitute, margarine. Thus, with the passage of time, inventories can increase because of the decline in demand associated with the introduction of substitutes (which were, in turn, encouraged by the price support program).

There is another practical problem in administering a price support and related government purchase program—support prices received by farmers should be adjusted to take account of the quality and location characteristics of individual lots of the commodity. When no price supports exist, markets determine the relevant price differences and appear to do so relatively efficiently (Chapters 7 and 8). But this is more difficult to do administratively, and with inappropriate administrative differentials, the government will acquire low-quality stocks in undesirable locations.

One approach to the problems associated with acquiring inventories is to try to limit production, and another is to try to subsidize exports. For example, if production can be held to the quantities consumers are willing to buy at the support price ($Q_1$ in Figure 11-2), then the government can avoid the costs of acquiring and storing inventories. Of course, the total revenue received by producers is less under an effective supply-control program than under a government purchase program, since a smaller quantity is produced and sold at the higher price. But farmers may be better off because they do not have to incur the costs that would be required to produce the larger quantity. Also, some government programs have compensated farmers for not producing. In general, supply management programs require considerable administrative apparatus to assure compliance with program rules.

## Direct Payments

Rather than attempting to influence supply or demand to raise farm prices, a program can be designed to make direct income payments to farmers. This type of program is perhaps more acceptable because it allows market forces to determine prices. Buyers of agricultural products pay market-clearing prices. Market prices allocate inventories, and the government does not have to manage surpluses.

If the market prices are below those deemed equitable to farmers, then direct payments are made to farmers. The magnitude of the payments is a political decision, and the funds come from taxpayers.[6]

The economic effects of an unlimited deficiency payment program (i.e., with no limits on production or the size of payments to individual farmers) are illustrated in Figure 11-3. In this figure, the price guaranteed to producers is $P_3$, which is above the equilibrium price, $P_2$. The quantity supplied, however, will be $Q_2$ because farmers receive $P_3$ not $P_2$, since the effective price received is the market price ($P_1$) plus the deficiency payment. Assuming the quantity $Q_2$ is

6. The United States has had several deficiency payment programs for major crops. At one time, payments were based on a difference between a "target price" and an "average market price," and the payment was based on historical production rather actual current production. A more recent program based the payment on the difference between the support price (loan rate) and a current market price (on a date selected by the farmer). The payment was based on proven actual production. Because program rules change and vary from country to country, we describe here a generic payment scheme.

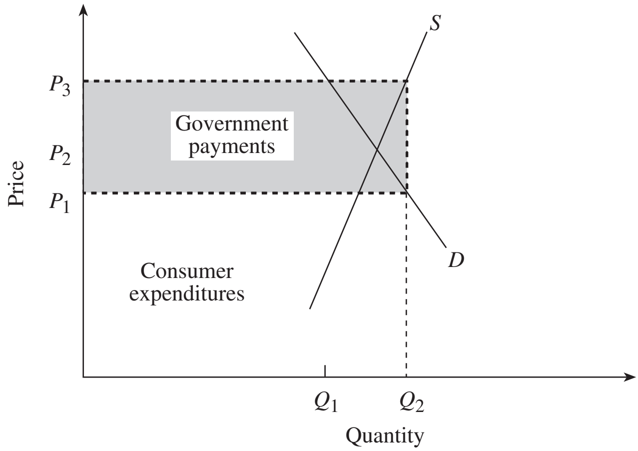



**Figure 11-3.** Consumer expenditures and government costs of supporting prices above equilibrium by making deficiency payments

realized (no yield risk), it is placed on the market, and the market-clearing price is $P_1$. Thus, the actual market price is below the equilibrium price that would have existed without the program.

Government payments will be equal to the difference between the guaranteed price ($P_3$) and the market-clearing price ($P_1$) multiplied by the total volume of production. This is represented by the shaded area in Figure 11-3. Consumers benefit from below-equilibrium prices, while producers receive above-equilibrium prices. Both production and consumption are subsidized. Of course, the payments come from tax revenues.

Analogous to the previous discussion of other programs, the slopes of the demand and supply schedules influence the cost to the government of the program. Shifts in the functions also influence the size of payments. In principle, demand could grow relative to supply, such that the market price rose above the guaranteed prices, and government payments would drop to zero. The reverse is, of course, also possible.

At the beginning of the twenty-first century, the prices of major grains in the United States were low relative to support levels, and the federal government made large payments to crop farmers. More recently, dairy farmers have received deficiency payments as part of U.S. dairy policy. But, as noted in Chapter 9, agricultural product prices have risen substantially since 2005, thus raising questions about the need for direct payments. In addition, deficiency payment programs are criticized by free-trade advocates because they have effects similar to export subsidies; i.e., exports are stimulated by the artificially lower domestic market price.

## Stabilization Schemes

The large volatility of agricultural product prices implies that the price risk faced by buyers and sellers of commodities is large. A price support program may help stabilize prices. It provides a floor price, and if demand should increase so that market prices rise above the support level, stocks can be released that reduce the upward movement in prices. Indeed, some programs such as the dairy price support program were originally intended to lessen the seasonal highs as well as lows in milk prices.

In principle, a program could be designed to reduce the amplitude of fluctuations without affecting the average level of prices. For storable commodities, this could be achieved by the management of inventories. Stocks are acquired when production is large and prices would otherwise be low. These inventories are then released when production is low, thereby preventing prices from rising as far as they otherwise would. Of course, it is not possible to maintain prices exactly at the long-run equilibrium level, but it might be possible to hold prices within a band. In practice, even this is difficult to do, partly because it is difficult to predict the long-run equilibrium price and hence to establish appropriate acquisition and release prices.

One can visualize the effects of a stabilization program on prices as shifts in a supply function along a static demand function. When a random shock (say, good weather) results in a large crop, this is depicted as a shift in the supply function to the right relative to the normal supply. The government agency acquires stocks, helping maintain prices; the supply entering market channels is then that associated with normal conditions. Subsequently, when a random shock reduces supply, inventory is released from government-held stocks, thereby reducing price relative to the level it otherwise would have been.

A reserve scheme that succeeds in reducing the amplitude of price fluctuations may help producers survive a period of depressed prices, but it will not necessarily lead to more stable returns for farmers. When prices are free to fluctuate in response to changes in production, high prices tend to offset low production, and vice versa. Price stabilization will reduce revenue instability, compared with what would happen in the absence of a stabilization program, only under certain conditions. In particular, if the price elasticity of demand is less than –0.5 (in absolute value), a stable price will reduce revenue instability compared with fluctuating prices, although revenue will still vary. If the elasticity exceeds –0.5 (in absolute value), revenue will be more unstable with a constant price than with prices that are allowed to adjust to changes in supply.

Likewise, average returns over a period of years may or may not be enhanced by reducing the amplitude of price fluctuations through a storage program. The effect on average returns of acquiring stocks and then reselling them later depends on demand conditions at the time of acquisition relative to the time of sale, the cost of holding commodities per unit of time, and the length of time they are held. For example, if demand remains stable, producers' aggregate revenue will increase over a period of years if the price elasticity of demand at the time stocks are acquired is more inelastic than when the stocks are released. This is indeed the case if the demand function is linear because, as one moves along the function from left to right, demand becomes more price inelastic (Chapter 3). (These ideas go back many years and are elaborated on in Eckstein and Syrquin 1971; Gislason 1959; Newbery and Stiglitz 1981.)

In sum, it is theoretically possible for governments to administer prices and reduce price variability, but this does not necessarily reduce the variability of returns to farmers. In general, attempts to stabilize prices through storage schemes have not been successful. One additional problem occurs in successive periods of consecutive high (or low) production, since in such cases inventories are at capacity (or near zero) and the government cannot acquire (release) additional stocks, making it impossible for the government to influence market prices.

## Cooperative Actions and Marketing Orders

Governments may influence agricultural product prices indirectly by providing the legal authority and climate for producers to act collectively. In the United States, the Capper-Volstead Act allows farmers to act collectively in marketing their products. Within some specified limits, farmers' cooperative actions are exempt from antitrust laws. Consequently, farmers may be able to use the principles of price discrimination (Chapter 5) in an attempt to enhance returns, cooperate to control the quality/quantity of the commodity marketed, and cooperate to collect funds to support an advertising program.

One vehicle for cooperative action in the United States is marketing orders. Milk marketing orders, for example, establish formulas for the price paid to farmers for milk. An order establishes prices for different classes of milk, which go for fluid and various manufacturing uses. Higher prices are established for the milk to be sold as fluid milk in retail stores, which presumably has a more price inelastic demand than the manufactured products such as butter and cheese. While processors pay these class prices according to how they use the milk, farmers receive a weighted average of these class prices based on the overall market utilization. Thus, classified pricing should increase total returns.

Other marketing orders influence prices by specifying the minimum grades and size of products that can be marketed. Sometimes the volume of produce moving into a market each week is controlled. The details of such programs are beyond the scope of this book (for more information, see Knutson, Penn, and Boehm 1996), but the general point is that farmers can organize to influence prices through controlling the quantity and quality of produce being marketed.

Another, indirect way of influencing the prices of commodities is through advertising. Although an agricultural commodity may have variable attributes, it is typically difficult to differentiate and brand the farm output. For example, milk can vary in its fat and protein content (and indeed the price received by farmers is influenced by these characteristics), but farmers do not attempt (typically) to brand their milk. Hence, few incentives exist for individual farmers to advertise.[7] Farmers might benefit, however, if they could organize and pool funds to advertise a commodity like milk. The hope is that the increase in aggregate demand resulting from advertising will raise farm prices sufficiently to more than cover the cost of advertising. This question can be answered only by empirical analysis on a commodity-by-commodity basis. Considerable research has been conducted on the economics of generic advertising, and farmers often appear to benefit from advertising (e.g., Kaiser et al. 2005, and references therein).

Thus, prices of agricultural products can be influenced by the permissible collective actions of farmers. These collective actions, in turn, depend on the legislative and administrative assistance of agencies like the USDA.

## Concluding Remarks

Economic growth and change provide incentives for changes in pricing institutions and/or the outcomes of pricing institutions. The demand for agricultural products is derived from many end uses, and this creates a need not only to establish base prices but also to accurately price the attributes of individual lots of commodities. Hence, many contractual arrangements have and are being developed, typically via a negotiation mechanism. But, even though the volume of transactions priced by older methods has declined in the United States, these arrangements do not disappear.

It is logical to think that a diversity of pricing arrangements persists because of a desire to minimize private transactions costs. This may result in market failures, in the sense that a large proportion of total transactions are priced by mechanisms that do not take account of the public costs that they induce. In particular, the growing use of private treaties reduces the amount of public information.

The role of governments in pricing farm commodities has also been changing. There is more emphasis on freer trade and lower levels of price support, at least in public debates. Nonetheless, government programs are still in place that influence agricultural prices. Also, governments are likely to continue to play a role in providing unbiased information about the economic forces influencing commodity prices (in crop reports, inventory reports, reported export quantities, etc.) as well as about prices themselves.

7. However, there have been instances of farmers challenging mandatory generic advertising programs, claiming that their product is different, e.g., higher quality than the generic commodity being advertising. For example, some raisin growers, which are using a harvesting technique called "dried on the vine," have challenged the California Raisin Marketing Board, asking to be exempt on the basis that their raisins are different from the ones being advertised.

## References


Becker, G. M., M. H. DeGroot, and J. Marschak. 1964. "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science* 3:226–232.

Coase, R. 1937. "The Nature of the Firm," *Economica* 4:386–405.

Eckstein, S., and M. Syrquin. 1971. "A Note on Fluctuations in Supply and Farmers' Income," *Am. J. Ag. Econ.* 53:331–334.

Gislason, C. 1959. "How Much Has the Canadian Wheat Board Cost the Canadian Farmers?" *J. Farm Econ.* 41:584–599.

Harwood, J., R. Heifner, K. Coble, J. Perry, and A. Somwaru. 1999. *Managing Risk in Farming: Concepts, Research, and Analysis*. USDA Ag. Econ. Rep. 774.

Hueth, B., E. Ligon, S. Wolf, and S. Wu. 1999. "Incentive Instruments in Fruit and Vegetable Contracts: Input Control, Monitoring, Measuring, and Price Risk," *Rev. of Ag. Econ.* 21:374–389.

Kaiser, H. M., J. M. Alston, J. M. Crespi, and R. J. Sexton, eds. 2005. *The Economics of Commodity Promotion Programs: Lessons from California*. New York: Peter Lang Publishing.

Klemperer, P. 2002. "What Really Matters in Auction Design?" *J. Econ. Persp.* 16(1):169–189.

Knutson, R. D., J. B. Penn, and W. T. Boehm. 1996. *Agricultural and Food Policy*, 6th ed. Englewood Cliffs, NJ: Prentice Hall Education, Career & Technology.

Lence, S. H., and M. L. Hayenga. 2001. "On the Pitfalls of Multi-Year Rollover Hedges: The Case of Hedge-to-Arrive Contracts," *Am. J. Ag. Econ.* 83:107–119.

MacDonald, J. M., and P. Korb. 2011. *Agricultural Contracting Update: Contracts in 2008*. USDA Econ. Info. Bull. EIB 72.

Milgrom, P. 1989. "Auctions and Bidding: A Primer," *J. Econ. Persp.* 3(3):3–22.

Newbery, D. M. G., and J. E. Stiglitz. 1981. *The Theory of Commodity Price Stabilization*. New York: Oxford Univ. Press.

Rogers, R. T., and L. M. Petraglin. 1994. "Agricultural Cooperatives and Market Performance in the Food Industry," *J. Ag. Coop.* 9:1–12.

Schroeder, T. C., and J. L. Graff. 2000. "Estimated Value of Increased Pricing Accuracy of Fed Cattle," *Rev. of Ag. Econ.* 22:89–101.

Sexton, R. J. 2013. "Market Power, Misconceptions, and Modern Agricultural Markets," *Am. J. Ag. Econ.* 95:209–219.

Stein, J. L. 1992. "Cobwebs, Rational Expectations and Futures Markets," *Rev. Econ. & Stat.* 74:127–134.

Tomek, W. G. 1980. "Price Behavior on a Declining Terminal Market," *Am. J. Ag. Econ.* 62:434–444.

Townsend, J. P., and B. W. Brorsen. 2000. "Cost of Forward Contracting Hard Red Winter Wheat," *J. Ag. & Applied Econ.* 52:89–94.

Unterschultz, J., F. Novak, D. Bresee, and S. Koontz. 1998. "Design, Pricing, and Returns of Short-Term Hog Marketing Window Contracts," *J. Fut. Mkts.* 18:723–742.

CHAPTER 12

# Price Relationships on Commodity Futures Markets

Markets for futures contracts are important pricing institutions for many of the major farm commodities in the United States and increasingly in other countries. These markets provide an opportunity for producers, inventory-holders, and users of commodities to shift the risk of adverse price movements and to forward price commodities. Futures markets also offer an opportunity to speculate and perhaps profit from changes in prices. This chapter focuses on the behavior of futures prices and the relationship of futures to cash prices. We begin with a discussion of markets for contracts. Then, basic models of price determination are introduced. The prices of options-on-futures contracts are also briefly discussed. The functions of futures markets are described in more detail in Chapter 13.

## Markets for Contracts

Cash prices reflect values of commodities for delivery "soon" or "immediately," and they reflect the unique quality and location attributes of the individual lots of the commodity (Chapter 11). In contrast, prices of futures contracts value the delivery of a commodity at a future date, and since the contracts have homogenous provisions, the transactions need establish only the price and number of contracts traded.

Markets also exist for options contracts that convey rights to buy or sell futures contracts. These markets establish the value of such rights, not of the commodity itself. It is also legal in the United States, provided certain requirements are met, to sell options contracts on the physical commodity (over-the-counter contracts). Thus, prices are established for immediate delivery, for forward delivery (forward contracts), for futures contracts, for options-on-futures contracts, and for options on the physical commodity.

## Futures Contracts

A futures contract is a legal instrument, enforceable by the rules of the exchange on which it is traded, to deliver or accept delivery of a specific amount of a commodity during a specified month at a price established in the market. For example, on October 1, trader A buys a May wheat contract at $7.00 per bushel from trader B. The two traders have entered into a legal *obligation* in October for delivery of 5,000 bushels of wheat within a designated time period in May. The contract calls for delivery of a specific grade of wheat at a specific location.

All contracts for a particular commodity on a given exchange are identical, but contract specifications differ from market to market (e.g., wheat contracts at Chicago, Kansas City, and Minneapolis are for different types of wheat). Most contracts permit the delivery of substitute grades at stipulated discounts or premiums; likewise, alternative delivery locations are permitted, with defined spatial discounts and premiums. Since the quality attributes are typically defined by ranges, the seller is expected to deliver the minimum level of attributes that meet the grade requirements of the contract. It is understood that the seller will use the least-cost delivery alternative. Thus, there is reasonable, but not perfect, certainty among buyers and sellers about what is being priced.[1]

Trading in futures contracts takes place under the auspices of formally organized markets, such as the Chicago Board of Trade (a division of the Chicago Mercantile Exchange Group). An organized market defines the contracts and related rules, facilitates the mechanics of trading, aids in disseminating market information, and helps regulate the business dealings of its members. Historically, trading took place by open outcry at a specific location during specified times, and as of the year 2013, some transactions in some commodity markets still occur via this method. Electronic (computerized) trading is, however, the dominant method for consummating transactions. Whatever the method of trading, futures markets are double-auction markets; traders have the right to both buy and sell. Most market participants establish accounts with a futures commission merchant (FCM), i.e., a broker, who provides access to a market (or markets). Trades can be executed by placing an order with a broker or via one's computer terminal. The FCM charges a fee for each transaction. Large traders may be members of those exchanges relevant to their business.

To start, a trader must make a margin deposit. The margin is analogous to a performance bond or security deposit rather than a down payment. As prices change, the margin account is either credited or debited, depending on the trader's position and the direction of the price change. The position is said to be "marked to the market." For example, if an individual has purchased a

1. Some futures contracts are settled by cash payments rather than by delivery, but cash settlements are uncommon for agricultural commodities. It is difficult to define an appropriate settlement price that can be accurately measured. Unlike the prices of financial instruments, it is costly to obtain high-quality observations of cash prices of commodities (Chapter 11).

contract (has a long position in futures) and prices rise, the account is credited with the increased value; if prices decline, the account is debited. If equity in the account falls below a "maintenance level," the margin funds must be replenished. This is referred to as a *margin call.*

The precise mechanics for determining and collecting margins are not important for our purposes, but it is essential to understand the function of margins. They provide insurance against defaults; margins cover losses on futures positions. With adverse price moves, calls to replenish the margin account can accumulate into a large sum. If margin calls are not met, the position will be liquidated. (FCMs require more than the minimum initial margin set by the exchange, with the "surplus" available in an account that can be used to meet margin calls.) Since the gains and losses are taken as they occur, the risk of default is negligible. The marked-to-market and margin system differentiate futures from forward contracts.

While trading in standardized contracts has the benefit of establishing prices for a well-defined entity, it also means that futures are typically used as temporary forward positions. Most traders do not intend to make or take delivery; initial positions are usually offset by taking the opposite position in the same contract before the maturity month arrives. This is done easily because futures markets use the device of a clearinghouse. The clearing organization of the exchange keeps records of the positions of its members,[2] and the clearing operation removes the individual responsibility of one member to another. FCMs are either clearing members or have a relationship with a clearing member. Thus, most individual trader's margins are administered via their FCM.

The net effect is, for example, that if A sells a May wheat contract to B and if B sells a May wheat contract to C, then A has a short position and C has a long position with the clearinghouse. Trader B, by first buying and then selling, has no position in May wheat. Traders may make numerous transactions through time, but they are responsible only for their net positions.

Those with a short position can elect to make delivery. Typically, the contract provides the seller with some flexibility about the timing of delivery within the maturity month. Notice of delivery must be given, and when a trader with a short position gives notice of delivery, the clearinghouse in turn notifies a trader with a long position. (The particular long receiving the notice is determined by the rules of the exchange.) The trader receiving notice of delivery may choose to sell futures and pass on the notice, or may choose to accept deliver. Most traders would not hold a long position into the delivery month unless they were willing to receive a delivery notice. The actual delivery method for the grains involves a transfer of title for a commodity at a particular location.

Although deliveries are a small fraction of the total volume of trading, they can be a large proportion of the positions open at the beginning of the delivery

2. The term *position* refers to the individual's status as a net buyer (long position) or a as a net seller (short position) of contracts.

month. It is delivery, or the possibility of delivery, that causes futures prices to converge (approximately) to cash prices for the maturity month. In other words, the potential for profitable arbitrage between cash and futures causes the two prices to be nearly identical at contract maturity. When contracts are settled by cash payments, the futures price converges to the cash settlement value defined in the contract. As emphasized repeatedly in this book, however, a variety of prices exist at a point in time for a given commodity, and although these various prices are highly correlated over time, they are not identical. Thus, inevitably, at contract maturity, the price defined by the futures contract will not be identical to most cash prices.

## Options Contracts

Since 1984, it has been possible to buy and sell options-on-futures contracts for agricultural commodities traded in U.S. markets. An option gives the buyer the right (privilege) to a position in a futures contract at a price fixed in the options contract. The buyer need not exercise the right. A *put* gives the buyer of the option the right to a short position in a particular futures contract. That is, the owner of a put has obtained a right to sell a futures contract at the price designated in the put; if the right is exercised, the buyer then obtains a short position in futures. A *call* option conveys the right to buy a futures (obtain a long position) at a specified price within a given time period.

The price of the underlying futures contract specified in the option is called the exercise or strike price. Options are traded for a large number of strike prices. An option has an expiration date, which is related to (but not the same as) the maturity date of the underlying futures. For example, the last day of trading for options on live cattle futures, traded on the Chicago Mercantile Exchange, is the first Friday of the contract maturity month (e.g., the first Friday of February),[3] while the last day of trading for the futures contract itself is the last business day of the maturity month. Most options traded in U.S. markets are "American options," which means that they can be exercised at any time from inception through the expiration date. European options can be exercised only on the maturity date.

Trading in options-on-futures contracts, like trading on futures, is done through the facilities of organized marketplaces. The price established in the marketplace is not the price of the commodity itself but, rather, a premium paid for the right granted by the option contract. The buyer pays this premium to the seller (called a writer) of the option.

Premiums for puts and calls are illustrated in Table 12-1. These prices are approximate values derived from one day in February 2013 in a corn market. (Actual historical prices can be obtained from a variety of Internet sites. Of course, traders see real-time prices on their computer screens.) The strike price,

3. This rule is an illustration. All contract rules are subject to change.

**Table 12-1.** Hypothetical premiums on a February day for May options, corn market

| | Premiums (¢/bu.) | |
|---|---|---|
| Strike price (¢/bu.) | Calls (long) | Puts (short) |
| 675 | 32 | 27 |
| 680 | 29 | 30 |
| 685 | 27 | 33 |
| 690 | 25 | 36 |

in the first column of the table, is the price of the underlying futures contract to the option owner if she or he elects to exercise the option. The figures in the second column of Table 12-1 show premiums for calls, while the third column contains premiums for puts. It would have cost 32 cents per bushel in February to obtain the right to a long position in a May futures contract at 675 cents per bushel.

At higher strike prices relative to the futures price, the premium for a call option declines and the premium for a put option rises. Intuitively, it is more expensive to purchase a right to sell at a price that is above the current market price for future delivery than it is to purchase an option to sell at the current market price.

Options contracts, like futures contracts, are used as vehicles for hedging and speculating. For example, a firm that owns an inventory of corn and wants to avoid the risk of a price decline could purchase a put option. If, between the purchase and expiration dates, the underlying futures price declines, the option to sell a futures contract at, say, 680 cents per bushel will become more valuable. The rise in the premium will offset the decline in the futures price. By selling the option for a higher price (premium), a trader can offset losses caused by the lower market prices for the commodity. Traders in put options will sell them by the expiration date if premiums rise and will not exercise the option if the futures price rises above the strike price.[4] In the latter case, the premium falls to zero.

From a buyer's viewpoint, futures and options markets differ in terms of their cash flow. A futures contract involves a legal obligation, and both buyers and sellers receive daily cash flows from the margin and marked-to-market features of the contract. For options, the buyer pays a premium and benefits only if the contract is exercised. Otherwise, the return from the option is zero. The maximum cost to the buyer is the initial premium (plus brokerage fee).

4. The benefits can be obtained by taking offsetting positions rather than by exercising the option. It is arbitrage between options and futures that keeps the respective prices in alignment and hence makes exercise unnecessary.

The seller of an option, called an option writer, receives the premium and must make a margin deposit. If prices move adversely (e.g., prices rise for the seller of a call), margin calls will occur. The writer of naked (unhedged) options is exposed to almost unlimited price risk, but the premium-margining system for options operates through a clearinghouse, like futures trading. Thus, as for futures, the risk of default on options on agricultural futures traded in regulated U.S. markets is small. It is precisely because the risk of default is large for unregulated options contracts that such contracts have a tainted history.

One issue in using futures and options markets for hedging is the alternative costs. Aside from the differences in cash flow just discussed, a potential difference in cost relates to whether a transaction influences price. By definition, in liquid markets, a large transaction has little effect on price; a buy order may cause prices to rise by the minimum tick. With low liquidity, the price effect is larger. This price effect is a cost to the hedger because (say) the purchase is made at a higher price and the subsequent sell order is made at a lower price than would have occurred in a liquid market. Options markets typically are less liquid than comparable futures markets. Another trade-off is between the up-front cost of an option and the possible margin calls from holding futures positions.

Premiums for put and call options-on-futures are related to the underlying futures prices, and futures prices are related to the underlying spot price. Firms considering hedging need to understand these relationships. In subsequent sections, we discuss models of futures prices and their relationship to cash prices. The determination of options prices is also briefly discussed.

## Establishing Prices for Futures Contracts

Prices are the outcome of the transactions executed through the auspices of the various futures markets. Traders bring a variety of motives to the marketplace. The total set of trades may be classified by their purposes, broadly as hedge or speculative positions. Although it is difficult to provide unambiguous definitions of hedging and speculation, classifying the types of traders helps identify the motives that participants bring to the marketplace.

### Types of Traders

*Hedgers.* An oversimplified but common textbook view of hedging is establishing a position in futures exactly opposite from the one held in the cash market. The implied motive is to shift the price risk. In a classic example of a selling hedge, the owner of grain inventory sells futures contracts to protect against a decline in the price of grain. Later, as the grain is sold out of storage, the merchant covers (offsets) the original position in futures by buying the equivalent contracts. Since cash and futures prices are positively correlated, a decline in the price of the cash commodity is more or less offset by the gain in the value of the futures contract.

Holbrook Working (1953) provided a broader and more accurate definition of a hedge—the use of a futures contract as a temporary substitute for a later transaction in the cash market. A selling hedge anticipates the future sale of a commodity in the cash market and starts with the sale of futures contracts. Thus, a hedge does not necessarily require an explicit position in a physical commodity. For example, when a U.S. farmer plants a soybean crop in May, he or she expects to harvest the crop in October. A harvest-time sale in the spot market could be anticipated by selling November soybean futures prior to harvest. The existence of a futures market allows the producer to price the commodity at any time over the life of the contract. Presumably, the producer establishes a price because it is profitable to do so.

Working (1953) classified such a transaction as an anticipatory hedge, and he emphasized the profit motive (rather than shifting risk) as a basis for hedging. Although a hedge can more or less assure a sale price, prices will change over the life of the futures position and will result in debits or credits to the margin account. In this sense, an anticipatory hedge has a speculative element, and a futures hedge of a growing crop also has yield risk.

A buying hedge starts with the purchase of futures contracts. For example, a processor or exporter may know that he or she will need to buy a commodity at a future date and can assure (approximately) the purchase price through a hedge. Typically, the commodity will be procured in the cash market and the long futures position offset at the same time to complete the hedge.

In sum, firms' motives for hedging are to establish a buying or selling price (hence, help assure a profit) and/or to shift price risk. In the terminology of economics, firms are assumed to maximize expected utility, and their utility is assumed to depend on profits and the riskiness of these returns. Firms presumably use futures markets to accomplish these objectives when hedging in futures is the least-cost way to do so. Recall (Chapter 11) that forward contracting can also be used to establish a price and manage risk, and although forward contracting is relatively popular with U.S. farmers, the existence of active futures markets for commodities implies that they are useful hedging vehicles for many firms. A corollary is that the continuation of active trading in futures markets depends on them remaining the low cost way to establish price and manage risk for some firms.

*Speculators.* Speculators take positions in futures with the expectation of profiting from price changes. It is convenient to categorize speculators into three groups: scalpers, position traders, and spreaders. These types of traders differ in the length of time they hold positions, the proportion of professionals (individuals who make their primary income from speculation), and the function they perform in the market.

Scalpers are professional speculators who trade frequently and hold positions for only short periods of time. They are an important source of liquidity in markets and are the equivalent of market makers in security markets. Scalping is essential because of the random order flow entering markets. It is unlikely that matching buy and sell orders will enter a market at the same time. Scalpers provide a bridge between the various orders. For example, if a large short hedge (sell) order enters the soybean market, a scalper will probably be willing to take the opposite side of the transaction (buy) at a price 0.25 cent below the previous price (where the minimum price change is 0.25 cent). Having purchased soybean contracts, the scalper expects to sell them, perhaps a few at a time, at a price above the purchase price. These sales typically occur in the next few minutes to hours. Scalpers often have a zero position in the market at the end of their trading day.

Scalpers are not forecasting long-run trends in prices. Basically, the simple, implicit model underlying their trades is that the mean price is a constant and that they can profit from the small variation in prices around the mean, analogous to a bid-ask spread, which is related to the transactions effects of the order flow. Of course, new information will enter markets and prices will adjust up or down to new levels. Successful scalpers must be able to identify these "trends" and minimize the losses from them. Not every trade made by a scalper will be profitable, but since scalpers are professionals who earn their primary living from trading futures, they must be successful on average. (Empirical evidence on the profitability of scalping is limited because it is private, but see Kuserk and Locke 1993.)

The price effect of placing and lifting hedge positions is a cost to the hedger, and the existence of liquid markets is essential for low-cost hedging. In the relatively liquid corn market, the bid-ask spread is less than 0.5 cent, even in recent years with highly variable prices (Wang, Garcia, and Irwin 2012). Thus, speculators, especially scalpers, play an important economic role in futures markets.

Position traders buy or sell futures based on their expectations (forecasts) about changes in price levels. If, for example, a trader thinks a price is going to rise, he or she buys contracts expecting to sell later at a higher price. Position traders are acting on implicit or explicit models of price behavior. Their forecasts of price changes are based either on analyses of fundamental supply and demand factors or on technical analyses of past prices (and perhaps of other variables like trading volume). Numerous methods of technical analysis exist, but all are designed to identify patterns of behavior in time series that can assist in forecasting the future (e.g., Kleinman 2001). Positions will likely be held for periods longer than a day, perhaps for weeks or months.

In addition to providing liquidity, position traders can potentially contribute to the stock of information determining prices by their analysis of data. But ill-informed traders may contribute to unwarranted price changes by, for example, relying on unfounded rumors. Although the evidence is limited, it appears that the majority of amateur position traders lose money. It also appears that professionals managing commodity pools (the equivalent of mutual funds for stocks) cannot consistently earn returns that exceed those in alternative investments.[5] It is possible, however, that a few traders have better private information than do most position traders. Thus, they profit from their better investments in relevant information and/or in better models.

Spreaders are arbitragers. They make offsetting purchases and sales of different futures contracts based on the expectation of a change in the *relative* prices of the different contracts. Again, the forecast may be based on an explicit model or an implicit informal model of price changes.

Given the large number of contracts and markets, the potential arbitrage opportunities are numerous. Purchases and sales of contracts may be made in different delivery months for the same commodity (an inter-delivery spread), between contracts for the same or similar commodity in two different markets (say, the Kansas City and Chicago wheat contracts), between different commodities, such as corn and oats, or between the commodity (say, soybeans) and its products (meal and oil). The latter two are called intercommodity spreads. Given the existence of markets for options-on-futures contracts, arbitrage opportunities also exist between futures and options contracts as well as between different options contracts.

Spreading serves the important purpose of keeping prices of the different contracts consistent with warranted economic values. For example, if a trader thinks that the difference in prices between the March and May corn contracts is too large relative to storage costs, then he or she will purchase March contracts (assuming this is the lower price) and sell May contracts. Later, the spreader will make offsetting sales and purchases. In this example, the actions of spreaders will tend to reduce the difference in prices by raising the March price (by purchasing them) and lowering the May price (by selling them). Both prices may rise or fall after the initial trades are completed, but the change in overall level does not affect the spreaders' profits or losses as long as both prices change by the same amount. The spreader gains or loses only if the relative prices change. In the foregoing example, the spreader gains if the forecast of a narrowing of the price difference occurs. The initial purchase of the March contract will be sold at a higher price and/or the initial sale of the May contract will be purchased at a lower price.

5. The evidence on returns to position trading is fragmented. A common belief is that the majority of amateurs dabble in futures markets and do not make money. A study of habitual traders—those who keep their accounts open for extended periods and trade regularly—found that their trades win more frequently than they lose, but that after taking account of brokerage fees, they are net losers (Canoles et al. 1998). Professional managers of futures funds also do not seem able to consistently earn returns that are competitive with other alternatives (Chapter 13).

## Discovering Prices

As the foregoing discussion suggests, hedgers or speculators may establish market positions by either buying or selling one or more contracts. It is not necessary to buy futures before selling. The seller simply has assumed a contractual obligation to deliver the specified commodity at the transacted price and to receive that price. The buyer is obligated to accept delivery and pay the contracted price. Of course, both buyers and sellers can offset their initial positions by subsequent transactions.

The number of contracts remaining to be settled is called the *open interest.* For example, if a trader with a net zero position buys a December contract from another trader with a net zero position, the open interest in the December contract increases by one. If this trader subsequently sells a December contract, the open interest is reduced by one contract. The open interest in the market is equal to the net number of either long or short positions. (The number of long and short positions must always be equal.)

The volume of trading is the total number of transactions in a given time period, such as a day or over the life of the contract. It may be quoted in number of contracts or in the physical quantity of the commodity represented by the contracts. A numerical example of volume of trading and open interest is given in Table 12-2.

Volume and open interest for a particular maturity have definite patterns over the life of the contract. When contracts first start trading, volume and open interest are small; over time they grow, often reaching a peak about the time of the expiration of the prior contract. For example, the volume of November soybeans peaks in September. Then, volume and open interest decline as the expiration date approaches. On the last date of trading, open interest must go to zero; open contracts must be offset, or delivery occurs.

A well-developed futures market comes close to exemplifying the economist's concept of a perfectly competitive market. There are many buyers and sellers dealing in a homogenous item, the futures contract. Of course, transactions costs are not zero; the contracts represent fixed quantities of the

**Table 12-2.** Illustration of open interest and volume of trading in a futures market for 3 "days"

| Day | Transaction | Volume during day[a] | Open interest end of day |
|---|---|---|---|
| 1 | A sells 5 to B | 5 | 5 |
| 2 | C sells 10 to D | | |
| | E sells 5 to F | 15 | 20 |
| 3 | F sells 5 to A | 5 | 15 |

[a] Number of contracts.

commodity; and traders do not have perfect knowledge. But, in U.S. commodity markets, much public information exists, and in principle, this information is equally available to all traders. For example, USDA crop-size estimates are released on known dates and time of day. Likewise, price quotations are widely available.

In practice, not every trader will access all available information at the same time, but the diversity of traders in markets suggests that new information will be quickly reflected in market prices. Specialization by traders in active markets assists price formation by bringing diverse knowledge to bear on decisions to buy and sell. Controversy persists, however, about possible over- or under-adjustments of futures prices to new information (more on this later). The point we wish to emphasis here is that prices are being established by many traders with diverse motives, and the price outcomes are quickly and widely disseminated.

Prices of futures contracts in liquid markets change frequently; these changes represent adjustments to the entry of new information into the market or to the transaction effects of the order flow and/or differences of opinion about existing news. The latter type of price changes is sometimes called "noise." A notation to summarize these points is $F_t = E[F_T|I_t]$, where $F_t$ is the current futures price, $E$ is the expected value, $I_t$ is the current information, and $F_T$ is the price at contract maturity.[6] In words, the current futures price is interpreted as the expected price at contract maturity (time $T$) conditional on the current (time $t$) information. Since the futures price at contract maturity should, in theory, equal the cash price for the par (deliverable) grade of the commodity, we can also write $F_t = E[P_T|I_t]$, where $P_T$ is the cash price at the maturity date, $T$. Thus, futures markets can make an important contribution to the price discovery process; they are potentially unbiased forecasts of forthcoming prices. If so, they are an observable measure of the rationally expected price at time $T$.

## Cash-Futures Price Relationships for Grains

Relationships between futures and cash prices for commodities like wheat and corn are explored in this section. They are seasonally produced, continuous-inventory commodities. Two aspects of the behavior of the constellation of prices are of interest: changes in the overall level of prices and changes in the cash-futures price differences. Previous chapters have dealt mainly with changes in the level of cash prices. Now, we add the determination of futures prices and



6. This equation needs to be modified if the current futures price includes a "risk premium." Whether commodity futures prices include a risk premium is a long-standing, controversial issue in the literature. In this chapter, we assume that futures prices do not contain a risk premium. The topic is discussed further in Chapter 13.

the relationship of futures prices for various maturity dates to the cash price. The differences in prices relate to the differences in dates of the maturity of the various contracts relative to the current cash price.

A difference between a futures price and a cash price is referred to as a *basis.* The commercial trade defines a basis by subtracting a futures price from a spot price. A considerable portion of the academic literature, however, defines a basis ($B_t$) as a futures price ($F_t$) minus a cash price ($P_t$), i.e., $B_t = F_t - P_t$. We justify using this definition as a learning device because it makes the economic incentives associated with basis behavior consistent with the economic intuition used in previous chapters. That is, a positive basis for an annually produced grain is associated with a positive incentive to carry inventory. Of course, the only difference in the two definitions is a sign reversal.

Many possible bases exist at any point in time. One reason is that cash prices vary by location and quality attributes. For example, a basis for corn at a specific location is defined as the futures price for a contract for a particular maturity month minus a local cash price (or plus the cash price if the commercial trade's definition is used). A basis is often reported for the price of the nearest-to-maturity contract; this is relevant for spot transactions where the purchase price is equal to the nearby futures price minus the basis. But bases can be computed for any contract. An offer price in a forward contract for harvest-time delivery would probably be based on the price of the immediate post-harvest futures contract adjusted for the estimated, relevant local basis at harvest.

## Normal Seasonal Price Relationships

We start with the assumption of fixed, known supply at harvest time, with inventories being allocated over the entire year to meet known demands. As explained in Chapter 9, cash prices in a perfectly competitive market are expected to rise seasonally by the amount of storage costs. This is represented by the line labeled "Cash" in Figure 12-1. In a perfect market, the prices for the various futures contracts will exactly anticipate the respective maturity months' cash prices. $F_t = E[P_T|I_t]$. Thus, the prices for the futures contracts are represented by the dashed lines drawn parallel to the horizontal (time) axis (Figure 12-1). The differences in prices are related to the costs of storage.

A basis at time $t$ is a measure of the cost of storage from time $t$ to contract maturity, time $T$. Hence, the basis for each contract narrows as the time to delivery diminishes (as $T - t$ gets smaller, for $t = 0, 1, 2, \ldots, T$). But both negative and positive bases are possible, depending on the date on the time axis and on the particular maturity month used to define the basis. In Figure 12-1, the negative bases are associated with the futures prices for the contract maturing in the next (new) crop year, i.e., after the next harvest. A negative basis implies a negative return to carrying inventory from the old crop to the new crop year, and we will have more to say about why positive inventories might be carried even though the return to storage is negative.

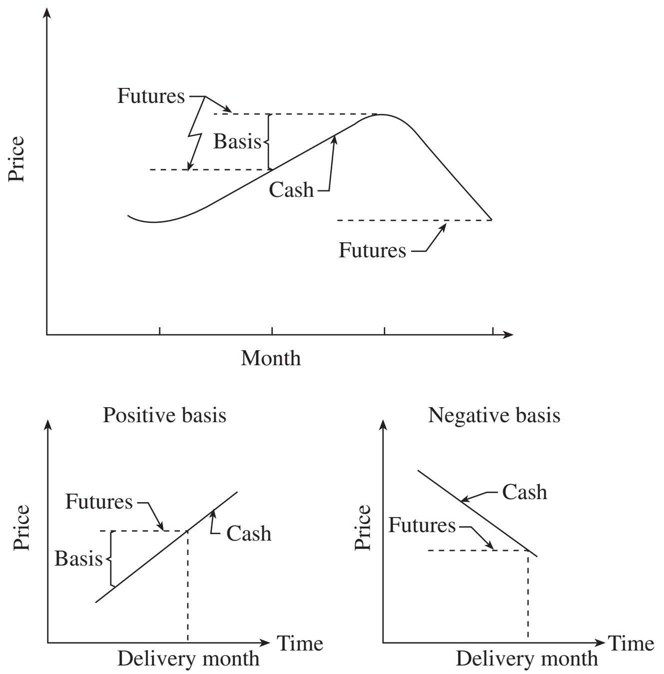



**Figure 12-1.** A seasonal pattern of cash prices with hypothetical futures prices superimposed

All the prices in Figure 12-1 are drawn assuming that they pertain to the same deliverable commodity. In practice, since bases can be computed for cash prices for various locations and grades, the difference between a futures price and a cash price will depend on the specific factors influencing the cash price.

## Determining Futures Prices

We combine models from Chapters 5 and 9, with slightly adjusted notation, to explain the determination of futures prices as well as the level of current consumption, inventory size, and cash prices. In Chapter 5, cash price was determined from the inverse demand function, given a predetermined supply: $P_t = D^{-1}(Q_t, Y_t)$, where $Y_t$ represents the variables influencing current demand and $Q_t$ is the current consumption. In this model, either $Q_t = A_t$ or $Q_t = A_t + I_{t-1} - I_t$, where $A_t$ is the production in time period $t$, $I_{t-1}$ is the beginning inventory, and $I_t$ is the ending inventory. (Note that the letter $I$ is also used to represent the stock of information; here, it is the inventory of grain. The definition of $I$ should be clear from the context of the discussion. We also abstract from the effects of imports and exports.)

In this model, production and beginning inventory are given by prior events, and the cash price reflects the demand for current consumption and indirectly for ending inventory. This is the simplest possible explanation of the price level. Prices change because of exogenous changes in production or in the level of demand.

As noted in Chapter 9, however, decisions about inventories depend on the relationship of the expected price for delivery in the future relative to the current spot price, and thus both prices are affected by current and expected economic conditions. When a futures market exists, the current quotations for future delivery can be treated as market-determined estimates of the expected spot prices for the respective maturity months. In principle, the entire constellation of spot and futures prices could be modeled.

Current production, $A_t$, is treated as predetermined by decisions made by farmers in prior periods (say $t - 1$). Beginning inventory, $I_{t-1}$, is also determined by past decisions. The derived demand for current consumption is modeled as a function the current spot price and exogenous demand shifters, $P_t = D^{-1}(Q_t, Y_t)$. To explicitly model the carrying of inventories, a supply-of-storage function is written in inverse form as $m_t = F_t - P_t = g^{-1}(I_t, Z_t)$, where $m_t$ is the marginal cost of storage over the relevant storage interval and $Z_t$ is the exogenous variables influencing the level of storage supply. With a demand-for-storage equation (to be discussed) and the identity, there are four equations, which determine four endogenous variables: current cash price, current futures price, current consumption, and ending inventories.

Since both spot and futures prices are determined simultaneously and are being influenced by the same set of explanatory variables, changes in these prices should be correlated. At the same time, the prices differ by a cost of storage; the difference between a futures and a cash price, the basis, is also referred to as the price of storage. It is the expected return from holding inventory and hedging. For example, if a merchant buys corn at the prevailing cash price, at the same time sells an equivalent quantity of futures contracts, and holds the grain until maturity (either delivering on the contract or buying futures and selling cash), the merchant earns the amount of the narrowing of the basis as a return to storage. In a competitive market, in equilibrium, this price of storage equals the marginal cost of storage.

The supply-of-storage concept is illustrated in Figure 12-2. As with conventional supply schedules, the quantity of inventory rises with an increase in the price of storage (the basis). Merchants and farmers are willing to increase the stocks they hold as the basis increases. The particular non-linear shape of the supply function is based on the hypothesized nature of the underlying cost functions. Recall that the length of the storage period depicted in the figure is constant; then, the marginal cost of storage pertains to the fixed storage interval implicit in the prices. The costs of storing over this interval depend on conventional costs like those discussed in Chapter 4: direct costs of inputs, such as labor and energy, used in the storage process, and opportunity costs, which depend on interest rates and the value of the inventory.

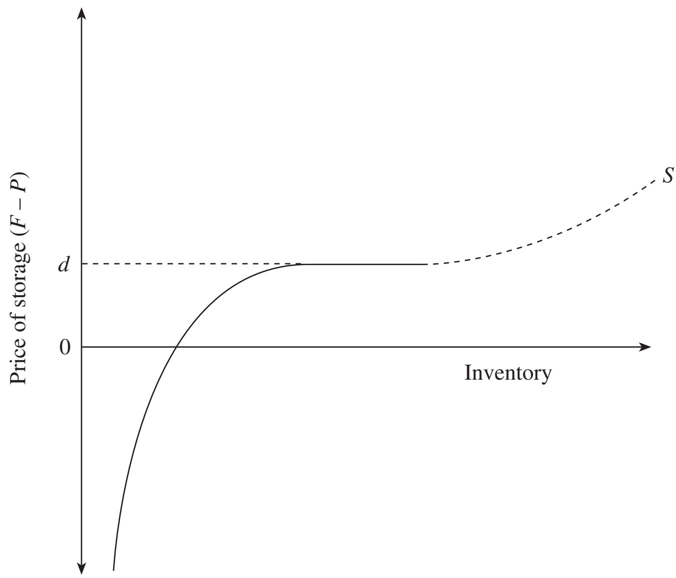



**Figure 12-2.** Hypothetical supply-of-storage relationship, where *d* is the marginal direct cost per bushel stored

In addition, convenience yield is typically incorporated into the cost function (Chapter 9). Convenience yield is a benefit that accrues from having "working stocks" to maintain the continuous operation of a processing plant and/or to meet unexpected demands. Thus, positive inventories are held even when the price of storage is negative. The common belief is that marginal convenience yield decreases and approaches zero as stocks increase. Hence, given that inventories cannot be negative, the supply-of-storage function is thought to be nonlinear (for an alternative explanation see Chavas, Despins, and Fortenbery 2000).

It is possible, however, that when the convenience yield becomes zero, the marginal cost function is relatively flat over a considerable range of inventories, as depicted in Figure 12-2. The addition of more grain to inventory may not increase marginal costs; i.e., diminishing returns are perhaps relatively unimportant in the storage process. But the size of inventories could approach the limit of available space, and if grain stocks must be held out of doors, then spoilage becomes a major potential cost. Thus, we have drawn the dashed portion of the supply function in Figure 12-2 to indicate that it is possible for the marginal cost of storage to turn upward as inventories get large. (In the United States, loss of grain in storage is typically a small and relatively constant proportion of the total inventory. This is less true in some other countries, particularly low-income ones.)

Historically, the supply-of-storage function was thought to be relatively stable (Working 1953; Brennan 1958). Presumably the variables that shift the level of supply changed relatively little, but this is much less true today. In particular, the level of the supply function depends on opportunity costs, which in turn depend on interest rates and the price of the commodity. Interest rates and commodity prices are variable. A rise in interest rates, for example, shifts the supply-of-storage function upward and to the left; i.e., smaller inventories will be held for a given price (basis). The opportunity cost of funds tied up in stored commodities rises when interest rates increase, and therefore an inventory-holder requires a larger basis to make storage profitable.

The model is completed by a demand-for-storage function (Telser 1958). This function depends on the expected future demand for consumption relative to the current demand. Current demand is written as $P_t = D^{-1}(Q_t, Y_t)$, and expected demand is $F_t = D^{-1}(Q_t^*, Y_t^*)$. Recall that $F_t$ is the expected spot price in time $t + 1$, and we use $Q_t^*$ as the expected consumption level in $t + 1$ based on information available in $t$. The second equation can be subtracted from the first to obtain the basis, $F_t - P_t$, as a function of the current and expected levels of the variables in the two equations. The definitions of $Q_t$ and $Q_t^*$ contain $I_t$, and thus with some algebra, a demand-for-storage equation can be written as

$$I_t = f(B_t \mid A_t, I_{t-1}, A_t^*, I_t^*, Y_t, Y_t^*).$$

This function says that the current quantity demanded for ending inventory in period $t$ depends on the current price of storage, $B_t$, holding other factors constant. As in a typical demand equation, the relationship between $I_t$ and $B_t$ is expected to be negative.

The demand shifters include changes in expected production relative to current supply as well as changes in expected demand variables relative to current demand. Thus, the demand for inventory-holding depends on current and expected supplies, as well as on the values of "typical" demand shift variables. Specifically, $A_t$ logically has a positive sign; large current production, *ceteris paribus*, results in a large demand for carrying inventories. On the other hand, $A_t^*$ logically has a negative sign; large expected output in the next crop year, *ceteris paribus*, reduces the demand for carrying stocks into the next year. Analogous arguments can be made for other demand shift variables.[7]

The possible effects of shifts of the supply and demand for storage are illustrated in Figure 12-3. For example, an increase in current production from one

7. We have simplified the discussion by considering expectations only for the next year or time period. In principle, expectations could be held about many periods into the future, but our simplification is probably realistic. One way to handle $I_t^*$, which is the expected ending inventory for period $t + 1$, is to assume that this expectation equals the average inventory in recent years. The USDA provides estimates of farmers' planting intentions, area planted, expected crop size, and current and expected inventories; an important source of information is the monthly *World Supply and Demand Estimates.*

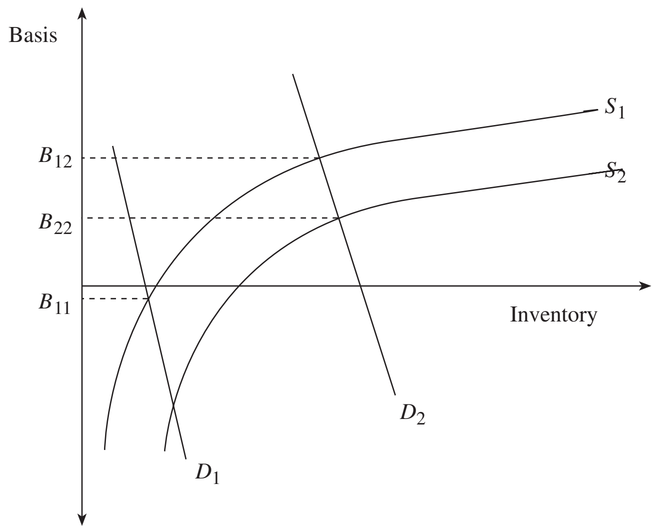



**Figure 12-3.** Changes in demand and supply for inventories

crop year to the next will shift demand to the right, say, from $D_1$ to $D_2$. An exogenous change in interest rates will influence the level of the supply-of-storage function. If interest rates decline, then the supply function will shift downward and to the right, say, from $S_1$ to $S_2$.

To summarize, a conceptual model of the determination of futures and cash prices (and related consumption and inventory variables) can be specified as

$P_t = D^{-1}(Q_t, Y_t)$, inverse demand for current consumption,

$Q_t = I_{t-1} + A_t - I_t$, identity,

$B_t = g^{-1}(I_t, Z_t)$, inverse supply of storage, and

$I_t = f(B_t, A_t, I_{t-1}, A_t^*, Y_t, Y_t^*)$, demand for storage assuming the ending inventory in $t + 1$, $I_t^*$ is a constant.

In this specification, $F_t$ is defined through the basis. In other words, the futures price is determined as a part of the price of storage. (Since $A_t$ depends on expectations held in the prior year, $t - 1$, supply can be modeled as $A_t = S(F_{t-1}, X_{t-1})$, where the futures price, $F_{t-1,}$ is the expectation in $t - 1$ of the spot price that will occur in time $t$.)

## Implications

Variants and extensions of this framework have been used for a variety of purposes. One purpose is simply to explain the year-to-year variability of a particular basis. For example, an analyst might want to explain the variability of the harvest-time basis, say, the variability of a local basis for corn in November. In this case, the basis is defined as the November price of the December futures contract minus the cash price in November, and then the research question is, why does this basis change from year to year? In principle, multiple-equation models could be developed to explain both the intra-year and inter-year behavior of the bases for all of the traded contracts for a grain. Agricultural economists have specified and estimated a variety of models of basis behavior; Sanders and Baker (2012) and Wilson and Miljkovic (2011) are just two of many examples.

It can also be useful to forecast the basis change over a storage period (from time $t$ to time $T$). This forecast provides inventory-holders with an estimate of the gross return to hedged storage. The simplest model, first noted by Working (1953), makes the basis change a linear function of the beginning (time $t$) basis. We illustrate this application in Chapter 16.

The conceptual model also contains general implications for the behavior of futures prices relative to spot prices. Consider, for example, shifts in the demand-for-storage along the supply-of-storage equation. At small inventory levels, it is clear that the cash price can be high relative to the price for future delivery; the basis is a large, in absolute value, negative number. Cash prices are high—they spike—relative to expected forthcoming prices. On the other hand, when inventories are large, arbitrage between cash and futures constrains the futures price to a level that cannot exceed the cash price by more than the marginal cost of storage.

In general, changes in cash and futures prices will be highly correlated. They depend on the same underlying variables.[8] Likewise, the prices for different maturity months are highly correlated. As the model suggests, prices for different time periods are connected through inventory behavior. If expected demand increases relative to current demand, then both cash and futures prices rise because the market is discouraging current consumption and encouraging relatively more stocks to be carried to the future. The linkage in daily price changes is illustrated in Figure 12-4 for four corn contracts. These prices change by almost equal amounts each day, but it is also possible for the relative prices to change as well. The relative price changes reflect new evaluations of the cost of storage from one maturity to the next.

When inventories are relatively small, however, the price of a new crop futures (say, for the December contract for corn as quoted on the prior May 1) can be nearly disconnected from the price of the current crop futures (the May contract for corn). That is, the nearby futures price is very high relative to

8. The model specified in the chapter can be viewed as a set of structural equations that can be solved for the reduced-form equations (see any econometrics textbook). The reduced-form equations make the current endogenous variables (e.g., cash and futures prices) a function of all of the predetermined variables (the $Y$'s and the $Z$'s) in the system. This demonstrates that the futures and cash prices are a function of the same set of variables.

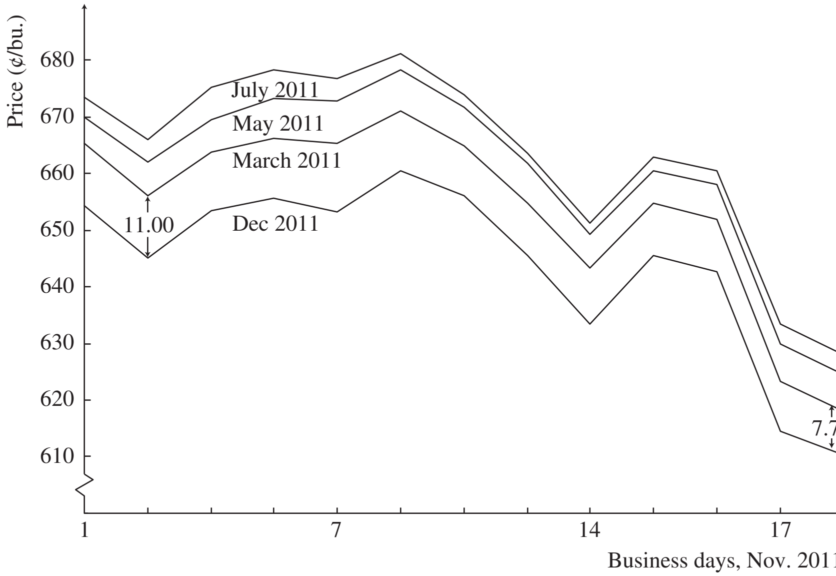



**Figure 12-4.** Settlement prices of four corn futures contracts, Chicago Board of Trade, November 1–18, 2011

the price of the new (after harvest) crop futures. This principle was not well understood by merchants who offered contracts that "promised" to lock in high prices for the new crop, based on the high prices for the current crop, by rolling positions in old crop futures into new crop futures. Nevertheless, as the foregoing conceptual model implies and as empirical work demonstrates (Lence and Hayenga 2001), a very high current price is associated with small inventories and with a negative basis. Thus, it is not feasible to maintain a high price by rolling over a hedge position from old crop futures to new crop futures.

Another implication of the model is that a futures price is an estimate of a conditional expectation; futures prices are a function of the collection of variables—the information set—that determine the price level and the price differences among maturity dates. If futures markets use this information efficiently, then forecasts from formal econometric models, available in the public domain, should not out-perform the forecasts provided by futures quotations.

Since for the grains, futures and cash prices are typically highly correlated, the use of cash prices (naive expectations) versus futures prices may make little difference in the empirical results obtained from (say) estimating a supply equation. The futures market prices are perhaps the conceptually preferred alternative for measuring expectations, but the two measures of price expectations often have similar behavior. An important exception occurs when, as noted above, current prices spike to high levels but the prices for new crop futures do not.

The model of futures prices for the grains can also be used to describe potential patterns of behavior in the *variances* that occur over the life of futures contracts. To crystallize thinking, the reader can visualize the year-to-year variability, say, of the November soybean futures contract by months. For example, we can consider the variability of November soybean futures prices the prior December, or the prior April, or at contract maturity in November. Also, one can visualize the variances of daily price changes as the one moves through time from the initiation of the contract until its maturity.

The variance of cash prices for grains changes over the year (Chapter 9). Prices have the lowest year-to-year variability at harvest time, and this variability increases over the storage season, peaking in the summer months. Thus, since prices of futures contracts converge to the behavior of cash prices at contract maturity, the year-to-year variability of futures prices at contract maturity should be similar to that observed for cash prices in that month (Peterson and Tomek 2005). For example, the variance of November soybean futures in November should be smaller than the same coefficient for the July contract. Also, as discussed in the section on daily price changes, the variance of futures price changes varies over the life of any particular contract. This too relates to the pattern of flow of new information into the market. In sum, the variances of futures prices are expected to have a time-to-maturity effect and seasonal patterns; these effects are well documented for grains and oil seeds in the empirical literature (e.g., Streeter and Tomek 1992).

An implication of the time-to-maturity effect is that the price of a post-harvest futures contract will be less variable at planting time than at harvest time. Thus, hedges that routinely sell post-harvest futures at planting time should reduce the variance of returns relative to routinely selling at harvest. (Few farmers hedge by routinely selling futures at planting time.)

In closing this section, we note that the foregoing discussion assumes no government intervention in the market. In the past, government programs in the United States and elsewhere have influenced the prices of major grains. If a government is supporting prices above equilibrium levels and holding large stocks, the price relationships discussed here will not hold. With government-held stocks, the market is unlikely to provide price incentives for private merchants to hold stocks.

## Price Relationships for Livestock

Futures markets first developed and are perhaps best understood for storable grains, but they also developed for semi-perishable commodities like butter, potatoes, and onions that have inventories. Starting in the mid-1960s, futures markets emerged in the United States for live cattle (grain-fed steers), feeder cattle, and hogs; still later, contracts were initiated for class III and IV fluid milk. Livestock are continuously produced and are perishable in the sense that individual lots of animals will meet the contract specifications only during a short time window. Fluid milk is clearly perishable. While livestock and livestock products have no inventories in the same sense as grains, inventories of processed products, like ham and cheese, exist and can be important in determining the prices of futures contracts for perishables like hogs and milk.

Models of price determination for livestock and livestock products are more complex than those for the grains. Each market must be modeled separately to capture the unique characteristics of the particular commodity. The following two subsections provide a general approach to understanding price relationships.

### Basis Behavior

A simple view of the basis for a livestock commodity is that the cash price reflects current supply and demand (*S* and *D* in Figure 12-5) and that the futures price reflects expected supply and demand (ES and ED) at the maturity date of the contract. Although not apparent from the figure, the cash price and the prices of the various futures contracts are related to each other through the nature of the production process and its economics. In particular, an inventory-like relationship exists in the production process, which is captured by a simple equation, the relationship between the total inventory of animals and production of meat,

$$I_t = (1 + g)I_{t-1} - A_t,$$

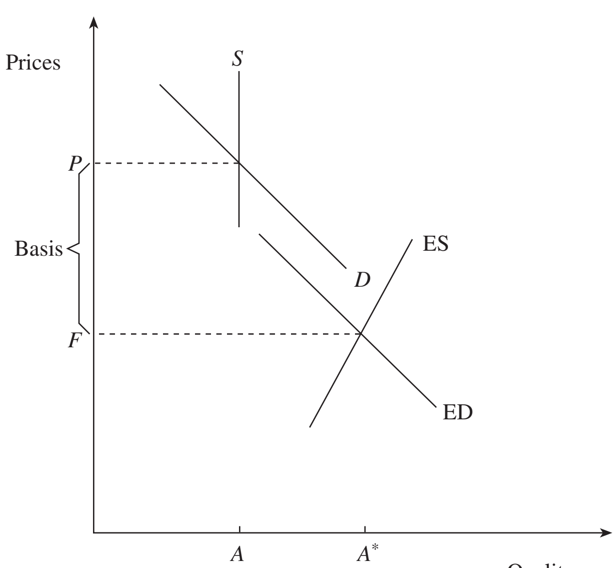



**Figure 12-5.** Current and expected supply and demand relations for live cattle at a point in time

where $I_t$ is the ending inventory of the herd or flock numbers, $I_{t-1}$ is the beginning inventory, $A_t$ is the slaughter (supply of meat), and $g$ is a constant representing the increase in the herd related to the birth, net of death losses, of animals. This equation merely says that ending inventory depends on the output from the stock of animals minus slaughter. Slaughter consists of culling older animals from the breeding herd and of the allocation of younger animals to producing meat.

The complexity of modeling arises from the decisions about culling from the breeding herd and about the proportion of young female animals to retain for the breeding herd versus to place on the track to produce meat. Clearly these decisions influence both current and future production. The economic principle for making these decisions relates to the expected returns relative to expected costs. Both the culling of older females and the retaining of young females depend in important ways on the returns expected from these animals now versus the future as well as on the expected costs of these decisions, including maintaining animals in the breeding herd.

For beef in particular, producers also have choices about the feeding regimes that may be followed and, hence, about how fast animals are grown to marketing weight. These decisions again depend on the expected returns and costs. We can represent current slaughter, given the biological time lags in animal production, as dependent on expectations held in a prior time period. The supply equation is specified as $A_t = S(F_{t-1}, X_{t-1})$. Here the subscript $t - 1$ is used in the generic sense of representing a lag; actual lags may be longer than one time period, depending on the time units. $F_{t-1}$ is, of course, the futures market's representation of expected price in time $t$ when the animals are ready for slaughter. $X_{t-1}$ represents a set of current and expected factors influencing supply; this set of variables can include futures prices for feed grains and protein (say, corn and soybean meal) and the current price of beef. In addition, the number of animal placed in feedlots is constrained by the inventory of animals available to be fed.

Given $A_t = Q_t$, since the output is treated as perishable, cash price is determined by the inverse derived demand function $P_t = D^{-1}(Q_t, Y_t)$, where as before $Y_t$ represents the set of factors determining the level of current demand. So, price is determined recursively by past decisions about supply. Since the decisions made in the past about current supply were influenced by expectations about the future, current price is influenced by expectations. In the context of beef production, current output depends on past decisions about the number of young females (heifers) to retain for the breeding herd relative to the number to be placed in feedlots for beef production. An increase in heifers retained for the breeding herd reduces meat output in the short run, but, *ceteris paribus*, it will increase beef output in future years and hence influence prices in future years. Since producers also have choices about feed rations, the rate of gain and expectations about prices are influenced by these decisions as well.

Using the same equations as before, but advancing the time by one period, $I^*_t = (1 + g)I_t - A^*_t$, implying $A^*_t < (1 + g)I_t$. The decision about the size of $A^*_t$ depends on the supply relation, $S(F_t, X_t)$, and $A^*_t = Q^*_t$. Thus, $F_t = D^{-1}(Q^*_t, Y^*_t)$. In all cases, the asterisks stand for expected levels, and the current futures price is depicted as a function of the expected levels of $Q$ and $Y$ at time $t$. In the terminology of Chapter 9, quantity and prices are being determined recursively, i.e., sequentially, but the current set of futures and cash prices are determined simultaneously. This can be seen by substituting the supply equation for $Q^*_t$ into the inverse demand equation; specifically, quantity consumed in the future, as well as currently, depends on current cash price, as well as other variables.

Based on this model, it follows that the constellation of cash and futures prices (and the corresponding bases) reflect expectations about supply and demand at various points in time. Since current and future supplies are influenced by both current and expected prices, the constellation of prices is expected to be correlated. But, because output is continuous and perishable and because producers have some flexibility in varying production rates (within the constraint of the total stock of animals), the correlations among livestock futures prices are usually smaller than for the grains.

Futures prices for live cattle for selected delivery months are plotted in Figure 12-6 for a 15-day period. The prices are clearly correlated, but prices on some

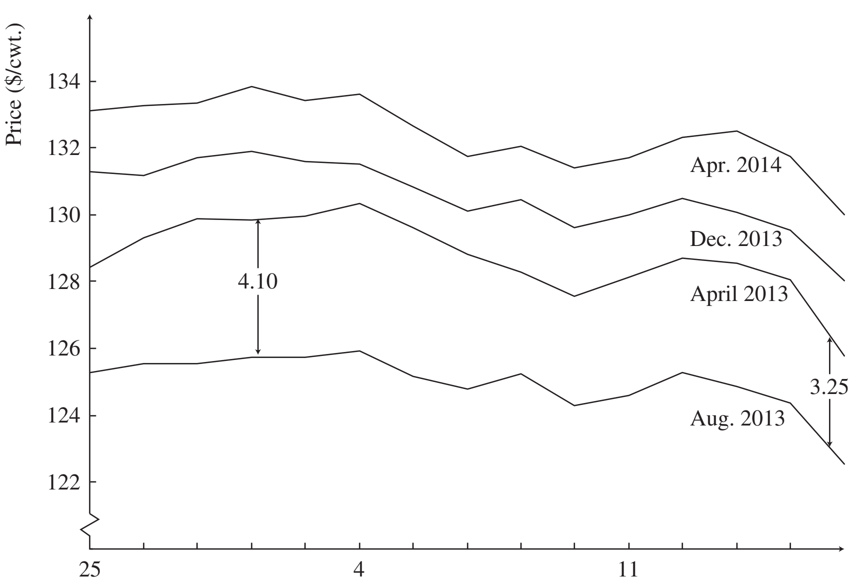



**Figure 12-6.** Settlement prices of live cattle futures contracts, Chicago Mercantile Exchange, February 25–March 15, 2013 (business days)

days move in opposite directions. Even when the changes are in the same direction, the magnitudes of the changes can differ.

Indirect measures of expected supply for various livestock commodities are available in the United States. Leuthold and Peterson (1983) took advantage of this fact to model futures and cash prices for live hogs. Hog contracts mature every other month, and a set of equations was fitted for different maturity dates using different measures of expected supply. For example, for a model of the current cash price and the futures price for the contract maturing 6 or 7 months in the future, expected supply was measured by two variables: the number of sows farrowing and the number of pigs in the weight category under 60 pounds. Current supply is measured by the number of hogs slaughtered in the current month. Since the inventory of processed pork products is likely to be an important determinant of prices, the quantity of cured and frozen pork in cold storage at the beginning of the month is also included in their model.

In the notation of this book, the Leuthold and Peterson model can be characterized as

$$P_t = f_1(A_t, A^*_t, I_{t-1}) \text{ and } F_t = f_2(A_t, A^*_t, I_{t-1}),$$

where $I_{t-1}$ stands for inventory of cold-storage holdings at the beginning of the month. In principle, variables to measure the current and expected levels of the demand for pork could also be included in the model. These equations suggest that the basis depends on the same variables.

Less empirical evidence is available about the probability distributions of prices of livestock and livestock products than for the grains. It is highly likely that the variance of prices of livestock futures has a time-to-maturity effect, namely that, *ceteris paribus* the variance of prices grows as the contract maturity date approaches. Historically, cash prices for livestock and livestock products have also had seasonal patterns, but they have probably dampened over the years (Chapter 9). To the extent that there is seasonality in supply and prices, one can expect seasonality in the variance of prices. Whether or not this seasonality remains statistically important under current conditions is an empirical question.

## Price of Feedlot Services

Bases for livestock commodities are not measures of a price of storage. Rather, the lapse between current time $t$ and contract maturity time $T$ involves a transformation or service. A small animal, a “feeder,” is transformed into a mature animal to be slaughtered for meat, and futures prices can be used to evaluate the expected returns from feeding cattle or hogs. Paul and Weston (1967) coined the term *price of feedlot services* to describe a current cash-futures price spread that can be interpreted as a return to a feedlot operator for providing the service of producing a market-weight animal from the inputs of the feed and feeder animals. Basically, it is a difference between the futures price of live (grain-fed) cattle and the principal variable costs of producing these cattle. Specifically, $S_t = F_t - C_t$, where $F$ is the quote for the live cattle futures that matures at the end of the feeding period, $C$ is the cost of feeder cattle and of the feed used to transform these feeders to market-weight animals, $S$ is the price of feedlot services, and $t$ is the time subscript indicating that all price are observed on the same date, presumably near or at the beginning of the production period.

These prices and returns are stated in dollars per hundred pounds of the fully fed (say, 1,200-pound) steers. For example, if a 700-pound feeder steer costs $900 and if the feed costs to transform it into a 1,200-pound marketable animal are $300 at time $t$, then these major variable costs sum to $1,200 per animal, or $100 per 100 pounds of a 1,200-pound steer. If the current futures price for delivery in 6 months is $125 per 100 pounds, then the price of feedlot services is $25 (= 125 − 100) per cwt.

Paul and Weston's main point is that all these prices are known at the beginning of the production period, time $t$. The residual amount, the price of feedlot services ($25 per 100 pounds) must cover other costs, like veterinarian fees, wages, return to the operator's labor, management, and capital. If a manager determines that the return is adequate, then it can be approximately assured through hedging, i.e., by selling live cattle futures at $125, assuming the feeder cattle and feed have been or will be purchased at current prices. (In practice, the futures price must be adjusted by an appropriate local basis to estimate the price being determined by the sale of futures.) If these major inputs have not been purchased, they can also be set by buying the relevant futures contracts.

The return for providing a feedlot service will decline as the length of the feeding period is reduced. For example, if a 950-pound animal is fed for 3 months, less service is provided than if a 650-pound animal is fed for 6 months. As the length of the feeding period is reduced, relative prices should provide a smaller return; Paul and Wesson (1967: 39) showed that this is indeed the case.

The prices of full-fed steers, feeder animals in various weight categories, and feed are interrelated. On average, relative prices should provide competitive returns to feedlot operators, ranchers, and grain producers. Price adjustments are such, however, that relative prices at various points in time may provide negative returns or positive rents above those expected under competition. This relates to the lengthy times required for arbitrage to be effective, given the biological nature of the production process.

The number of placements of feeder cattle into feedlots is expected to be positively related to the price of feedlot services; i.e., the number of placements increases as the margin increases. But placements are limited by the total inventory of animals, in various weight categories, available to put on feed at any point in time. Indeed, if margins are wide, young female animals may be withheld from the market and placed into breeding herds, thereby reducing the short-run supply of beef (as previously discussed). When profitable margins exist, futures markets permit this margin to be realized. The producer need not wait until the cattle reach market weight before pricing them. But, as noted before, existing price relationships are not always profitable.[9]

## Daily Price Changes

In an efficient, competitive market, $F_t = E[F_T|I_t]$. Pricing efficiency means that current prices reflect current information (Chapter 13). Since this is true each day, it follows that $F_t = E[F_{t+1}|I_t]$, where the subscripts denote different days. Alternatively, $E[F_{t+1} - F_t|I_t] = 0$. That is, if the market is efficient in the sense that the current price reflects current information, the expected price change is zero.

Of course, in practice, new information will enter the market on day $t + 1$, and consequently, prices change from day to day (and within the day). The notation merely states that the current price reflects current information and, hence, that no reason exists for expecting (forecasting) any particular price change. Since prices will change from day to day, this is denoted by an error term, as follows: $F_{t+1} - F_t = e_{t+1}$. From the previous discussion, the expected value of this error term is zero, and the error term has a probability distribution.

Two broad questions arise in analyzing and using the time-series observations on futures price changes. First, what is the nature of the flow of new information into the market, and second, what is the process of adjustment to this new information? Although daily price changes can be viewed as adjustments to new information, the flow of information and the process of adjustment can differ from market to market. Hence, specific answers to the two questions require an empirical analysis of market data, and the statistical principles and methods involved are beyond the scope of this book. Nonetheless, the questions are important. Can a historical series of price changes be used to make profitable forecasts of forthcoming changes?

### Price Responses to New Information

Models of price changes often start with the notion of a perfect market and assume information is available at no cost. Current price reflects all known information, including information about expected supply and demand conditions. New information, which occurs randomly (otherwise it could be forecast), is reflected instantly and correctly into a new price.

9. In theory, it would seem that producers holding rational expectations would take countercyclical action to offset (arbitrage away) the cycles. But forecasts are imperfect, and the costs of adjustment are considerable. The price quotations for futures contracts are a type of forecast, but all sources of forecasts tend to be imprecise for periods more than 3 or 4 months in the future (Tomek 1997).

Although futures markets are not perfect, they are usually competitive. Thus, it is plausible to assume that changes in daily prices (observed 24 hours apart) are a random series. This implies that complete adjustment to new information takes place within a day. If so, the best forecast of tomorrow's price is today's price. $F_{t+1} = F_t + e_{t+1}$. This equation, with some added assumptions about the error term, is called a *random walk* model. In a pure random walk model, however, prices would be free to wander anywhere, but prices cannot be negative. Futures prices must terminate at a positive value at contract maturity. For this and other reasons, empirical analyses of futures prices often use the logarithms of prices. For simplicity, we continue to use notation for the observed prices.

The foregoing model is not appropriate if markets are pricing inefficient. An inefficient market exists if price changes do not fully reflect information changes. Of course, there are costs of arbitrage. Hence, a more complete definition of an inefficient market is one in which profitable arbitrage opportunities persist; current and past information can be used to make economically significant forecasts. This implies, for example, that a model like $F_{t+1} = a + b_0F_t + b_1F_{t-1} + u_{t+1}$ has been estimated from historical data, that it is a statistically significant relationship, and that, taking account of transactions costs, the equation can be used to make profitable trades.

The principal reason for inefficiency is perhaps that poorly informed traders dominate price formation, but this case should be differentiated from appropriate decisions based on erroneous information. Also, "inefficient" price adjustments may occur for reasons other than poorly informed traders (discussed in Chapter 13).

Empirical studies of pricing efficiency have produced mixed results. It is perhaps fair to say that the majority of academic analyses find that futures markets are relatively efficient. If so, it is not possible to make speculative profits consistently by using trading rules based on historical patterns of price behavior. Nonetheless, most of the non-academic literature on speculation is built on the premise that profitable trading rules can be found in historical price patterns (e.g., Kleinman 2001). An analogous debate relates to whether or not farmers can enhance profits by following speculative pricing strategies (e.g., Wisner, Blue, and Baldwin 1998 versus Zulauf and Irwin 1998).[10]

## Effects of Information Flows

The discussion thus far has focused on price responses to new information. The nature of the flow of information is an equally important variable in

10. The definition of *pricing efficiency* used in this book does not preclude a few traders from having superior knowledge and profiting from it. They may obtain information sooner and/or have better models than the vast majority of traders. This is private information, and the profits from such information are likely to be earned within minutes or a day, not over many days.

influencing the probability distributions of daily price changes. It is not a contradiction to say that individual pieces of information (events) are random but that the overall flow of information events has patterns. For an annual crop, the amount of information about expected crop size follows a seasonal pattern. By definition, events influencing crop size occur during the growing season, and many such events can occur. The individual events have little or no predictability, but their importance and frequency have a seasonal pattern.

Inventories of seasonally produced crops decline, of course, over the marketing year, and year-ending inventories vary. This implies that the magnitude of the effect on price of a given piece of new information (about current or expected supply and demand) can vary within the marketing year and from year to year. The smaller the inventory, the smaller the flexibility to respond to, say, new information about expected demand. It follows that the probability distributions of prices can have an increasing variance and skewness as the marketing season progresses.

Information is likely to have a time-to-maturity component as well (Samuelson 1965). This hypothesis arises from the commonsense observation that the amount of information impacting the price of a particular contract tends to increase as the maturity date approaches. Thus, the variance of commodity futures prices can trend upward as maturity approaches.

These patterns do not affect the expected price change, which is zero, but they do affect the variance and probably the skewness and kurtosis parameters of the distributions of price changes. Consistent with earlier discussions, the variances of daily price changes are probably related. For example, if we observe the variance of the December corn futures on days in the prior January and then follow this variance through time until maturity, it is probable that the variance will have a seasonal peak in mid-summer, which is imposed on a generally rising trend. It is also possible that the skewness increases from winter to summer and then declines as the new harvest approaches (Peterson and Tomek 2005).

In sum, daily changes in futures prices are *uncorrelated* but *not independent.* Today's price change is not related to yesterday's price change, but the variance of the today's price change is related to yesterday's. This is illustrated schematically in Figure 12-7. A plot of a variance against time is given for two futures contracts for a grain. Both variances have an upward trend and a seasonal component. The contract maturing in the summer has a larger variance at maturity than does the contract maturing after harvest. The kurtosis and skewness parameters may also have systematic behavior. Thus, although it would be convenient statistically if the probability distributions of price changes for commodity futures contracts were normal, they probably are not. If prices are converted to logarithms and the differences of the logarithms are used, this allows, to some degree, for skewness in the original prices, and in principle, one could model the variance of a log-normal distribution.

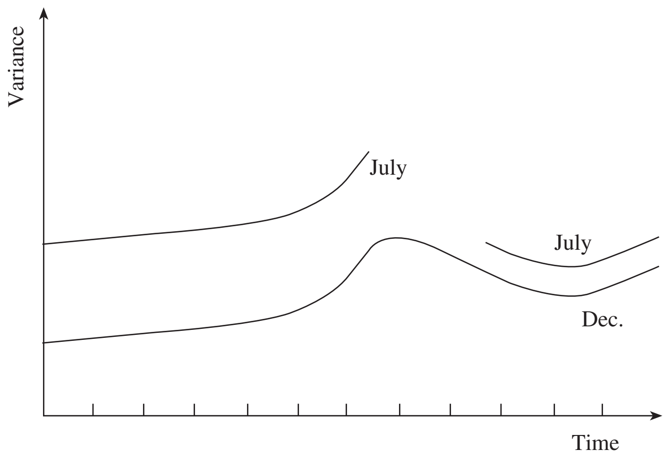



**Figure 12-7.** Patterns of behavior of the variances of price changes for two grain futures contracts

We also remind readers of the need for clarity in distinguishing between the behavior of cash and futures prices. Daily cash prices are likely to be related through time; both the prices and their variances have systematic patterns (Chapter 9). In contrast, in this chapter we argue that the expected daily changes of futures prices are uncorrelated for any single contract, although the variances do have systematic behavior. The cash price has a clear pattern of behavior, but the individual futures prices anticipate the respective maturity-month prices (Figure 12-1).

Confusion can arise in analyzing futures prices when prices for different contracts are combined to create a long time series. A common procedure is to link the prices of the nearest-to-delivery contracts. Since futures prices are the expected spot prices at maturity, the use of nearby contract prices creates an approximation of the underlying cash prices. Analysts may find patterns of behavior in the pooled futures data that mimic cash prices. The difficulties of studying systematic behavior in futures prices is further discussed in Irwin, Zulauf, and Jackson (1996).

## Prices of Options Contracts

A large literature on the pricing of options has emerged since the fundamental papers of Black and Scholes (1973) and Merton (1973) were published.[11] Although the concepts have broad applicability to the many types of options

11. Scholes and Merton received the Nobel price in economics in 1997 for their contributions to the theory of pricing options, which emphasizes the importance of their paper. Fischer Black died in 1995; otherwise, he would undoubtedly also have been a co-recipient of the prize.

that exist in any economy, much of the literature emphasizes the pricing of options on individual corporate stocks and financial instruments. The literature on pricing options on commodity futures (or on physical commodities) is less extensive. In this section, we discuss the linkage between the prices of commodity futures contracts and prices of options-on-futures contracts, but we do not discuss a formal model.

An option's price (premium) depends on four factors: the level of the strike price relative to the current price of the underlying asset, the length of time until the option's expiration (the magnitude of $T - t$), the magnitude of the variability of the prices of the underlying asset, and an interest rate. Options pricing models formalize these factors. For example, the Black-Scholes model assumes, *inter alia*, that the price (returns) of the underlying asset has a log-normal probability distribution with a constant variance, transactions costs are zero, no riskless arbitrage opportunities exist (competitive markets), and the risk-free rate of interest is constant over the life of the option.

The strike price specified in the option can be below, equal to, or above the current price of the underlying futures contract. If, for a put option, the strike price is above the current futures price, the option has "intrinsic value," or it is said to be "in the money." A put is "out of the money" when the strike price is below the current market price. Without arbitrage between markets, the buyer of an in-the-money put could exercise it, thereby obtaining the difference between the strike price and the futures price as a profit. But it is precisely the ability to arbitrage between the two markets that helps determine the price of an option. The higher the strike price relative to the current futures price, the larger the premium for a put option, other factors held constant. For a call option, the opposite is true (Table 12-3).

The longer the time to the expiration of the contract (the larger $T - t$ is), the larger the probability that the price of the underlying asset will change before the expiration of the contract, *ceteris paribus.* Such a price move could mean that an option that initially was out of the money becomes profitable to exercise or that an option that was already in the money moves more deeply into the money. If the probabilities of price increases and decreases are equal, then the time value should be the same for puts and calls, and, *ceteris paribus*, the longer the time to maturity, the larger the option premium (Table 12-3). The seller of the option must be compensated for the increased risk that the option might be exercised.

The probability that a futures price will change over any fixed length of time depends on its variability, i.e., on its probability distribution. If price changes are log-normally distributed, then the volatility of the futures prices is measured by a variance (or standard deviation). An increase in volatility, *ceteris paribus*, results in an increase in the price of the option, and vice versa. Volatility and options prices are positively related because the probability of exercise increases with the variability of the underlying price. In pricing options on

**Table 12-3.** Prices of options on two soybeans futures at alternative strike prices, February 25, 2013 (cents per bushel)

| | Maturity month | |
|---|---|---|
| | May 2013 | July 2013 |
| Futures close | 1,452¼ | 1,435¼ |
| Strike prices | | |
| Calls | | |
| 1,420 | 63⅝ | 75¼ |
| 1,440 | 52¼ | 65⅜ |
| 1,460 | 42⅜ | 56⅞ |
| 1,480 | 34⅛ | 49⅛ |
| 1,500 | 27 | 42⅜ |
| Puts | | |
| 1,420 | 31⅜ | 60 |
| 1,440 | 40 | 70⅛ |
| 1,460 | 50⅛ | 88½ |
| 1,480 | 61⅞ | 93¾ |
| 1,500 | 74⅝ | 110¼ |

commodities, the characterization of the probability distributions is, as stated earlier, a complex issue, and the assumption of log normality is probably an inadequate characterization of commodity price behavior. Hull (2009: 282f.) provides an introduction to estimating a volatility parameter from historical data given the assumption of a log-normal distribution with a constant variance. Plausible models for options on commodity futures contracts probably differ for different commodities.

A common simplifying assumption in options pricing models is that an option can be exercised only on its maturity date—a European option. In contrast, options on commodity futures traded in U.S. markets can be exercised at any time in their life, called an early exercise feature or an American option. If the early exercise feature is ignored, then it is relatively easy to understand that the current price of an option is the present value of its expected benefit to the buyer on its expiration date. An interest rate is relevant in options pricing to determine a present value. The interest rate and the option's price are inversely related. The larger the interest rate, the lower the present value of a given expected return at the fixed future date.

The available empirical evidence suggests that the magnitude of the interest rate is relatively unimportant in determining option premiums. Typically the time to maturity of an option on futures is relatively short (months, not years). The early exercise feature of American-type options presumably raises the options premiums slightly relative to comparable prices of European options.

Premiums for put and call options for soybean futures traded on the Chicago Board of Trade are illustrated in Table 12-3. All the prices pertain to the settlements for February 25, 2013, a point in time. The table presents premiums for selected strike prices at 20-cent intervals. Options exist for higher and lower strike prices than those shown and sometimes are traded for 10-cent intervals. The first point to note, therefore, is that options can be traded for a wide range of alternative strike prices.

On February 25, 2013, the price of a call on May soybean futures at a strike price of 1,460 cents per bushel was 42.75 cents. A buyer of this option would have paid a $2,137.50 premium ($0.4275 × 5,000 bushels) for the right to buy 5,000 bushels of soybeans at 1,460 cents per bushel. The strike price is a little above the current market price of May soybeans, 1452.25 cents per bushel. The other important factors determining the premium for this call were the time remaining to maturity and the volatility of soybean prices for the May contract. The price of the comparable put is a bit larger, 50.124 cents, because the put with a 1,460 strike was slightly in the money; it was conveying the right to sell at 1,460 when the market prices was 1452.25 cents.

The data in Table 12-3 demonstrate that the price of a call declines as the strike prices rise relative to the current market price of the futures contract, while the reverse is true for puts. The table also illustrates that the longer the time to the expiration of the option, the larger the option premiums. For example, the near at-the-money call option on July futures (strike of 1,440) had a price premium of 65.375 cents compared with the nearest to at-the-money call on May futures (strike of 1,460), which had a premium of 42.375 cents. Of course, since the table reports prices for a fixed date, it is merely a snapshot and cannot illustrate the consequences of changes in expectations about the volatility of prices. As the expiration date approaches, the premium will decline because the probability of price changes that would result in exercise declines.

Formal models are available to determine the theoretically appropriate price of an option, given the assumptions of the model (as in Hull 2009). Such prices can be compared with the market-determined prices of traded options. Differences between the model's estimated price and the market's price can occur for two reasons: either the model is inappropriate for the particular market application or the market is inefficient. Traders in options markets use options pricing models to identify potential arbitrage opportunities, but they know that the models are not perfect.

Market prices of options can also be used to derive information about the implied volatility of the underlying asset. If a model of an options price is assumed to be correct, the market-determined price is also accepted as correct, and a relevant interest rate and the known time to expiration are given, the only unknown in the model is the volatility parameter. Thus, the model can be solved for this unknown. In other words, instead of estimating a volatility from historical data and estimating an option premium, the market-based premium is accepted as correct and the model is solved for the implied volatility. (Since options prices exist for different strike prices, there is also the issue of the internal consistency of the implied volatilities estimated from the various prices, but this topic is beyond the scope of this book.)

Finally, we remind readers that the prices of the physical commodity, prices of futures contracts, and prices of options-on-futures are surely interrelated. Fackler and Tian (1999) estimate the time structure of the volatility of futures prices and the volatilities implicit in option premiums from an underlying model of spot price behavior. The results are consistent with the time-to-maturity and seasonal effects alluded to earlier.

## Summary

Futures and options markets are organized trading venues for standardized contracts. Futures contracts carry a legal obligation to make or take delivery, while the buyer of an option contract obtains a right, but not an obligation, that can be exercised within a specified time period. Futures and options markets are usually competitive, but of course they are not perfect.

Price changes arise because new information enters the market. If futures markets are efficiently using information, the market's price for a particular maturity month is equivalent to a forecast that should perform as well as, or better than, forecasts from formal statistical models available to the public.

The relationship of futures prices to cash prices is defined in terms of basis relationships, and an understanding of basis behavior is important to managers who use futures or options-on-futures contracts to hedge. The next chapter elaborates this point as it discusses the functions of futures markets.

## References

Black, F., and M. Scholes. 1973. "The Pricing of Options and Corporate Liabilities," *J. Pol. Econ.* 81:637–659.

Brennan, M. J. 1958. "The Supply of Storage," *Am. Econ. Rev.* 48:50–72.

Canoles, W. B., S. Thompson, S. Irwin, and V. G. France. 1998. "An Analysis of the Profiles and Motivations of Habitual Commodity Speculators," *J. Fut. Mkts.* 18: 765–801.

Chavas, J-P., P. M. Despins, and T. R. Fortenbery. 2000. "Inventory Dynamics under Transactions Costs," *Am. J. Ag. Econ.* 82:260–273.

Fackler, P. L., and Y. Tian. 1999. "Volatility Models for Commodity Markets," in *Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*, T. C. Schroeder, ed. Chicago. Pp. 247–257.

Hull, J. C. 2009. *Fundamentals of Futures and Options Markets*, 7th ed. Upper Saddle River, NJ: Prentice Hall.

Irwin, S. H., C. R. Zulauf, and T. E. Jackson. 1996. "Monte Carlo Analysis of Mean Reversion in Commodity Futures Prices," *Am. J. Ag. Econ.* 78:387–389.

Kleinman, G. 2001. *Commodity Futures and Options: A Step-by-Step Guide to Successful Trading*. London: Prentice Hall.

Kuserk, G. J., and P. R. Locke. 1993. "Scalper Behavior in Futures Markets: An Empirical Analysis," *J. Fut. Mkts.* 13:409–431.

Lence, S. H., and M. L. Hayenga. 2001. "On the Pitfalls of Multi-Year Rollover Hedges: The Case of Hedge-to-Arrive Contracts," *Am. J. Ag. Econ.* 83:107–119.

Leuthold, R. M., and P. E. Peterson. 1983. "The Cash-Futures Price Spread for Live Hogs," *North Central J. Ag. Econ.* 5:25–30.

Merton, R. C. 1973. "Theory of Rational Options Pricing," *Bell J. Econ. & Mgt.* 4:141–183.

Paul, A. B., and W. T. Wesson. 1967. "Pricing Feedlot Services through Cattle Futures," *Ag. Econ. Res.* 19:33–45.

Peterson, H. H., and W. G. Tomek. 2005. "How Much of Commodity Price Behavior Can a Rational Expectations Storage Model Explain?" *Ag. Econ.* 33:289–303.

Samuelson, P. A. 1965. "Proof That Properly Anticipated Prices Fluctuate Randomly," *Ind. Mgt. Rev.* 6:41–49.

Sanders, D. J., and T. G. Baker. 2012. "Forecasting Corn and Soybean Basis Using Regime-Switching Models," in *Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. St. Louis, MO. www.farmdoc.illinois.edu/nccc134/paperarchive.html.

Streeter, D. H., and W. G. Tomek. 1992. "Variability in Soybean Futures Prices: An Integrated Framework," *J. Fut. Mkts.* 12:705–728.

Telser, L. G. 1958. "Futures Trading and the Storage of Cotton and Wheat," *J. Pol. Econ.* 66:233–255.

Tomek, W. G. 1997. "Commodity Futures Prices as Forecasts," *Rev. of Ag. Econ.* 19:23–44.

Wang, X., P. Garcia, and S. H. Irwin. 2012. "The Behavior of Bid-Ask Spreads in the Electronically Traded Corn Futures Market," in *Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. St. Louis, MO. www.farmdoc.illinois.edu/nccc134/paperarchives.html.

Wilson, W. W., and D. Miljkovic. 2011. "Dynamic Inter-relationships in Hard Wheat Basis Markets," in *Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. St. Louis, MO. www.farmdoc.illinois.edu/nccc134/paperarchives.html.

Wisner, R. N., E. N. Blue, and E. D. Baldwin. 1998. "Preharvest Marketing Strategies Increase Net Returns for Corn and Soybean Growers," *Rev. of Ag. Econ.* 20:288–307.

Working, H. 1953. "Hedging Reconsidered," *J. Farm Econ.* 35:544–561.

Zulauf, C. R., and S. H. Irwin. 1998. "Market Efficiency and Marketing to Enhance Income of Crop Producers," *Rev. of Ag. Econ.* 20:308–331.

CHAPTER 13

# Functions of Commodity Futures Markets

The general function of futures markets is to facilitate various types of resource allocation through hedging and through the provision of forward prices. This chapter discusses both the hedging and price discovery roles, and in the process, some issues related to price behavior are also discussed.

## Hedging

In an uncertain world, economic activity has risks. For example, a farmer who commits seed, land, and other resources to wheat production is subject to yield and price risks. Actual yields may be below expected yields, and likewise, the price expected at planting time may not be realized at harvest time. A variety of economic institutions exist to price commodities and ameliorate risks. The wheat farmer perhaps can negotiate a forward contract at planting time that contains a selling price that assures a competitive profit. This shifts the price risk from the farmer to the buyer, and the buyer can, in turn, use a futures market to cover the price risk of offering a fixed price in a forward contract.

Futures markets are simply a particular kind of forward contracting institution; standardized contracts are traded under well-known rules (Chapter 12). If the market is functioning well, it provides the least-cost way for some (but not necessarily all) firms to enter into and exit contracts. In the process, prices for future delivery are established in a competitive bidding process, and this information is widely available to all interested parties. Since the contracts are standardized, hedging in futures typically involves a temporary position in futures contracts in anticipation of the transactions that will occur in the cash market (Working 1953).

The early literature emphasized that the principal role of hedging in futures markets was to shift price risk. Subsequently, Working stressed the profit motive. Now, economists often assume that a firm's objective is to maximize utility.

Utility may be viewed as function of both expected profits and the riskiness of these returns. The addition of a futures position to a firm's portfolio can help manage risk by assuring (approximately) a given level of returns. The next several subsections discuss alternative types of hedges and introduce the notion of "optimal" hedge levels.

### Arbitrage Hedges: Temporal Allocation of Inventories

Since futures trading in the United States originated for seasonally produced commodities, with continuous inventories, perhaps the most emphasized role of commodity futures markets has been that of helping allocate inventories through time. At any point in time, markets provide price quotations for a constellation of futures contracts as well as the current cash prices. The differences in prices for different maturities are the "prices of storage," which can be compared with the marginal costs of storage for the various time intervals (Chapter 12). Thus, decisions about carrying inventory and hedging can be based on observable futures prices relative to prices for immediate delivery.

The price of storage is a return that is essentially assured through hedging; it is an incentive or disincentive to store. A hypothetical example illustrates the point. A merchant observes the following price relationships for wheat on September 1:

cash at \$7.00 per bushel, and
March futures at \$7.40 per bushel.

The positive price of storage for 6 months (September 1 to March 1) is 40 cents per bushel. If the expected convergence of the two prices by March 1 is sufficient to cover storage costs, then the merchant can assure the return by hedging. The mechanics of the hedge are to buy and store cash wheat and simultaneously sell the equivalent quantity of March wheat futures. Subsequently, when the wheat is sold out of storage, the futures position is offset by buying the March futures.

If the cash price paid by the merchant pertains to the par grade and location of wheat specified in the futures contract and if the cash and futures markets are perfectly competitive, then the two prices will converge to the same amount at contract maturity. The merchant is assured of the 40-cent return by hedging. If, for instance, the price was exactly \$7.40 on March 1, the return from the futures position is zero (the short position is sold for \$7.40), and the cash price will have risen from \$7.00 to \$7.40. But the 40-cent return is not contingent on a seasonal rise in cash prices. If the two prices decline to \$6.75, the hedger still earns the 40-cent return. The final transactions on March 1 are as follows:

cash wheat sold at \$6.75 for a 25-cent loss, and
March futures purchased at \$6.75 for a 65-cent gain.

The 25-cent loss in the spot market is offset by the 65-cent gain on the futures position; so the return is still 40 cents per bushel.

The same result occurs if spot prices rise, but in this case, the hedge prevents the firm from earning the full amount of the increase in cash prices. For example, if the spot price rises from \$7.00 to \$7.60, the 60-cent gain is partly offset by 20-cent loss on the futures contract, whose price will have risen from \$7.40 to \$7.60 at contract maturity.

In practice, the typical merchant does not intend to hold inventory until contract maturity and make delivery on the contract. Rather, the merchant will be seeking opportunities to sell wheat at a gain above that implied by the initial basis. This possibility exists because the passage of time increases the probability of a buyer appearing who needs grain of the quality and in the location that the merchant can supply. The demand and supply of the particular attributes can be matched. In other words, as time passes, not only does the average level of market prices change, but the penumbra of individual transactions prices fluctuates around the central tendency. All merchants are seeking to conclude their transactions at a relatively favorable basis change. The term *arbitrage hedge* means that the decision to hedge is based on the expectation of a favorable change in relative prices over the storage interval.

Likewise, the expected change in relative prices may not provide an incentive to hedge. It is possible that storing *and* hedging would lock in a negative return. A small inventory will be carried, however, in the face of negative returns to hedged inventory because a few merchants and processors see potential benefits from storage (Chapter 12). This line of reasoning, it should be noted, assumes that some firms will see a value in holding inventory based on the possibility of favorable price changes in the future, notwithstanding the apparent negative return implied by the relative prices.

In sum, the variability in the basis (the price of storage) provides varying incentives for carrying inventories. A positive basis that equals the marginal cost of storage provides an incentive to hold stocks and to hedge. Negative bases reflect small current inventories relative to forthcoming supplies and a disincentive to carry stocks.

The economics of storage for an individual firm can be formalized through a simple profit equation.

$$\text{Profit} = (P_{t+k} - P_t)\mathrm{I}_t + (F_t - F_{t+k})X_t - mI_t,$$

where $P$ is the cash price; $F$ is the futures price; $I$ is the physical quantity stored; $X$ is the size of position in futures; $m$ is the per unit cost of storage over the interval $t$ to $t + k$; and the subscripts indicate two time periods, designating the beginning and ending of the storage period. Clearly $k$ can vary, and if it does, then $m$ will change; $m$ pertains to a fixed interval.

The setup of the equation assumes that merchant potentially will buy grain at $P_t$ and simultaneously sell futures at $F_t$. This is the logical transaction if the prices are positively correlated and if the price difference reflects a profitable storage opportunity. If $P_{t+k} = F_{t+k}$ and if $X_t = I_t$, then with some algebra,

$$\text{(1)} \qquad \text{Profit} = (F_t - P_t)I_t - mI_t,$$

where $F_t - P_t = B_t$, the basis at time t. Because of the assumptions, namely perfect convergence and known cost of storage, a hedge will exactly earn the time *t* basis. In terms of economic principles, a profit-maximizing firm in a perfectly competitive market stores and hedges when marginal revenue equals marginal costs. In this example, marginal revenue is defined by the initial basis and marginal cost is equal to *m*. Put another way, taking the first derivative of the profit equation with respect to *I* gives $B_t = m$. This simple equation assumes that costs are a linear function of *I*. As noted earlier, this is not likely to be true over the full range of *I*, but it may be approximately true within a considerable range of inventories (review Figure 12-2).[1]

In practice, convergence will not be exact. Typically, the cash price faced by individual firms will not be identical to the cash price pertaining to the par commodity defined by the futures contract. Moreover, because of the costs of arbitrage and because of the delivery options specified in futures contracts, convergence will not be exact even for the par grade and location.[2] As a consequence, hedges do not give perfect results. There is basis risk; namely, the actual magnitude of convergence will differ from the expected level of convergence. An analyst can forecast expected convergence (see Chapter 16), but there will be forecasting error. In addition, convergence problems occasionally occur (discussed later in the chapter); failure of convergence is a serious problem for hedgers.

In Equation (1), for example, basis convergence is assumed to be exact; consequently, if the initial basis (marginal revenue) equals the marginal cost of storage, profit is maximized by setting the size of the futures position exactly equal to the size of the inventory (with opposite signs). However, if there is basis risk, then the optimal decision is not necessarily to have a short

1. The equation over-simplifies the returns in a number of ways. Perhaps the most important is that the equation ignores the costs of hedging and the cash flow (marking to market) associated with futures positions. The equation assumes that entering and exiting the futures positions are costless and that only the beginning and ending prices are important. A more realistic specification would take account of the brokerage fees, initial margin, and cash-flow effects.

2. The delivery options for contracts in U.S. markets typically allow delivery over a period of days in the maturity month. Also, the seller may have choices about the grade delivered and the location of delivery. For these reasons alone, convergence will not be exact on the first day of the maturity month. Moreover, there are costs of making and taking delivery of a physical commodity. These costs create a zone in which futures and cash prices can move independently.

position in futures that is exactly equal in size to the long position in the spot market.

This observation has resulted in a huge literature on optimal hedges, which derives from portfolio theory. Using Equation (1) as an example, we can treat the basis risk as being related to the uncertainty associated with the changes in the cash and futures prices. These are the two random variables in the equation, assuming costs are known and the quantity stored, *I*, is given (exogenous). Then, assuming further that risk can be measured by the variance of returns, this variance is

$$(2) \qquad \mathrm{Var}(R_{t+k}) = I^2\mathrm{Var}(p_{t+k}) + X^2\mathrm{Var}(f_{t+k}) + 2IX\mathrm{Cov}(p_{t+\mathrm{k}}, f_{t+k}),$$

where *R* is the returns, the lowercase letters define the price changes over the interval $t$ to $t + k$, and *X* and *I* are as defined before. Given the foregoing assumptions, Equation (2) can be used to find the level of *X* at time *t* that will minimize the variance for a given level of *I*. The mathematics involves taking the first derivative of the expression relative to *X*, setting it to zero, and solving for *X* (see the appendix to this chapter). The result is that the variance-minimizing hedge depends on the covariance between the price changes relative to the variance of the futures price change.

The foregoing discussion introduces the ideas of maximizing expected profits or minimizing the variance of expected profits from holding and hedging inventories. The literature on optimal hedges is, however, large because many possible definitions of the objective function (i.e., defining what is being optimized) exist, both because of varying assumptions about which variables are assumed to be risky and how risk is to be measured and because of issues about how to estimate the unknown parameters in the objective function. We illustrate some of these points in the appendix to this chapter, but this literature is mostly beyond the scope of this book.

## Operational Hedges

The previous subsection considered the role of hedges in carrying inventories. Futures markets can also assist firms in assuring a return from their merchandising, processing, and exporting activities. For example, an exporter may bid on a contract to export wheat without owning a physical stock of wheat. If the bid is successful, the firm will ultimately have to purchase wheat for export, but the cost of wheat can be more or less assured at the time the forward contract is concluded by buying wheat futures. The bid is constructed by using the price of a relevant wheat futures contract (near the time when the wheat would be purchased) adjusted by a merchandising margin. If the bid is successful, the firm has a short position through the forward contract to export wheat and a long position in wheat futures. When the wheat is purchased, the futures position is offset, and the firm realizes the return built into the bid.

A merchant can offer farmers forward contracts to buy their grain while assuring a positive return through hedging. The price set in the forward contract is linked to a relevant futures price. The precise nature of the price risk faced by the merchant in offering the contract will depend on its terms. A straightforward example is a merchant offering a forward contract to a corn producer at planting time, say in April. The contract price offer can be based on the current price quote of the immediate post-harvest futures contract (the December contract for corn); the price fixed in the forward contract will equal the futures price adjusted for the local basis. When the forward contract is consummated, the corn buyer will sell an equal quantity of December corn futures as a hedge. These two transactions presumably are constructed to provide a merchandising margin for the buyer. When the corn is delivered at harvest, the merchant pays the contracted price and offsets the futures position. The futures market is being used routinely to lock in a margin for the firm and also to protect against the risk of a price decline.

The availability of futures markets for soybeans, soybean meal, and soybean oil allow processors to lock in a processing margin. A 60-pound bushel of beans is processed into about 48 pounds of meal, about 11 pounds of oil, and approximately 1 pound of "waste." If the futures price of soybeans is $14.25 per bushel, of soybean meal $420 per ton, and of soybean oil $0.50 per pound, then buying bean futures and selling oil and meal futures at these prices assures a gross margin of $1.33 per bushel of beans.[3] The existence of futures markets gives soybean processors considerable flexibility in pricing their input and output. The processor may fix the purchase of price of beans in the local market via a forward contract, as discussed in a previous example, and/or make forward contracts for the oil and meal output (as well as using futures). The general point is that the existence of these markets provides processors with a variety of opportunities for establishing positive margins for their businesses. (Of course, margins vary through time and can be negative.)

Operational hedges are an integral part of the business, and many types of agribusinesses routinely use futures markets. The existence of futures markets allows firms to construct bids, offer forward contracts, and establish prices. The firms' returns come from the operating margin established by the hedges. These varied potential uses of futures markets further illustrate Working's definition of a hedge in futures as a temporary substitute for a later transaction(s) in the cash market.

3. The estimated margin is computed as follows. A price of meal of $420 per ton equals $0.21 per pound; multiplying by 48 gives a value of meal of $10.08 per bushel. The oil price is $0.50; multiplying by 11 gives $5.50 per bushel. The total value of the products is $15.58, and the cost of beans is $14.25. Subtracting, the gross margin is $1.33 (15.58 − 14.25) per bushel. This margin must cover all costs of processing, and since basis risk exists, the exact margin is not guaranteed via a hedge.

## Anticipatory Hedges

The existence of a futures market allows a farmer to price a commodity before it is produced. Similarly, a processor can price an input before the actual purchase of the physical commodity. An anticipatory hedge is based on expected future actions, such as the production and sale of the commodity. Anticipatory hedges are typically done selectively. The potential hedger uses decision rules about whether or not to hedge and the magnitude of the hedge. (In contrast, the operational hedge is typically a routine part of the business.)

Indeed, advisory services (consultants) are available, for a fee, to assist farmers with marketing decisions. Their advice often includes recommendations to sell a portion of the crop before it is harvested by selling futures and/or using put options. (The farmer could also forward contract.)

An anticipatory hedge can lock in a profit when price relationships are favorable. A corn producer, for example, will know the approximate cost of producing corn. Thus, if before or during the growing season the price of December corn provides an acceptable return above costs, a hedge using futures contracts assures this return. The farmer need not take the price risk associated with waiting until the crop is harvested and available for sale before pricing it. Of course, if prices rise after the hedge is placed, the producer will face margin calls and may regret the initial decision.

Contracts for some grain futures in the United States are traded more than 36 months in advance. Thus, a farmer could, in principle, price not only the current crop but up to three subsequent crops via futures. In March 2013, contracts for corn on the Chicago Board of Trade were being traded for delivery as far out as December 2016. Clearly, a corn producer has considerable flexibility in establishing prices, not only for the current crop but for future crops. It should be noted, however, that prices are not always at a level that fully covers costs.

Although anticipatory hedges provide flexibility in pricing a commodity, they also have risks. To see this, we can write a profit function as

$$\text{(3)} \qquad \text{Profit} = P_{t+k}A_{t+k} + X_t(F_t - F_{t+k}) - \text{production and hedging costs,}$$

where $A$ is the output, $P$ is the cash price, $X$ is the futures position, $F$ is the futures price, and the subscripts denote time periods.[4] The time $t + k$ can be defined as an immediate post-harvest date. If this date is at contract maturity and if prices pertain to the par commodity in a perfect market, then $P_{t+k} = F_{t+k}$, but, of course, in practice the two prices will not be identical. Thus, the first point to notice is that the net price from a hedge will depend on the basis at time $t + k$. A decision about placing a hedge at time $t$ requires an estimate of the basis at the completion of the hedge. This information must be relevant to the particular individual farm firm contemplating the hedge.

4. Again, the model is simplified by ignoring transactions costs and the marking-to-market feature of a hedge in futures.

For example, if the harvest-time basis for corn is 40 cents in a rural (non-par) location, then selling December futures at 570 cents per bushel is equivalent to a local price of 530. As explained in Chapter 12, the basis is not a constant. It will vary from year to year. If, in our example, the hedger expected a 40-cent basis but it turns out to be 45 cents, the actual return is 525 rather than 530. Deviations from forecast or "normal" bases will result in unplanned profits or losses. (We encourage readers to work out empirical examples to understand this point.) Successful hedging depends on the quality of forecasts of the relevant basis at time $t + k$. This risk can be measured by the variability of actual bases around the prior forecasts (expected levels) of the bases; one such measure is the standard deviation of forecast error, to be discussed in Chapter 15.

Basis risk can be small relative to price-level risk, but it need not be for every potential user of futures markets, and there have been incidents of futures and cash prices not converging as expected, resulting in poor performance of many hedge positions. The lack of convergence is likely to be a consequence of contract-design problems related to the delivery process, not to "excessive speculation" (Irwin et al. 2011; Paul, Kahl, and Tomek 1981).

A second risk relates to yield and hence to the quantity $A_{t+k}$. In the inventory hedging example (the previous sub-section), the size of inventory is known and the risk of loss in storage is usually small. Thus, the quantity $I_t$ is treated as known. In contrast, output may be influenced by pests, diseases, and the weather. The quantity, to be realized at time $t + k$, is not exactly known at time $t$ and is a random variable at time $t$. This has implications for hedging using futures contracts. If a farmer sold 100,000 bushels of corn in futures but produced only 80,000, then 20,000 bushels of the futures position is effectively a speculative position. Moreover, if a lot of other farmers have suffered the same damage and aggregate production falls as a consequence, then prices will rise, and the farmer-hedger will take a loss on the position in futures. Under these circumstances, the return from the increase in cash price does not completely offset the loss from the increase in the futures price because of the 20,000-bushel difference in the two positions. Thus, the futures position should be less than the expected production. In principle, both yield and price risk can be considered in models that estimate the optimal size of the futures position. Because of yield risk, farmers might also consider the use of put options for hedging (briefly discussed in the next sub-section). Also, in the United States, the federal government has subsidized crop and revenue insurance, which provide alternatives to using futures and options markets in managing risk. Another consideration is possible government price-support programs. If market prices are at or near the support level, then the down-side price risk is small and little incentive exists to hedge. On the other hand, a processor realizes that prices cannot go much lower and may want to assure a low procurement price via buying futures or by purchasing call options.

One of the controversial features of making decisions about hedges, which is illustrated by Equations (1) and (3), is whether changes in futures prices can be forecast. As can be seen from these equation, the effect of the hedge on returns depends, in part, on the change in the futures price over the hedge interval. In theory, the expected futures price change is zero: $E[F_{t+k} - F_t|I_t] = 0$, where $I_t$ in this equation represents the information set at time $t$ (Chapter 12). If this is true, then returns cannot be increased, on average, by taking positions in futures.

On other hand, if at time $t$ models exist that provide information for profitable speculation (forecasts of the price change), then this information should be taken into account in the determining the quantity to be hedged. From a farmer's perspective, if prices are forecast to decline from time $t$ to $t + k$, then the proportion of the crop to be sold at $t$ increases; if prices are forecast to increase, then, *ceteris paribus*, the position in futures will be smaller than it otherwise would have been. Wisner, Blue, and Baldwin (1998) suggest that pre-harvest marketing strategies exist that will enhance net returns for corn and soybean farmers in the United States. Basically, they argue that the pre-harvest prices of corn and soybeans are biased upward. Their proposed strategies assume that prices, on average, decline from early in a contract's life until maturity just after harvest.

Finally, we note that the existence of futures markets permits diversification of the sale of a grain crop through time. For example, 10 percent of a farmer's crop could be sold at 10 different points in time (hence, at 10 different prices), and some of these sales can be made prior to harvest by using futures contracts. Some evidence for the U.S. corn market suggests that such strategies are not especially effective in reducing price risk (Peterson and Tomek 2007). The reason is that the variance of prices is not a constant over the marketing year (Chapter 9). Thus, the potential benefits of diversification—low prices in some periods counter-balanced by higher prices in other periods—are offset by moving sales from months with low price variability to months with large price variability.

## Using Options Contracts

The existence of options-on-futures contracts for commodities (and the development of forward contracts with options features) provides farmers and other business people with more alternatives for hedging. Indeed, given the existence of puts and calls, each with a range of strike prices, a firm can create a large number of possible portfolios, i.e., combinations of positions in the spot market and in various futures and options contracts. The alternative portfolios will give different patterns of returns (Hauser and Eales 1987). Here, we briefly introduce the possibilities. Readers interested in this topic should seek an in-depth treatment elsewhere.

Perhaps the simplest example of using options to hedge is a farmer's purchase of a put to assure a floor price for the crop. This is a potential alternative to using futures contracts to hedge a growing crop. The returns portion of this hedge can be defined as

$$\begin{aligned}\text{Return} &= A_{t+k}P_{t+k} - X_t M_t + X_t(F_t - F_{t+k}), \quad \text{if } F_t > F_{t+k}, \text{ or} \\ &= A_{t+k}P_{t+k} - X_t M_t + 0, \quad \text{if } F_t < F_{t+k},\end{aligned}$$

where $A$ is the production, $P$ is the spot price, $X$ is the size of put option position (which is equivalent to a futures position), $M$ is the options premium, $F$ is the futures price, subscripts $t$ and $t + k$ represent two time periods, and the strike price for the put is assumed to equal the current ($F_t$) futures price, i.e., an at-the-money put. By buying a put, the producer has effectively established a floor price; the floor will be $F_t$ minus the options premium and the basis at time $t + k$. If the price level declines, the option will be exercised; having purchased the right to sell at $F_t$, the farmer will be able to buy the position back at the lower $F_{t+k}$. (The equivalent transaction is to sell the put at $M_{t+k}$ because, in a perfect market, $M$ will increase by the amount of the decrease in $F$.)

If the price level remains the same or increases, the option will not be exercised. The return will simply equal the cash sale minus the cost of the option. An increase in $F$ means, of course, that $P$ will increase. The use of the put permits the farmer to realize the benefit of any increase in price; in contrast, a hedge in futures fixes the price. If yields are unexpectedly small and prices rise, the farmer will obtain the benefit of the higher price for whatever quantity may have been produced. Since the put's total cost depends on the size of the position, the transaction is still influenced by yield risk. The larger the quantity hedged ($X_t$), the larger the cost, and if the quantity "insured," $X_t$, is large relative to the realized crop size at $t + k$, then the yield risk reduces returns. Options-on-futures have the same basis risk as the underlying futures contract.

The premiums for options are market determined, and the premiums reflect the potential risk to the sellers that the option may be exercised (Chapter 12). If a potential hedger is concerned about the cost of at-the-money options, an out-of-the-money options can be used, which provide less protection but also have lower premiums (review Table 12-3).

To illustrate these points, we assume that a soybean processor is concerned about an increase in the price of soybeans. The current level is, say, 1,450 cents per bushel. The premium for an at-the-money (1,450 strike price) call, which will expire in 2 months, is 32 cents. The cost per 5,000-bushel contract on the Chicago Board of Trade is 32 cents multiplied by 5,000 bushels, i.e., \$1,600 per contract. In contrast, the premium for an out-of-the-money call (1,480 strike price) is 20 cents per bushel, or \$1,000 per contract. If the firm purchased the call with a strike price of 1,480, the cost is reduced but the hedge protects only against increases in prices above 1,480 cents.

Another way to lower the cost of hedging with options is through more complex spread transactions. The firm could, for example, execute a *vertical* *bull call spread* by buying the at-the-money call and selling the out-of-the money call. The net cost is 12 cents (32 – 20) per bushel. For this hypothetical example, the maximum protection would occur if prices rose to 480, giving a return from the options transactions of 1,480 – 1,450 – 12 = 18 cents per bushel. If soybean prices increased to 1,500 cents per bushel, the return to the purchase of the at-the-money call would be 50 cents (1,500 – 1,450) per bushel, but having sold a call at a strike of 1,480, the firm would lose 20 cents (1,500 – 1,480) on this position. For prices above 1,480, the gains in one position would offset by losses in the other. Thus, the spread provides a limited amount of protection to the soybean processor but at a reduced cost.

The type of protection in the two examples differs. The purchase of the out-of-the-money call does not protect against price increases until prices rise above the strike price. The spread position provides some immediate protection against prices that rise above the at-the-money level but not against large increases above the out-of-the-money level.

The general point to be gained from these illustrations is that the diversity of inter-related futures and options contracts, with many strike prices, makes many combinations of positions possible. Firms can structure a variety of portfolios, including those that involve speculative positions, that meet the firm's objectives. Combinations of positions in options contracts have terms like *butterflies* and *straddles.* Some strategies recommended by consultants and brokers are speculative and involve large risks. Opportunities for speculative profits and losses will depend on possible changes in price levels or in relative prices of the derivative instruments (Hauser and Eales 1987).

Clearly potential users of these markets must understand the relationships among prices of the spot commodity and of the various derivatives and the related benefits and risks. One of the barriers to the use of options contracts and structuring portfolios by small firms is the perceived complexity of the transactions and their potential outcomes. There is a cost involved in learning about the use of derivatives and of following the various market prices. The cost is not just the premiums on the options but also the opportunity cost of the time required to manage a portfolio (or paying someone to manage it). If, as previously suggested, some strategies are speculative, it is not surprising that many farmers are skeptical about the potential benefits of using options markets. A second barrier for some firms is that the premiums are "too large." For soybeans, the premium for an at-the-money option maturing in 6 months can be over 6 percent of the value of beans. Nonetheless, one must recognize that the premiums are market-determined prices for the options; obviously, there are willing buyers and sellers at these prices.

## Price Discovery

Price discovery is a by-product of futures trading, but nonetheless it is an important function of futures markets. The views of many buyers and sellers are focused on a single market, and as explained in Chapter 12, price is being established for a well-known item, which is precisely defined by the contract. Futures prices can be used as reference prices for establishing spot prices and represent a market consensus on the expected price in the future based on current information.

The introduction of a new futures market attracts speculators who do not trade in the cash market. Thus, prices are being discovered by numerous traders with a variety of motives. An uninformed view of speculation is that it has no economic benefit. Indeed, speculation is sometimes seen as having detrimental effects on price discovery because speculation is thought to contribute to unwarranted price instability. In this section, we elaborate further on the role of speculators in futures markets (and by implication in options markets) and comment on the influence of futures trading on cash prices.

## Functions of Speculators and Returns to Speculation

Speculators contribute liquidity and information to the price discovery process. Speculators provide two kinds of liquidity. One is intra-day liquidity to accommodate the uneven order flow into the market. Buy and sell orders of equal magnitude do not arrive in markets simultaneously, and as discussed in Chapter 12, scalpers provide the service of bridging between orders. Their potential profits are a return for making a market.

A liquid market is one in which the price effect of a transaction is small, say, the minimum price change allowed by the market. If, instead, a transaction creates a large price change, this is costly to a hedger. For example, if a merchant wished to sell 100,000 bushels of corn as a hedge and if the sale reduced the price 1 or 2 cents per bushel (rather than 0.25 cent), then clearly the cost of the hedge would be increased. Likewise, offsetting the short position would cause an equivalent price increase, and the round-trip transaction could cost 2–4 cents in adverse price changes. Hedging in futures is facilitated by its potential low cost, and having low-cost hedges depends in important ways on the liquidity provided by scalpers.

Assuming no new information is entering the market, intra-day transactions prices tend to have negative autocorrelation—they bounce up and down around the mean. Scalpers earn their income from this variation. They are willing to buy a sell order at minimum price decline below the previous price, expecting to sell the contracts perhaps a few at a time at a higher price.

Another potential need for liquidity relates to the seasonality in the production and use of commodities. For grains, the demand for short hedge positions in futures is seasonally large immediately after harvest when inventories are largest. Firms owning inventories and hedging these positions will use short futures. As inventories decline, the short positions decline. To the extent that short hedges are not offset by long hedges, speculators willing to take long positions are needed. In other words, speculators provide a necessary offset to any imbalance between the demand for short and long hedge positions. If hedgers are net short, then speculators are net long.

This type of speculation potentially involves taking positions in futures and holding them. A trader that is long futures is speculating that prices will rise. This raises the question, as noted in Chapter 12, of whether changes in futures prices can be forecast. Also, although economists believe in efficient markets (defined later in the chapter), some have hypothesized that prices in efficient markets could have a systematic bias related to the existence of risk premiums. This bias, it is argued, is essential to provide speculators with an incentive to take risky positions in futures contracts.

Since futures markets first arose for seasonally produced commodities and since hedging inventories is thought to have an important role of these markets, it is natural to conclude that the risk premium is positive and that futures prices are biased downward. That is, at harvest time, when hedgers are demanding short positions in futures, speculators are assumed to require a "payment" to take the other side of the contract. This payment could come from a downward bias in prices. On average, this bias would permit speculators to buy low and sell high, as illustrated schematically in Figure 13-1.

In Chapter 12, we assumed that $F_t = E[F_T|I_t]$, but if a risk premium exists, then presumably it is a market-determined rate that changes this equality. This can be viewed in several ways. One is that $E[F_T|I_t] = F_t + r_t$, or $F_t = E[F_T|I_t] - r_t$. The current futures price is the expected price at maturity minus a risk premium ($r$), where $r$ is in the same units as $F$. Usually, the risk premium is defined in percentage terms as a proportion of $F_t$, and since the premium is uncertain (will

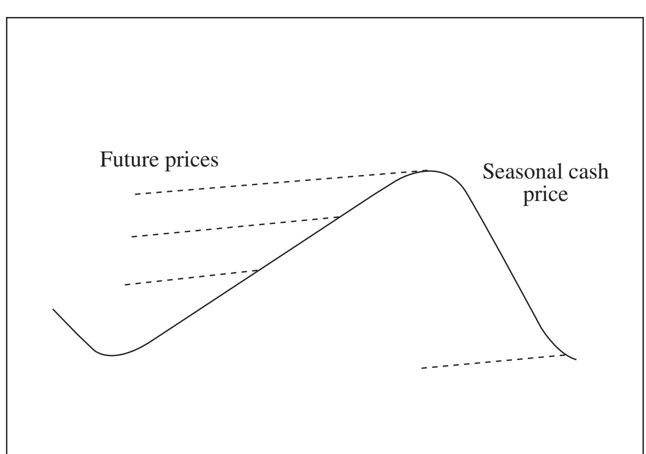



**Figure 13-1.** Hypothetical biased relationship of futures prices to cash prices under the assumption of a risk premium

not be earned until the speculative position is completed), it can be viewed as an expected value. If so, the model can be written as $F_t = E[F_T|I_t] - E[r_T|I_t]F_t$.[5] In words, the current futures price is the expected price at maturity minus the risk premium expected at the time the speculative position is taken.

Keynes (1923) and Hardy (1923) apparently were the first to discuss risk premiums. Keynes believed that speculators earned a risk premium; Hardy (1923: 225) did not. Subsequently, much empirical and theoretical research has been conducted on the topic, but the controversy persists. Keynes imagined that a particular risk was being shifted from hedgers to speculators and did not consider other components of hedgers' and speculators' portfolios. In contrast, the Capital Asset Price Model specifies that the rate of return for asset *j* depends on a risk premium that is proportional to the contribution of the individual asset to the risk of the return from the total portfolio (wealth) of the trader (Leuthold, Junkus, and Cordier 1989: 111–113). If adding a futures position to a portfolio lowered the overall risk, then a risk-averse trader would actually be willing to pay for this benefit. Conversely, if the futures position added to risk, then the trader would have to be compensated for taking it. It is also possible that the futures position has no effect on the risk of the total portfolio, in which case the risk premium is zero.

Thus, although economists believe that price risk is important in an individual's decision making (assuming most decision makers are risk averse), this does not necessarily imply that positive risk premiums are required for individuals to take particular speculative positions in futures contracts. The question of whether a *market* risk premium exists is further complicated by the issue of aggregation over individual traders, who may have positive, negative, or zero premiums.

In addition, the types of hedgers are more diverse than they were in the 1920s. The demand for long hedging has grown since Keynes wrote. Grain exporters, processors, and other merchants are as concerned about the risk of price increases as inventory-holders are about the risk of price decreases. Consequently, the need for long speculative positions to balance short hedge positions is probably smaller now than when Keynes was writing. When this consideration is added to the possibly different risk premiums of individuals, it need not be the case that futures prices have a measurable risk premium. As a generalization, treating the current futures price as equal to the expected price at maturity appears to be a reasonable assumption for commodity markets.

In this context, a positive return to speculative positions in futures contracts depends on the speculator having superior forecasting skills. As suggested in Chapter 12, the majority of speculators appear to be losers, or at least the return from speculation in futures is less than the return from alternative investments.

5. The equation multiplies the risk premium by the current futures price, but since the speculator needs to deposit only a margin and since the position will be marked-to-the-market daily, it is perhaps not appropriate to use the futures price as the value of the "investment." Our emphasis here is on the concept and not on its precise measurement.

This then leaves a question, what is (are) the economic incentive(s) for firms or individuals to speculate in futures contracts? At least three hypotheses exist. First, a few traders are likely to have superior forecasting ability. These are traders that have private information and are able to use it to generate profitable trades. Second, as previously suggested, futures positions may contribute to some individuals' portfolios by reducing the overall risk. The returns from the futures positions may have a negative covariance with returns from other assets in the portfolio. Third, some individuals may view speculation in futures as comparable to the purchase of a lottery ticket, which has a small probability of a large gain (Telser 1967). In this case, the motive is equivalent to that for gambling. Such traders are said to be risk preferring.

As noted previously, Wisner, Blue, and Baldwin (1998) argue that corn and soybean prices for the post-harvest contracts are biased upward prior to planting and early in the growing season. They associate this bias with a yield risk premium. An unexpectedly small crop would raise prices, and producers, who are potential short hedgers of the anticipated crop, are concerned that prices might rise because of a small crop. The upward bias would benefit short hedgers because, on average, the futures position would profit from a price decline. Of course, a lower futures price implies a lower cash price, and it is unclear that the possible bias is explainable by a yield risk premium. In any case, it is difficult to measure risk premiums empirically, even if they can be justified theoretically. The typical historical sample is short by statistical standards, and empirical analyses will be influenced by the sample's length. The addition or subtraction of one or more years to the sample can potentially have a large effect on the empirical results.

The other major category of speculation relates to spreading, i.e., taking positions in two or more contracts in anticipation of changes in their relative prices (Chapter 12). Successful spreading depends on economic analysis to identify the potentially profitable arbitrage opportunities. Professionals who understand the economic fundamentals of relative price changes are likely to profit from this type of speculation. But arbitrage opportunities are usually short-lived, and thus amateurs are likely to find it difficult to do profitable spreads. It is arbitraging by speculators, however, that keeps the prices of the numerous contracts and the cash market in economic alignment. Returns to spreading are returns for providing economic information to the market.

In sum, speculation and hedging transactions in combination keep prices consistent with economic conditions. Concerns exist, however, about the potential detrimental effects of speculation, and this topic is addressed in the next sub-section.

## Pricing Efficiency

Markets are not perfect, but they can be efficient in the sense that prices correctly reflect current information. From a practical point of view, this means that profitable arbitrage opportunities do not persist in light of the available information. Much information is available to all traders: current and past prices and data provided by government agencies. A few traders probably have private information. Multinational commodity firms like Cargill probably have access to new, fundamental information sooner than the typical trader. This information also may be more accurate, and the firms' research departments estimate the effects of changes in fundamental factors on prices. Those who invest in collecting high-quality data and analyzing it presumably profit from their investments.

Using the terminology of Fama (1970), futures markets are semi-strong-form efficient if current prices reflect all publically available information. They are strong-form efficient if current prices reflect all public and private information. Commodity futures markets may be semi-strong-form efficient but are probably not strong-form efficient. In this context, price adjustments to new information will not be instantaneous, but they are likely to occur quickly. If one or a few firms obtain important new information, their transactions will result in changes in prices, and the price changes become a part of the publically available information. Rather soon, all interested traders will become privy to the new information. If a market is semi-strong-form efficient, arbitrage opportunities will be short-lived.

In appraising market efficiency, analysts should keep two points in mind. First, current information may turn out to be erroneous; it can be superseded by better information. If information turns out to be wrong, then prices were wrong; they will adjust to the revised data. But this does not mean that the market was inefficient. Efficiency merely means that prices correctly reflect what is known at the current time even if that information is flawed. Second, arbitrage is costly. These costs include the collection of data and its analysis as well as transactions costs. Also, since forecasts are not perfect, there is risk associated with arbitrage. A firm may have a model that has statistically significant relationships, but this does not guarantee that a forecast price change will provide economically significant benefits. The forecast change may be small relative to the costs of acting on the forecast.

Markets may not always be efficient, and this raises the question, why can markets be inefficient? That is, why can seemingly profitable arbitrage opportunities persist? We identify four possibilities. First, price formation may be dominated by uninformed or misinformed traders. These traders collectively have the economic power to determine current prices. The inefficiency can then be of two types: under- or over-adjustment to existing information. Under-adjustment, sometimes called price continuity, means that prices take long periods to adjust to new information. Rather than the prices' adjustment to news in a manner of minutes or hours, the adjustment takes days. For whatever reason, information is transferred slowly through the market. In this case, current and past prices can be used to forecast tomorrow's price.

Over-adjustments, also called bandwagon effects, fads, speculative bubbles, or price reactions, occur when prices over-adjust to new information and this over-adjustment persists for considerable periods of time, perhaps weeks or even months (for an explanation of bubbles in terms of fundamental economics, see Garber 1990). Ex post, the distortions of prices from fundamental values are clearly observable, but while the bubble is occurring, the knowledgeable arbitragers do not have the financial resources to offset the effects of misinformed traders. Ultimately, an increasing proportion of traders become better informed, the bubble "bursts," and prices adjust to warranted levels.[6] In sum, the first possible reason for inefficiency is that prices reflect uninformed or misinformed trading. If a large proportion of traders are in this category, prices can be distorted from equilibrium levels for considerable periods of time.

The second possible reason for inefficiency is market concentration. There has been a dramatic increase in the volume of speculative trading arising from managed commodity pools, exchange trade funds (ETFs), and other commodity-based instruments. Investors can place funds with professionally managed commodity pools (which are somewhat analogous to actively managed mutual funds for stocks and bonds), or they can purchase ETFs (which are "passive" instruments reflecting the value of the components of the fund).

For commodity pools, a concern is that the trading advisors for the pools use similar data and trading rules; hence, they may reach similar conclusions and take the same actions at about the same time. This does not involve formal collusion, but the effect of a large number of similar decisions taken at about the same time could cause unwarranted short-run fluctuations in prices. Empirical analyses have not, however, supported this hypothesis (Irwin and Yoshimaru 1999).

ETFs are based on indexes of prices of a basket of futures contracts. Many different baskets can be constructed, including those for sets of agricultural commodities. These financial instruments reflect the prices of nearby futures contracts used in the index, with rollover rules as contracts expire. An ETF's market value is based on long-only positions in each of the commodities in the basket of contracts. In this case, the concern (hypothesis) is that the increased volume of trading in long-only financial products affects the underlying futures prices. It is important to remember, however, that the net returns to all futures

6. This phenomenon can be defined by a mean reversion model. It assumes that prices are initially influenced by uninformed traders but that prices ultimately revert to the equilibrium level. Thus, a deviation of the current price from the true level has predictive power. This can be written as $F_{t+1} - F_t = a_0 + a_1(F^R{}_t - F_t) + e_{t+1}$, where $F$ is the futures price, $F^R$ is the true mean level, $e$ is the equation's error term, and the subscripts $t$ and $t + 1$ relate to time periods. Although $F^R$ is not observable, analysts use a proxy variable for it. (The results can be sensitive to the choice of the proxy.) If markets are efficient, then the coefficients $a_0$ and $a_1$ should be zero. Irwin, Zulauf, and Jackson (1996) discuss models and methods; their empirical results suggest that the five commodity markets analyzed are efficient.

positions must be zero; each long position is necessarily offset by a short position. (The returns are net negative if transactions costs are considered.) Thus, the long-run benefit (return) or price-distortion effects of holding long-only positions is theoretically about zero.

Sanders and Irwin, in analyzing such products, write, "Historical portfolio returns are not statistically different from zero and are driven by price episodes such as that of 1972–1974" (2012: 515). They note, as we have, that unlike investing in equities, futures markets produce no net earnings (see also Irwin and Sanders 2011). Moreover, the data suggest that, while long speculative positions (open interest) were growing, so were short hedge positions. The ratios of speculative to hedge positions were within the range of historical experience, as of the years 2006–08 (Sanders, Irwin, and Merrin 2010).

The third possible reason for inefficiency is that most commodity futures markets have rules that limit the size of the daily change in price. In early 2013, the daily change in the price of corn on the Chicago Board of Trade was limited to a 40 cents per bushel increase or decrease from the previous day's close. (Such rules are subject to change, and a sequence of limit-price moves triggers a relaxation of the rule.) The intent of such rules is to provide time for the market to appraise and digest a major change in information and to allow time to assure that margin calls are met. The consequence is that warranted price adjustments can be delayed for a day or two; because trading is suspended, no arbitrage is possible.

The fourth possible reason for inefficiency is that prices may be deliberately manipulated. In the United States, federal regulation of futures markets has minimized price manipulation and other fraudulent activities. Nonetheless, given human nature (greed), attempts to manipulate prices can occur. It is, however, difficult to manipulate prices in active markets. When markets have a large volume of trading and when most traders have good information, then by definition large financial resources are required to move prices. Moreover, if a trader is successful in moving prices in one direction (say, raising prices by a large volume of buying), the price is likely to move back to the original level as the trader attempts to take "profits" by offsetting transactions. And, of course, there are transactions costs associated with the large volume of trading.

Manipulation near the expiration of a contract is relatively easier because of the potential costs of making and taking delivery and the declining volume of trading. The classic attempt to corner a market involves one or more traders taking large long positions in futures contracts for a particular month while simultaneously having control of the physical supplies that could be delivered at low cost to fulfill these contracts. Thus, the manipulator is in a position to force prices up as sellers attempt to cover their positions either by buying contracts or by attempting to buy the commodity to make delivery. It is precisely because the open interest in futures must go to zero at contract maturity (either by offset or by delivery) that the exchanges themselves and the regulatory agency in the United States (the Commodity Futures Trading Commission) are especially vigilant as contracts mature.

## Futures Trading and Cash Prices

This chapter and Chapter 12 have emphasized that futures markets are price discovery mechanisms. Many buyers and sellers are brought together to trade in standardized contracts. Consequently, futures prices are sometimes used as reference prices for establishing prices in spot markets. If so, buyers set cash prices equal to the price of the nearby futures price plus or minus an appropriate differential (basis), and logically changes in futures prices will lead changes in cash prices. Analyses of cash-futures price relations are, however, handicapped by the complexity of price relationships and by the difficulty of collecting high-quality cash price series. As emphasized throughout this book, transactions prices in cash markets for agricultural commodities reflect the diverse attributes of the transactions, and price reporting can be costly and imperfect. Moreover, local cash markets can be imperfect. Thus, results of empirical analyses are often difficult to evaluate (e.g., Fortenbery and Zapata 1993).

Whatever the empirical results, economic logic suggests that cash and futures markets must be related. In this context, a question that has been asked is, does the introduction of a futures market have beneficial or adverse effects on cash prices? The question is most often asked, however, when cash prices are thought to be "too high," "too low," or "too variable." Trading in onion futures in the United States was made illegal by act of Congress in 1958 because of the alleged adverse effects of futures markets on cash prices. (Onion prices are still highly variable; see Chapter 9.) The spike in petroleum prices during the Gulf War of 1991 was alleged to have been caused by speculators on the New York Mercantile Exchange. The increase in commodity prices that started in 2006 has been attributed, at least in part, to speculators (as summarized in Irwin and Sanders 2011). In all these cases, it seems likely that futures prices reflected reasonable responses to current and expected economic conditions; in contrast, the allegations imply that, if futures markets had not existed, the spot prices would have been less volatile, or declined less, or increased less.

The hedging, made possible by the existence of futures markets, can influence prices in a beneficial way. A futures market can lower the cost of shifting risk for many firms. If the cost of managing risk is reduced, this can influence the level of the supply and demand functions. So far as we know, no complete conceptual model exists of these effects, but clearly, if producers' costs of managing price risk are reduced, then, *ceteris paribus*, the supply function shifts to the right (Turnovsky 1983). As noted previously, hedging can help assure returns (manage risk) for storing, processing, and merchandising commodities. A cost-reducing institution is a benefit to the economy. Another benefit of futures markets is that they provide decision makers with a consensus forecast of prices in the future, which is useful for making production and marketing decisions.

With respect to price discovery, futures trading attracts more traders to the market. In principle, the amount of information available in the market is increased, and the analysis of the available information may be improved. If this is true, the systematic component of price behavior is more precisely determined when a futures market is introduced. In addition, information may be transmitted faster, so cash prices adjust faster to new information.

Against these potential benefits, one must place two possible adverse effects (which were discussed earlier): deliberate manipulation and the effects of ill-informed traders. Nevertheless, many empirical studies have been conducted using data from the pre-electronic trading era, and most of these analyses do not confirm the allegations that futures trading leads to more unstable prices (e.g., Streeter and Tomek 1992; Irwin, Zulauf, and Jackson 1996; Irwin and Yoshimaru 1999).

A relatively new issue is the potential effects on price behavior of the high-frequency trading that is now possible using existing computer algorithms. Concerns exist that high-frequency trading may cause extra noise (more intraday random variability) in price behavior and/or unintended cascades in prices. As of this writing, empirical analyses of the effects of high-frequency trading in commodity futures markets is limited; moreover, it is difficult to evaluate whether the effects of such trading are transmitted to the behavior of the underlying cash prices (for a review of literature, see Fabozzi, Focardi, and Jonas 2011).

It is true that spot and futures prices have become increasingly variable since 2005 (and that they were already variable). However, the role of futures markets, if any, in this variability is difficult to appraise. Increased price variability implies, *ceteris paribus*, an increased demand for hedging (but basis risk has also increased), and increased hedging and price variability attract speculation. So, speculative activity can be viewed as a consequence, rather than a cause, of price variability, and empirical analyses of the direction of causation can be difficult to interpret.

No doubt, incidents of aberrant futures price behavior have occurred, but since they occur randomly and have only short-run effects, they are difficult to evaluate empirically. Given the general benefits and the seemingly small problems, futures and options markets are thought to be beneficial on balance. Economic institutions develop and persist to the degree that they provide net economic benefits. Markets fail when their benefits are smaller than their costs. If a futures market has persistent imperfections and does not provide for low-cost pricing and hedging activities, businesses will turn to better alternatives (for an example of a market failure, see Paul, Kahl, and Tomek 1981).

## Summary

Futures and options markets can assist businesses in pricing commodities, assuring margins from their business operations, and in managing risk. These markets necessarily attract various kinds of speculation. In principle, the existence of many traders, some with specialized knowledge, should result in efficient markets, and it appears that many futures markets do indeed work well. Like any human institution, these markets are not perfect, but they typically appear to provide net economic benefits.

## Appendix: Optimal Hedges

A potential hedger can view a futures position as part of a portfolio of assets. These assets may yield returns, but the returns are likely to be risky. In this context, the notion of an optimal portfolio relates to maximizing the investor's utility, which is a function of the returns and their risks. Adding an asset to a portfolio can influence the overall rate of return and the riskiness of the return from the portfolio. The effect of a adding or subtracting an asset depends on the correlations (covariances) among the returns.

In this appendix, we consider simple portfolios—positions only in futures and cash markets—and simple optimization problems. There is a huge literature on portfolio concepts and optimal hedges, and this appendix provides a short introduction. We illustrate that, if cash and futures prices are correlated, as theory suggests, then taking a position in a futures contract can help offset the risk of changes in the cash price. Specifically, if the two prices have a positive covariance, then a short position in futures can offset the price effects of a long position in the cash commodity.

For our purposes, an optimal hedge is the futures position (quantity) that maximizes a trader's expected utility, given the trader's position in the cash market. Expected utility is characterized in terms of expected returns and the riskiness of the returns, and risk is often defined by the variance. That is, a mean-variance framework is used. (The representation of expected utility by a mean-variance framework, although common, is controversial because upward as well as downward deviations in profits are considered equally negative for the decision maker.)

A simple example is the case of the firm holding an inventory of grain, where the potential returns from hedged inventory can be written as

$$R_{t+j} = (P_{t+j} - P_t)I_t + (F_t - F_{t+j})X_t - mI_t,$$

where $P$ is the cash price, $F$ is the futures price, $I$ is the quantity of inventory, $X$ is the size of the futures position, $m$ is the cost of storage per unit of $I$, and $t$ and $t + j$ subscripts indicate the beginning and ending periods for storage. In this equation, $I$ is known and given, and the storage interval is fixed. (The fact that $I$ is known implies that the inventory-holder expects the basis convergence over the storage interval to cover costs.) It is assumed that the cost of storage over the fixed interval is known. The model further assumes that $P$ and $F$ are positively correlated, which is consistent with the theory presented in Chapter 12.

An objective function to be optimized can be written as

$$L = E[P_{t+j} - P_t]I_t + E[F_t - F_{t+j}]X_t - \lambda/2[I^2\mathrm{Var}(p) + X^2\mathrm{Var}(f) + 2IX\mathrm{Cov}(pf)],$$

where $E$ is the expectation operator, and the variance of expected returns is specified in the brackets. (These are conditional expectations, but we omit the conditioning notation for simplicity.) The random variables are the price changes; thus Var($p$) equals the variance of the spot price change, etc. The term $\lambda/2$ is a constant related to the individual decision maker's risk preferences. In the equation, the negative sign implies that risk (variance) reduces utility. Thus, if $\lambda > 0$, the decision maker is risk averse; if $\lambda < 0$, the decision maker is risk preferring; and if $\lambda = 0$, the decision maker is risk neutral (the variance drops out of the expression). In this context, the only decision variable is the level of $X$. If $\lambda < 0$, the decision maker desires risk, which is not likely to be the case for most farmers or other firms.

Given the assumptions, the first-order condition for the optimum is found by taking the first derivative of the expression with respect to $X$, setting the expression equal to zero, and solving for the optimum $X$.

$$dL/dX = E[F_t - F_{t+j}] - 2(\lambda/2)\mathrm{Var}(f)\mathrm{X} - 2(\lambda/2)\mathrm{Cov}(pf)I = 0.$$

$$X^* = \mathrm{Cov}(pf)\mathrm{I}/[\mathrm{Var}(f)] - \{E[F_t - F_{t+j}]\}/[\lambda\mathrm{Var}(f)], \text{ or}$$

$$X^*/I = \mathrm{Cov}(pf)/\mathrm{Var}(f) - \{E[F_t - F_{t+j}]\}/[\lambda I\mathrm{Var}(f)],$$

where $X^*$ is the optimal position in futures, conditional on the various assumptions, and can be viewed as a proportion of the given level of inventory, $I$, at time $t$. The ratio depends on the covariance of the price changes relative to the variance of the futures price change and on the ratio of the expected change in the futures price to a term defined by $\lambda$, $I$, and Var($f$).

The last term on the right-hand side is often assumed to be zero. In theory (Chapter 12), $E[F_t - F_{t+j}] = 0$. Also, if the decision maker is highly risk averse, this implies that $\lambda$ will be large. The larger $\lambda$ is, the smaller is the ratio defined in the last term. If the last term is zero, then the optimal hedge ratio is given by the covariance of the price changes relative to the variance of the futures price change, and in this case, the optimal hedge minimizes the variance of returns, given the assumptions of the model.

For hedging inventory, another approach is to recall that a merchant can use the expected basis change as a guide to whether to store and to hedge. If the expected return from the hedge is positive, then the second step is merely to select the futures position that minimizes the variance of returns, given that the expected return will cover costs. Thus, the minimum variance objective can be justified directly.

The size of the variance-minimizing ratio can be estimated from the regression model,

$$P_{t+j} - P_t = \alpha_0 + \alpha_1(F_{t+j} - F_t) + e_{t+j}$$

because the estimate of the slope coefficient $\alpha_1$, using the least squares estimator, is an estimate of the ratio of the covariance of the two variables to the variance of the regressor. Many such models have been fitted in the literature, and they tend to give large hedge ratios. For example, $\alpha_1 = 0.95$ implies that the futures position should be 95 percent of the size of the inventory position.

Although such estimates have been popular in academia, largely because of the simplicity of estimation, they are probably unrealistic. One reason is that objective functions, such as the one given, are too simple. They ignore the costs of hedging and do not include sufficient alternatives in the portfolio (e.g., Brorsen 1995). Also, the setup assumes that the parameters (variances and covariances) are constants, and as indicated in Chapter 12, they probably are not. Since specifying more realistic objective functions and estimation procedures is difficult, it is not surprising that the academic literature on this topic is large.

## References

Brorsen, B. W. 1995. "Optimal Hedge Ratio with Risk-Neutral Producers and Nonlinear Borrowing Costs," *Am. J. Ag. Econ.* 77:174–181.

Fabozzi, F. J., S. M. Focardi, and C. Jonas. 2011. "High-Frequency Trading Methodologies and Market Impact," *Rev. of Fut. Mkts.* 19:7–38.

Fama, E. F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work," *J. Finan.* 25:383–423.

Fortenbery, T. R., and H. O. Zapata. 1993. "An Examination of Cointegration Relations between Futures and Local Grain Markets," *J. Fut. Mkts.* 13:921–932.

Garber, P. M. 1990. "Famous First Bubbles," *J. Econ. Persp.* 4(2):35–54.

Hardy, C. O. 1923. *Risk and Risk-Bearing*. Chicago: Univ. of Chicago Press.

Hauser, R. J., and J. E. Eales. 1987. "Option Hedging Strategies," *North Central J. Ag. Econ.* 9:123–134.

Irwin, S. H., P. Garcia, D. L. Good, and E. L. Kunda. 2011. "Spreads and Non-Convergence in Chicago Board of Trade Corn, Soybean, and Wheat Futures," *Applied Econ. Persp. & Policy* 33:116–142.

Irwin, S. H., and D. R. Sanders. 2011. "Index Funds, Financialization, and Commodity Futures Markets," *Applied Econ. Persp. & Policy.* 33:1–31.

Irwin, S. H. , and S. Yoshimaru. 1999. "Managed Futures, Positive Feedback Trading, and Futures Price Volatility," *J. Fut. Mkts.* 19:759–776.

Irwin, S. H., C. R. Zulauf, and T. E. Jackson. 1996. "Monte Carlo Analysis of Mean Reversion in Commodity Futures Prices," *Am. J. Ag. Econ.* 78:387–399.

Keynes, J. M. 1923. "Some Aspects of Commodity Markets," *Manchester Guardian Commercial*, section 13, March 29.

Leuthold, R. M.., J. C. Junkus, and J. E. Cordier. 1989. *The Theory and Practice of Futures Markets*. Lexington, MA: D.C. Heath & Co.

Paul, A. B., K. H. Kahl, and W. G. Tomek. 1981. *Performance of Futures Markets: The Case of Potatoes*. USDA Tech. Bul. 1636.

Peterson, H. H., and W. G. Tomek. 2007. "Grain Marketing Strategies within and across Lifetimes," *J. Ag. & Res. Econ.* 32:181–200.

Sanders, D. R., and S. H. Irwin. 2012. "A Reappraisal of Investing in Commodity Futures Markets," *Applied Econ. Persp. & Policy* 34:515–530.

Sanders, D. R., S. H. Irwin, and R. P. Merrin. 2010. "The Adequacy of Speculation in Agricultural Futures Markets: Too Much of a Good Thing?" *Applied Econ. Persp. & Policy* 32:77–94.

Streeter, D. H., and W. G. Tomek. 1992. "Variability in Soybean Futures Prices: An Integrated Framework," *J. Fut. Mkts.* 12:705–728.

Telser, L. G. 1967. "The Supply of Speculative Services in Wheat, Corn, and Soybeans," *Food Res. Inst. Studies* 7(Supplement):131–175.

Turnovsky, S. J. 1983. "The Determination of Spot and Futures Prices with Storable Commodities," *Econometrica* 51:1363–1387.

Wisner, R. N., E. N. Blue, and E. D. Baldwin. 1998. "Preharvest Marketing Strategies Increase Net Returns for Corn and Soybean Growers," *Rev. of Ag. Econ.* 20:288–307.

Working, H. 1953. "Hedging Reconsidered," *J. Farm Econ.* 35:544–561.

# PART IV

# INTRODUCTION TO EMPIRICAL PRICE ANALYSIS

All models are wrong, but some are useful.
G. E. P. Box, "Robustness in the Strategy of Scientific Model Building"

This part of the book—Chapters 14, 15, and 16—stresses two broad topics: (1) the specification of models to explain the behavior of agricultural product prices and related variables, and (2) the use and appraisal of the results of quantitative price analysis. Developing useful price analysis models has proved difficult, and we therefore try to get a sense of the depth of scholarship necessary to obtain useful results. We provide a discussion of model building that shows how the economic principles discussed in earlier parts of the book can be combined with statistical methods to provide useful empirical results (Chapter 14). An additional objective is to enable students to interpret and evaluate empirical studies (Chapter 15). This material is reinforced by selected examples presented in Chapter 16. We do not include a formal discussion of statistical inference, and consequently, students with training in statistics will benefit most from these chapters.

CHAPTER 14

# Background for Price Analysis

The term *price analysis* refers to the quantitative study of price behavior, especially demand-supply-price relationships. Much of price analysis is applied econometrics, but quantitative analyses may range from the construction of tables and graphs to the use of a variety of advanced mathematical tools such as mathematical programming and simulation models. No attempt is made here to survey all available quantitative methods. We discuss some of the issues in linking empirical studies to the concepts discussed earlier in the book.

The two reasons most frequently given for engaging in price analyses are (1) to estimate specific coefficients (parameters) such as price and income elasticities and (2) to provide forecasts of prices or the variables affecting prices. In both cases, the results are intended to aid decision makers. For example, research has been devoted to studying the effects of advertising programs for generic (not branded) commodities: does an increase in advertising expenditures by farmers increase the demand for (say) milk sufficiently to cover the costs of advertising? The answer to this question requires an estimate of the elasticity relating the level of demand to advertising expenditures, and the model must control for the other factors influencing the level of the demand for milk. Moreover, if advertising shifts the level of demand, this will affect price and subsequently the quantity supplied. Thus, to determine the net effects of advertising on farm income, a complete model of the dairy sector is needed.

Since, however, it has proven difficult to obtain robust, useful empirical results, the principal contribution of most previous price analyses has perhaps been to generate hypotheses that require additional analyses and to increase understanding of agricultural markets. Even after over 85 years of study (for a classic study, see Warren and Pearson 1928), much remains to be learned about commodity markets. A part of the problem is that, while economic principles do not change, economies are growing increasingly complex. Among other things, this means that models must adjust to changes in the economy.

Moreover, the data available for analysis are not always sufficient to keep up with the increased complexity.

This chapter is devoted to a discussion of procedures for and problems in specifying price analysis models. (An evaluation and use of such models are covered in Chapter 15.) Price analyses often use time-series data, i.e., variables observed with the passage of time. These data, which are generated by the economy, are called *observational (non-experimental)* data, and a fundamental question is whether they are of adequate quality for the purposes of the research. For some problems, data are available from a cross section (sample) of households or firms or from controlled experiments. Although useful for the analysis of many problems, sample or experimental data have less to say about the behavior of prices with the passage of time.

Although this chapter introduces many of the issues faced by price analysts, it is not possible to provide a comprehensive discussion of all the methods and problems of analysis (for a sense of the vast range of the literature, see Bessler et al. 2010). We hope, however, that students gain some appreciation of the complexities of price analysis based on time-series data. In the past, analysts have probably under-estimated the difficulty of obtaining high-quality empirical results.

## Alternative Models and Techniques

The first step in research based on the scientific method is a precise problem definition. The models, data, and methods should follow from the problem statement. We start, however, with an outline of alternative models and associated techniques to give the reader a context for the remainder of the chapter. An implicit assumption throughout this chapter is that it is helpful to have the results of empirical studies based on formal models. Such studies help organize thinking and make relationships among variables explicit. Expert judgment also can be helpful, and empirical studies should be subjected to careful evaluation, including the judgment of knowledgeable individuals. Nevertheless, judgment alone typically is insufficient to solve problems or make precise forecasts.[1]

Given the research objective and a tentative model, an early step in analysis is to collect, organize, and summarize the relevant data. The USDA, for example, provides considerable information about major commodity markets. For the grains, data are provided on inventories, production, and uses. This can be organized into a balance sheet of quantities; the total supply (beginning inventories and production) must equal the sum of the various uses and ending inventories. In terms of a formal model, the quantities constitute an identity; the ending inventory must equal the beginning inventory plus production minus all uses.

1. There is a considerable literature, for example, on composite forecasts. The idea is that combining forecasts from various sources, including qualitative judgments by experts, can improve the precision of forecasts (Granger and Newbold 1986).

**Table 14-1.** Feed grain balance sheet, United States (million metric tons)

| Item | 2010–2011[a] | 2011–2012 (prelim.) | 2012–2013 (est.) |
|---|---|---|---|
| Beginning stocks | 48.1 | 32.3 | 27.8 |
| Production | 330.0 | 323.6 | 285.8 |
| Imports | 2.4 | 2.7 | 5.2 |
| Total supply | 380.5 | 358.5 | 318.8 |
| Food, seed, industrial | 169.9 | 170.1 | 156.0 |
| Feed and residual | 127.7 | 119.6 | 120.7 |
| Exports | 50.7 | 41.0 | 23.2 |
| Total use | 348.3 | 330.7 | 299.9 |
| Ending stocks | 32.2 | 27.8 | 18.9 |
| Average price of corn ($/bu.)[b] | 5.18 | 6.22 | 7.1* |

*Note:* Feed grains included: corn, sorghum, barley, and oats. Asterisk (*) indicates midpoint of forecast range 6.75–7.45.

[a] September 1–August 31 marketing year.

[b] Marketing-year weighted average price received by farmers.

An example for U.S. feed grains is shown in Table 14-1; the data include information on imports and exports as well as domestic supply and uses. This data set is based on information available on March 8, 2013. It represents a revision of information available in February, and it too has been revised in light of new information available in subsequent months. The USDA also reports season-average prices associated with the historical data and provides forecasts of prices associated with the estimated levels of supply and demand (Table 14-1). The forecasts are based on models used by the USDA and presumably modified by judgment.

Analysts in the private sector use official government statistics as a benchmark, but they may adjust the data based on their own information. The private-sector analysts have their own models and judgments. Expert-judgment forecasts may involve reasoning by analogy; the analyst tries to identify past years that look the same as (have economic conditions similar to) the expected conditions for the forecast period. This is sometimes called the balance-sheet approach to forecasting. In other words, the organization of the data has an implicit supply-demand framework, and the technique involves judgment and reasoning by analogy. Often, the analyst uses statistical models as well. We provide a simple illustration of a statistical model in Chapter 16, using data like those shown in Table 14-1.

Empirical models of economic data can be categorized broadly as either deterministic or stochastic. Deterministic models, by definition, assume that the price (or other variable) does not have a random error component. The variability through time is exclusively a deterministic function of past levels of the variable. One type of deterministic model, which has a long history in price analysis, is moving average specifications. For example,

$$P_{t+1} = [P_t + P_{t-1} + P_{t-2}]/3,$$

where $P$ is a price and the subscripts represent time periods. In this case, three prices are used in the average, and each receives equal weight. The forecast for period $t + 1$ is merely the average of the current and two past prices.

The selection of the number of prices to include in the average and whether or not they should receive different weights is a matter of judgment, perhaps based on an empirical search of alternatives. The search would involve examining combinations of lags and weights that seem to provide the "best" forecast. Plots of moving averages, say, of daily prices, are often associated with the technical analysis used by traders for speculating in futures and options markets. But it should be noted that if an analyst searches long enough, he or she is likely to find a specification that fits a set of historical observations. The problem is that this historical fit is not likely to have any predictive ability; the analyst has merely found an artifact of the particular historical record.

Non-linear deterministic models, called chaos models, are a somewhat more recent development in the literature (for a relatively non-technical discussion, see Savit 1988). An example is

$$P_{t+1} = bP_t - bP_t^2 = bP_t(1 - P_t).$$

Price is specified as a non-linear function of the previous price. The precise path of price through time will depend, however, on the value of $b$ as well as the initial starting price, $P_0$. Many possible paths are possible, and some of the deterministic paths look much like the "random" movement of futures prices. Technical analysis methods, such as moving averages, are perhaps crude, but possibly useful, ways of exploring the non-linear qualities of data (Clyde and Osler 1997). Although technical analysis is rather widely used by practitioners, we do not cover these methods in this book. Detailed explanations are available in numerous sources.

Stochastic models are statistical models containing a random error term. In addition, they typically include a systematic component. In simple notation,

$$P_t = S_t + e_t,$$

where $P$ is the price, $S$ represents the systematic component, $e$ is a random error term, and the $t$ subscript represents the time dimension. The subscripts can vary because $P$ may depend only on lagged systematic components. The models and techniques of analysis will depend on the specification of the right-hand side of the equation. Clearly the right-hand side of the equation must be consistent with the observed behavior of the left-hand-side variable, and in practical applications, this can be accomplished via a number of alternative specifications. Competent analysts, using the same data to analyze a problem, may specify different models. As noted in the quotation at the beginning of this chapter, it is not possible to specify the uniquely correct model for an economic sector, but it may be possible to obtain empirical results that are useful relative to a particular research problem.

We can further sub-divide the stochastic category into *structural* models and *associational* models. Structural models specify the supply-demand relationships relevant to the research problem; the systematic component attempts to capture the causal economic relationships. The specification of $S_t$ involves explicit measures of economic variables. A complete model may involve numerous equations, including demand equations for various uses of the commodity, such as domestic demand, export demand, and inventory demand. The intent is to better understand the economic structure and to obtain estimates of parameters, like price elasticities of demand, that are useful for decision making.

Associational models summarize the empirical regularities in the data. In the early literature, this was done by making the variable of interest a function of deterministic variables that capture trend and seasonal components of the variable (as well as a random component). A simple linear trend equation is

$$P_t = \alpha_0 + \alpha_1 t + e_t,$$

where $t = 1, 2, \ldots, T$. This equation assumes that price, $P_t$, is increasing (or decreasing) by $\alpha_1$ amount each time period. $P_t$ is also varying with the random component, $e_t$. The most common set of assumptions about $e_t$ is that it has mean zero, has a constant variance, has zero lag covariances,[2] and is normally distributed. If the linear trend model were correct, then the mean of price would be trending but with random variability around the trend.

Although using deterministic time-series variables like trend is common in empirical models, modern time-series econometrics stresses autoregressive moving average (ARMA) specifications and their extensions (Hamilton 1994). A first-order autoregressive process is defined by

$$P_t = \alpha_0 + \alpha_1 P_{t-1} + e_t.$$

In contrast to the chaos model, this equation is linear; it is a difference equation with an appended error term. The error term is assumed to have the properties just listed (zero mean, constant variance, etc.). If this model is correct, then $\alpha_1$ is expected to lie between −1 and 1. If $\alpha_1 = 1$, then the time series is said to be "integrated of order 1," or to have a "stochastic trend," or to follow a "random walk."[3] For $\alpha_1 = 1$, the autoregressive model can be rewritten as

2. An example of a lag covariance is the covariance between $e_t$ and $e_{t-1,}$ sometimes written $\mathrm{Cov}(e_t e_{t-1})$. Although these errors are usually assumed to have a zero covariance, in alternative model specifications, they are assumed to be correlated, called autocorrelation.

3. In Chapter 9, we argue that it is unlikely that cash prices of commodities have stochastic trends. It is possible, however, that futures prices for a given contract have a stochastic trend (Chapter 12). This is consistent with futures markets being efficient.

$$P_t - P_{t-1} = \alpha_0 + e_t,$$

which is a random walk with drift.

A first-order moving average model is exemplified by

$$P_t = \alpha_0 + e_t + \beta e_{t-1},$$

where $\alpha_0$ and $\beta$ are parameters of the model, $e$ is an unobservable error term, and the subscripts represent time periods. The two specifications can be combined to obtain an ARMA process. The first-order specification (ARMA (1,1)) is

$$P_t = \alpha_0 + \alpha_1 P_{t-1} + e_t + \beta e_{t-1}.$$

More complex models would involve longer lags in $P$ and/or $e$. An obvious generalization is the second-order autoregressive model, which will include $P$ lagged two periods.

The various specifications have varying implications for the time path of the variable, $P_t$. In the simplest first-order autoregressive case, if $\alpha_1$ lies between 0 and 1, then $P_t$ is positively autocorrelated and will follow a dampened wave pattern. If $\alpha_1$ lies between −1 and 0, then $P_t$ is negatively autocorrelated and will alternate up and down while dampening. (The simplest cobweb model produces a dampened cycle with negatively autocorrelated prices; see Chapter 9.) The implications of the various specifications of ARMA models for the time-series properties of the random variables are covered in books such as Hamilton (1994).

In this and the following two chapters, structural modeling is emphasized. We examine linkages between the economic models of earlier chapters and empirical analyses. This does not mean that structural models are more important than associational models. Indeed, as implied in earlier chapters (and footnote 3) there should be a logical relationship between the structural and associational representations of economic variables. The time-series properties of the random variables in a structural model—the prices and quantities—should follow from the economic structure of the market under analysis.[4] A potential criticism of some time-series results in the literature is that the analysts have not considered the implications of their empirical results for the underlying structure. For example, is the finding that a commodity price series is integrated of order 1 (has a unit root) consistent with the logic of predictable systematic behavior in the price series?

In structural modeling, it is common to use linear or log-linear statistical models (although non-linear models are also feasible). For instance, consider a

4. Hamilton (1994) provides a comprehensive, but advanced, treatment of time-series analysis. Shorter introductions are provided in econometrics textbooks (e.g., Johnston and DiNardo 1997).

linear regression model that makes a dependent variable ($Y$) a linear function of a set of explanatory variables ($X$'s):

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \cdots + \beta_K X_{tK} + e_t,$$

where $Y_t$ is the observable dependent variable; $X_{tk}$ are the observable explanatory (independent) variables, $K$ in number; $e_t$ is the unobservable error or disturbance term; the β's are the unknown population parameters; and $t = 1, 2, \ldots, T$ observations on the variables. The equation assumes that $Y_t$ depends linearly on the observed $X_{tk}$ and on the unobserved errors. The statistical problem is to estimate the parameters given the data and perhaps to test hypotheses about the unknown parameters.

For any given research problem, numerous plausible model specifications exist. A particular specification constitutes a researcher's view of how the data have been generated by the economy. The researcher must determine which variables will appear in the equation, whether the relationship is linear,[5] what assumptions may be reasonably made about the nature of the error term and about the distributions of the $X_{tk}$, and so on. Moreover, more than one equation may be required to analyze a particular economic sector. (Some of these details are covered in subsequent sections.)

Then, the coefficients of the model must be estimated, and several alternative estimation methods exist. By far the most common estimator, historically, is the method of least squares, sometimes called ordinary least squares (OLS) to distinguish it from other least-squares-type estimators. This book is not about econometric methods; hence, we do not concern ourselves with the details of alternative estimators. It should be noted, however, that OLS, although simple and convenient, is not always the preferred estimation method. The choice of the estimator depends on the assumptions made about the processes generating the variables and the unknown errors (e.g., Johnston and DiNardo 1997). The estimator should be consistent with the assumptions of the model. Whatever the estimation method used, it is important that the empirical results be carefully evaluated. We have more to say about evaluation in Chapter 15.

5. The term *linear* is used in several ways in econometric models. A model can be linear in the variables and in the parameters. This is implied by the linear regression model equation; $Y$ is a linear function of each of the $X$'s. From a statistical estimation point of view, however, the key question is whether the model is linear in the parameters. No estimation problems need arise from the use of non-linear functional forms. For example, the analyst can make $Y$ a function of the logarithm of $X$, and this equation, although non-linear in the variables, is linear in the parameters. The equation is said to be "intrinsically linear," and the analyst can specify various curvilinear relationships (Table 14-2, later in the chapter). Non-linear estimators are required when the model is "intrinsically non-linear," i.e., non-linear in the parameters. An example of an intrinsically non-linear model is when the parameter of one regressor is the exact square of the parameter of a second regressor. To fit the model, the restriction of the equality between the level and the square of the parameter should be imposed. Such estimators are not of concern for the purposes of this book.

## Getting Started

### Background Comments

High-quality empirical results depend on serious scholarship. Casual analyses are unlikely to produce useful results; hence, the starting point for price analysis is a precise problem definition. What is the focus of the research? This is perhaps clarified by asking, what decision will be influenced by the analysis?

An example of a question that is too general is, did the demand for beef in the United States have a structural change in the 1975–2012 time period? Superficially, the question sounds specific, but much experience has demonstrated that the answer is conditional on the model specification (Davis 1997).[6] To evaluate the model, the researcher needs to ask, why do we need to know whether or not a structural change has occurred? If, for example, a group of farmers were interested in the effect of advertising the generic commodity beef, then one focus of the research would be to obtain a stable estimate of the coefficient(s) relating the variable(s) measuring advertising expenditures (or some other measure of advertising effort) to the demand for beef. This estimate should be obtained conditional on the other important factors influencing beef demand. In this context, it would be important to know that the estimate was stable, i.e., that the advertising parameter was constant (or changes in a known way) and hence was applicable to future decisions about advertising.

Given a specific research objective, the analyst must obtain the necessary background for conducting high-quality research. This includes a good command of economic theory and associated techniques; a deep knowledge of the economic sector under analysis, including any special events or government policies affecting the sector; knowledge about past research on the topic; and an understanding of the attributes of the data to be analyzed. Among other things, it is helpful to confirm key past empirical results in the literature (Tomek 1993). The researcher should build on past work and make sure that the new results have the potential to add to knowledge.

Model building may be viewed as having two parts. The first part involves the specification of the economic model, that is, the general economic relationships. Economic theory can be written in terms of functions and the variables they contain. The second part of model building involves the explicit definition of the equations that are to be estimated. This includes many details, some of which are discussed later.

6. A common definition of a *structural change* is that one or more parameters of the model have changed. Indeed, a common implicit assumption of the typical linear statistical model is that all the parameters are constants over the sample period, i.e., are invariant over time. A statistical test of the null hypothesis of "no change in parameters over the sample period" is conditional on the correctness of the model specification. If the null hypothesis is rejected, this may mean that one or more parameters have changed, but it also may mean that the conditioning information is wrong—the model is mis-specified. If a logical reason exists for believing that one or more parameters have changed at some point in time, this can be modeled and the hypothesis tested.

A common practice has been to conduct an empirical search for the "best" model using a fixed data set. The technical term is *pretesting* and is sometimes called "data mining." It has long been known that if an analyst selects a model by pretesting, the level of type I error for tests of hypotheses based on the final model will be very large (Wallace 1977). A test conducted at the nominal 5 percent level of significance, for example, may have a true significance level of 40 percent, 50 percent , or more. From a practical point of view, this means that if the analyst searches over a sufficient range of alternative models, one can be found that fits the particular data set. But the result may very well be an artifact of the particular data set.

In this context, we highly recommend that price analysts develop a modeling philosophy and follow it carefully. Philosophies consistent with both classical and Bayesian statistical methods exist, but space limitations do not permit their discussion here (see Geweke and McCausland 2001; Spanos and McGuirk 2001). The point we wish to emphasize is that model specification should *not* involve sitting at a computer screen and conducting a mindless empirical search. The modeling process should have a logical foundation. If the researcher cannot reconstruct the exact process leading to the final model, then a problem exists.

Given the equations to be estimated, the next step is to obtain observations on the variables and to estimate the parameters of the model. In practice, the specification of the model is likely to be influenced by the data available. For example, if observations on certain concepts are not available, proxy variables may be constructed. In a supply analysis, for example, producers probably respond to their expectations about prices, but their expectations are not observable. Thus, the researcher must be concerned about minimizing errors in measuring variables and about possibly omitting important effects. We will have more to say about data quality in a subsequent section.

A final step in price analysis is the evaluation of the results. Modern practice requires that the preliminary final model be subjected to a battery of statistical tests of adequacy (e.g., McGuirk, Driscoll, and Alwang 1993). The results become final only if the proposed model can pass the tests. Of course, the ultimate test of the results is the ability to meet the research objective. Given the current state of economic analysis, there is no absolute criterion for judging the model. Rather, results are judged relative to their intended use.

## Elementary Model Specifications

The economic principles discussed in earlier chapters provide a general guide to the functions and variables that are appropriate for a particular research problem. For example, in a market demand relation, quantity is a function of the product's own price, the prices of other products, income, and population size. Theory also suggests that a product's quantity and price are inversely related (negative sign) and that for most products, income and quantity are positively related. Theory does not tell us the precise functional relationship among variables, but it may provide some general guidance about functional form. Linear relationships are the simplest specification, but non-linear functions are logical in some applications (discussed further below).

To illustrate the use of models, we begin with a simple model of competitive price equilibrium using linear equations. A static equilibrium, which simultaneously determines price and quantity, can be defined by two equations, with a third defining that quantity supplied equals quantity demanded, i.e., an equilibrium condition.

$$A_t = \alpha_0 + \alpha_1 P_t \text{ (supply equation).}$$

$$A_t = Q_t \text{ (quantity supplied equals quantity demanded).}$$

$$P_t = \beta_0 + \beta_1 Q_t \text{ (inverse demand equation),} \quad \text{and } \beta_1 < 0.$$

The $t$ subscripts indicate that price ($P$) and quantity ($Q$ or $A$) are observed within some specific time period. They are simultaneous determined, and with no other information, we would not expect price and quantity to change.

An initial modification to this model is to permit shifts in the level of the supply and demand functions because of observable changes in the determinants of the functions. For instance, we assume that the variable $R_t$ has a positive effect on the supply of the commodity, and we further assume that the level of $R$ in time $t$ is not influenced by the commodity's price or quantity in time $t$. Perhaps $R$ is an index of weather conditions such that favorable weather increases output; clearly the weather is not influenced by the product's price. The supply equation now is

$$A_t = \alpha_0 + \alpha_1 P_t + \alpha_2 R_t.$$

As $R$ increases, the level of $A$ increases, and vice versa (Figure 14-1). Other potential supply shifters are the prices of inputs, the prices of other commodities competing for the same resources, the effects of price risk, and technology (Chapter 4).

Theory does not exhaust the factors that explain the level of supply in a particular time period. Logic and knowledge of the economic sector can suggest that lagged effects are important, that government policy has influenced supply, etc. In the general-to-specific modeling philosophy (Hendry 1995), the analyst starts with a general, feasible specification, and a logical process is used to see whether simplification of the specification can be accomplished without losing important information. A Bayesian philosophy attempts to explicitly incorporate non-sample information into the estimation procedure (Geweke and McCausland 2001).

No specification, however, can find a set of observable information that exhaustively explains (say) the level of supply. Thus, an error term is appended to the structural equations, and the foregoing supply equation can be rewritten

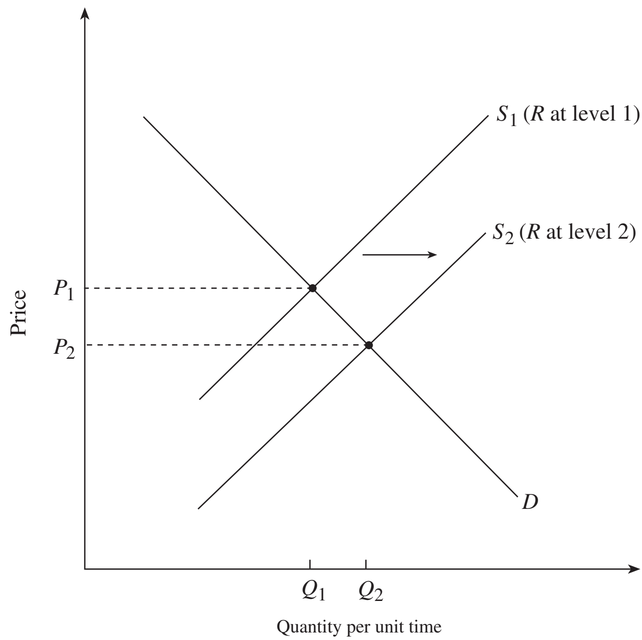



**Figure 14-1.** Changes in price and quantity as the result of changes in the variable $R$

$$(1) \qquad A_t = \alpha_0 + \alpha_1 P_t + \alpha_2 R_t + e_t$$

The level of supply depends not only on observable variables but also on unobservable random events.

If the analyst is considering a grain market, then changes in inventories and changes in imports and exports must be considered in a realistic model. An identity implied by the balance sheet in Table 14-1 may be written

$$(2) \qquad I_t = A_t + M_t + I_{t-1} - X_t - Q_t,$$

where $I_t$ is ending inventory, $I_{t-1}$ is beginning inventory, $X_t$ is exports during time $t$, $Q_t$ is domestic consumption in time $t$, $M_t$ is time $t$ imports, and as before, $A_t$ is current production. This makes it clear that a complete model probably must model the effects of inventory changes and the international market, as well as domestic demand and supply.

In the terminology of econometrics, Equation (1) is an example of a behavioral equation. Demand, consumption, and investment functions are other examples of behavioral equations. Behavioral equations contain error terms, and the structural parameters must be estimated. Equation (2) is a definitional equation, or identity. Just as 1 + 1 = 2, Equation (2) states that, by definition, the equality of the various quantities must hold.

Returning to the model depicted in Figure 14-1, the equations are

$$(3) \qquad A_t = \alpha_0 + \alpha_1 P_t + \alpha_2 R_t$$

$$(4) \qquad A_t = Q_t$$

$$(5) \qquad P_t = \beta_0 + \beta_1 Q_t, \quad \beta_1 < 0.$$

If the model is a correct representation of the real world, then price does not determine quantity nor does quantity determine price; rather, the two are simultaneously determined within time period $t$. When $R$ changes, both $P$ and $Q$ change; over a discrete time interval, $P$ and $Q$ interact to determine the new equilibrium. The variable $R$ is a measure of weather conditions, and consequently, it is determined outside of the market under analysis. It influences prices and quantities, but it is not in turn influenced by them.

By appending error terms to the supply and demand Equations (3) and (5), a simple simultaneous system of equations can be obtained. The variables that are determined simultaneously, $P$ and $Q$, are called current endogenous or jointly determined variables. These are variables that are being explained by the model. Variables that influence the endogenous variables but whose own level is determined by factors entirely outside the economic sector under consideration are called exogenous variables. By definition, $R$ is exogenous in the model.

If the supply equation also contains lagged price, this is a lagged endogenous variable, which is also sometimes defined as a predetermined variable.[7] Its value in time $t$ has been determined by what happened in time $t - 1$. With the lagged endogenous variable, the supply equation becomes

$$A_t = \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \alpha_3 R_t + e_{t1},$$

where the subscripts on $e$ represent the $t$th time period and the first equation in the model.

The cobweb model, defined in the appendix to Chapter 9, is an example of a recursive model. In a recursive model, the current endogenous variables are determined sequentially rather than simultaneously. If in the immediately foregoing supply equation $\alpha_1 = 0$, then

7. A variety of terminology exists in the econometrics literature about exogeneity. The variable $R_t$, as defined in this chapter, is strictly exogenous; its value at time $t$ has a zero covariance with all the errors, past, present, and future. A full discussion of various types of exogenous variables is beyond the scope of this book. The statistical issue in fitting the model is whether the right-hand-side (explanatory) variables are or are not related to the error term. The OLS estimator assumes that these variables are unrelated to the true error term—they are the so-called independent variables.

$$A_t = \alpha_0 + \alpha_2 P_{t-1} + \alpha_3 R_t + e_{t1}.$$

This specification is appropriate if a time lag exists between the decision to produce and the realization of output and if producers use naive expectations. With $A_t = Q_t$,

$$P_t = \beta_0 + \beta_1 Q_t + e_{t2}, \quad \beta_1 < 0.$$

Thus, as explained in Chapter 9, price and quantities are determined in a sequential process, but in the foregoing model, the levels of prices and quantities are also influenced by the random shocks of the $e_{tm}$'s and the effect of $R_t$. The paths through time will no longer be exact.

Recursive systems, although little discussed in general economics, are quite important for agricultural commodities. This is because the production side of the market has significant time lags (Chapter 4). The quantity produced does not depend on the current price. It still may be the case that the allocation of production and beginning inventories among domestic uses, exports, and ending inventories are simultaneously determined by prices. Likewise, depending on the length of the time unit (the frequency) used in the analysis (say, a quarter or a year), the level of imports may depend on the current price. The general point is, then, that a commodity model may contain important recursive components but that simultaneity also may exist. These are decisions that must be made by the analyst in the model specification.

## Organization and Data

The structure of markets for agricultural commodities is obviously more complex than the elementary models discussed so far. This complexity presents analysts with challenges. Many decisions are required before estimation begins. We have already mentioned the need to specify equations and the variables in them. We have also implied that the variables need to be classified as current endogenous and predetermined or exogenous. This, in turn, is related to the nature of causality among variables: what variables are explained by the model, and what are the explanatory variables? Further, the specification needs to recognize the unique factors of importance in the market under analysis. The analyst also must make decisions about many details like functional form, lags, and deflating.

The foregoing means, as already noted, that the researcher should have a method for organizing his or her thinking and making decisions. A tentative model should be outlined that is consistent with the research objective. This might be done by developing a diagram to describe relationships among variables and/or using general functional notation to write out a proposed model. Then, careful attention must be given to variable definitions and data selection. The model specification is a statement of the analyst's total body of knowledge; from a statistical perspective, the model is the "maintained hypothesis" that is thought to be correct. Subsequent statistical inferences are conditional on this specification. Model building is as much an art as it is a science. This is why the researcher should be expert in the commodity or sector being investigated.

The analyst should have a good understanding of the data. Although seemingly uncommon, we believe that data description is a helpful, early step in research. Time-series observations can be plotted against time, the mean and variance can be estimated, and/or the probability distributions can be constructed. Indeed, a structural model should be able to reproduce the properties of the endogenous variables.

Scatter diagrams of important relationships among variables are a potentially useful tool in price analysis (e.g., Figure 14-2). The simple relationship between two variables can help identify possible shifts in the relationship. In Figure 14-2, the price-quantity observations for pork in the 2006–2012 period lie to the left and below those from earlier years (continuing the shifts apparent in the comparable figure in the fourth edition of this book). Indeed, if we include even earlier observations, it appears that the demand for pork in the United States has been declining since 1975. The issue is how to model these changes.

A scatter diagram depicts only the simple relationship between two variables; the net relationship, after taking account of the influence of other variables, may differ from the simple two-variable relation. Thus, the two-variable diagram is

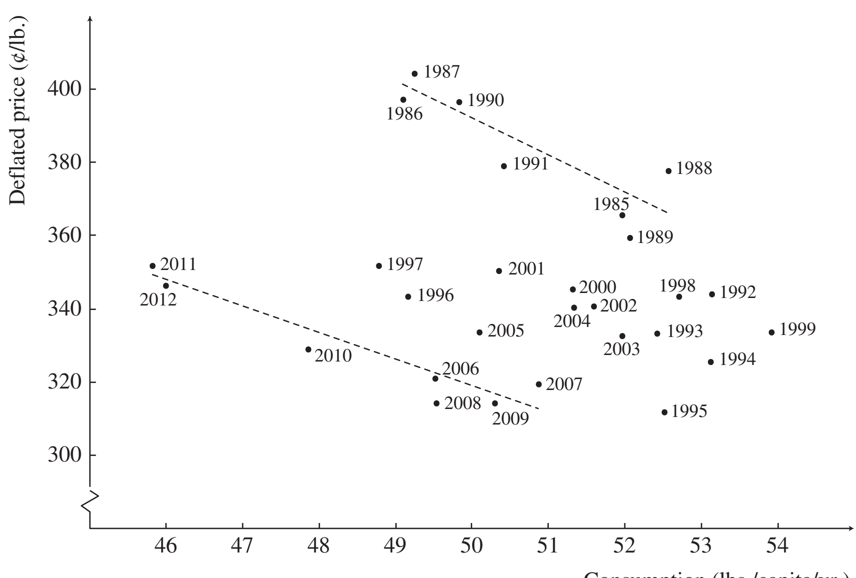



**Figure 14-2.** Relationship between retail price of pork (deflated by the CPI, 1982–1984 = 100) and per capita consumption of pork, United States, 1987–2012

misleading if the analyst is interested in the relationship between two variables net of the effects of other variables. It is possible, with modern computer software, to obtain plots of relationships between two variables that are net of the effects of other (specified) variables, called partial regression leverage plots. They can help identify influential observations and outliers.

Observational data generated by government agencies (or other secondary sources) should not be accepted passively. The careful analyst will have learned how the data are constructed. Beef consumption, for example, might be better labeled beef disappearance. In the United States, it is computed as a residual from a balance sheet of quantities. The USDA has estimates of inventory changes, net exports, and production; these data are used to compute "consumption," i.e., what has apparently disappeared into the domestic market.

Understanding how data are constructed helps identify the potential weaknesses of the data. Gardner (1992) provided examples of how data can mislead. Among them, he described why the Minnesota-Wisconsin price for grade B milk, reported by the USDA's National Agricultural Statistics Service, is not an average of prices actually paid for milk. As Gardner states, "the supine acceptance of data can cause scientific mischief of the worst kind—believing we know things that aren't so" (1992: 1066).

Agencies compiling data often revise the observations and sometimes change the definition of the variable or the method of computing it. Thus, the researcher must be sure that the data are internally consistent over the sample period. The most recent data sources should be used, and the analyst should work backward, as necessary, checking that observations for the earlier years are consistent with the current data. Data revisions alone can cause considerable changes in the estimates of parameters (Tomek 1993). One of the reasons that it is difficult to duplicate published empirical results in economics is that the original data set is difficult to duplicate. If the original analyst has not saved the data and if the data have subsequently been revised, then the person trying to confirm the earlier results cannot do so.

## The Identification Problem

If the research problem requires the estimation of structural equations and if observational data must be used, the question arises, is it possible to obtain estimates of structural parameters from secondary data? For example, if the researcher wants to estimate the price elasticity of demand for pork and if the data consist of a series of annual average prices and annual consumption (perhaps on a per capita basis), as in Figure 14-2, can these observations be used to estimate the elasticity? Each price and quantity pair may be viewed as representing an average equilibrium for the specified time period, and the identification problem is the question, can a demand equation be estimated from these data? It certainly is possible to compute coefficients using a statistical

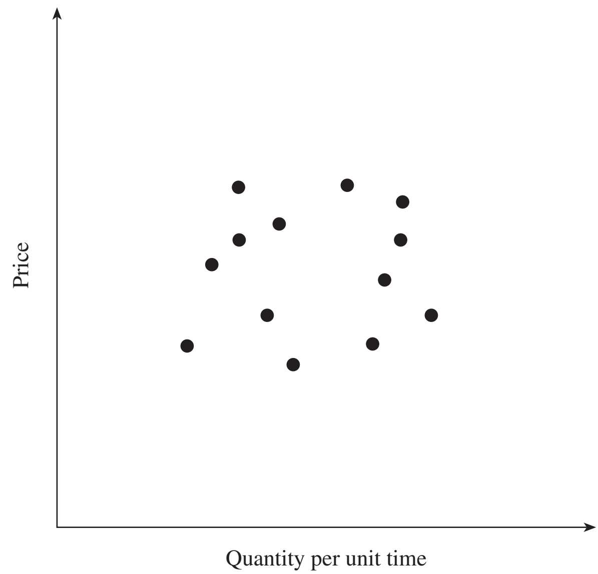



**Figure 14-3.** Scatter diagram with no identifiable relationship

estimator, but this does not guarantee that the coefficients can be interpreted as estimates of the desired demand parameter(s) that meet the research objective.

To illustrate an answer to the question, a competitive market structure for a single commodity is assumed. Then, if the economy were perfectly static, the observed outcome would be a single price and quantity—the point of equilibrium of static supply and demand functions. In this case, it is impossible to estimate either a demand or a supply function. An infinite number of functions can be constructed to go through the single point of equilibrium.

Of course, the economy is not static, and with the passage of time, supply and demand functions are expected to change (shift). These changes may help identify one or the other or both of the functions, but this is not guaranteed. The price-quantity observations might form a "shotgun" pattern, as in Figure 14-3, and without additional information about how the data are generated, neither a supply nor a demand equation can be inferred from the data. The scatter of observations is presumably the outcome of many combinations of changes in demand and supply, and this means that additional information must be obtained (or assumed) about how the observations were generated. In other words, the analyst must have a theory about the data-generating process for the endogenous variables, i.e., a model.

If the demand curve shifts through time and the supply curve remains stable, then the observations trace out a supply curve, as in Figure 14-4. If the supply curve shifts and the demand curve remains stable, then the observations trace

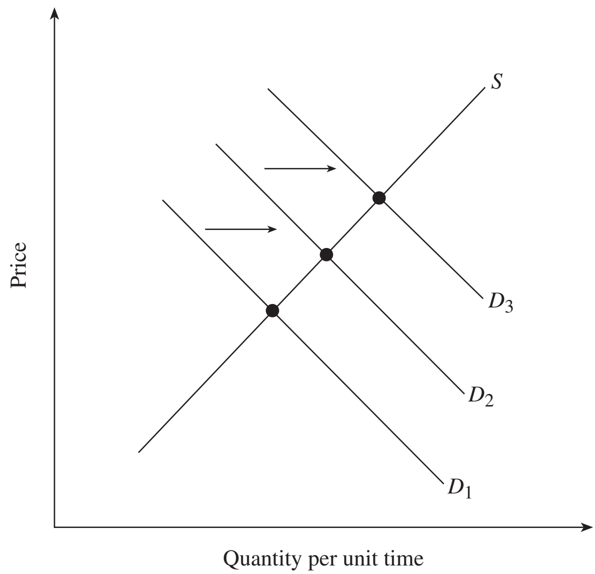



**Figure 14-4.** A supply relationship identified

out a demand curve, as in Figure 14-1. That is, from the viewpoint of Equations (3) and (5), it is the systematic shifts in the exogenous variable, $R_t$, in the supply equation that identifies the demand equation. The hypothetical data-generating process assumes that prices and quantities change only because of the changes in the level of supply.

The additional information necessary to achieve identification in the foregoing cases is the knowledge that one function is stable while the other is changing. In principle, this could also arise from information about the effects of the random error terms ($e_{tm}$), which was the approach taken by one of the earliest discussions of the problem (Working 1927). It is also possible to show that, if the effects influencing the level of supply and demand are correlated, this complicates achieving identification. In Figure 14-5, the shifts in demand and supply have a perfect negative correlation. The resulting data points have a negative slope, but the slope of the data points is not the same as the slope of the demand (or supply) function.

Our discussion is based on over-simplified supply-demand specifications. Nevertheless, the discussion illustrates two points. First, to achieve identification, the supply and demand functions must have different determinants, and second, these determinants cannot be highly correlated with each other. In other words, the data-generating process must be such that it is possible to disentangle the exogenous factors influencing the levels of the different functions. In principle, this seems possible. If income can be treated as an exogenous factor

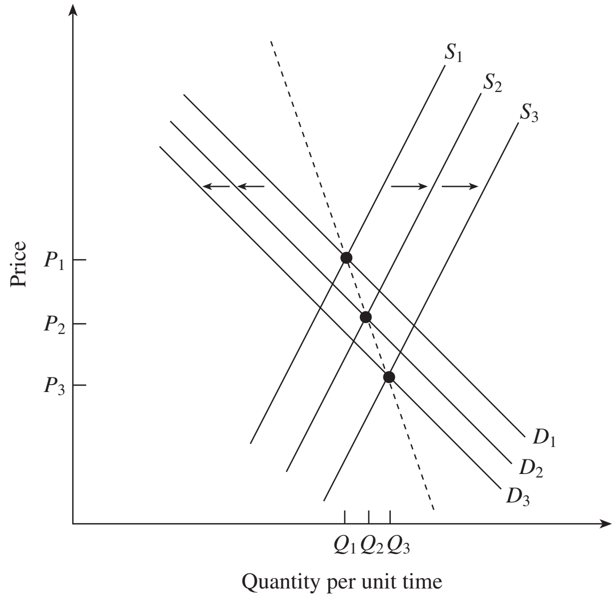



**Figure 14-5.** Correlated changes in supply and demand, and neither a supply nor a demand curve identified

influencing the level of demand and if technological change can be treated as an exogenous factor influencing the level of supply, then it would seem that both supply and demand are identified. But there are at least two potential complications. It could be that, in reality, income changes and technology changes are highly correlated; their separate effects may be difficult to disentangle. Moreover, it is not necessarily the case that these two variables can be treated as exogenous. The level of income, for example, is dependent on the prices and quantities of the goods and services in the economy.

The identification problem is a general one in the estimation of structural economic relationships, but it is usually considered explicitly only in simultaneous systems of equations (e.g., Johnston and DiNardo 1997: 309–314). A common approach in a simultaneous system is to specify that exogenous variables in the system appear only in selected structural equations of the system, i.e., are specified to have zero parameters in other equations of the system. These are called exclusion restrictions. In the earlier example, $R_t$ appears in the supply equation but not in the demand equation. Thus, one way of thinking about the problem is that identification requires having a sufficient number of correct exclusion restrictions on variables that are truly weakly exogenous (see footnote 7) to achieve identification for each equation. Certainty about identification requires knowing the true model, which of course is impossible. Even though reasons exist to doubt whether or not structural equations are truly identified in the context of time-series observations (Sims 1980), it is rather common for analysts to proceed "as if" the equations are identified.

The identification problem arises in the context of estimating the model in its structural form, sometimes called the primary form. The problem is presumably to estimate the structure. Assuming the equations are linear in both the parameters and the variables, it is straightforward to solve the primary form for the reduced form of the system. The *reduced form* makes each current endogenous variable in the system a function of all the predetermined variables of the system—the lagged endogenous and exogenous variables—that are all assumed to be weakly exogenous.

As an example, we can write a supply and demand model where each equation contains one excluded exogenous variable and where we assume that quantity supplied equals quantity demand ($A_t = Q_t$). Thus, we write the structural form as

$$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 R_t + e_{t1}$$

$$P_t = \beta_0 + \beta_1 Q_t + \beta_2 Y_t + e_{t2}, \quad \beta_1 < 0,$$

where $Q$, $P$, and $R$ are as defined previously and $Y$ is an additional exogenous variable that appears in the demand function but not the supply function. It is commonly assumed that each of the equation error terms has a zero mean, constant variance, and zero lagged covariances, as discussed earlier. Now, with a system of equations, we also need an assumption about the possible relationship among the error terms of the different equations. For a simultaneous system, it is commonly thought (and assumed) that the covariances are not zero. In other words, it is assumed for our simple example that $e_{t1}$ is related to $e_{t2}$ in time period $t$.

The reduced form of the model is

$$Q_t = \pi_{10} + \pi_{11} R_t + \pi_{12} Y_t + w_{t1}$$

$$P_t = \pi_{20} + \pi_{21} R_t + \pi_{22} Y_t + w_{t2}.$$

The $\pi$'s depend on the $\alpha$'s and $\beta$'s (and likewise the $w$'s are related to the structural equation error terms and parameters). Readers are encouraged to do the explicit algebra to show that this is true.

One reason for solving for the reduced forms is to demonstrate that the data-generating process for the current endogenous variables can be viewed in alternative ways, and it is possible to write out the reduced forms even if one or more of the structural equations are not identified. If the example model is true, the analyst is claiming that the two endogenous variables depend linearly on changes in the exogenous variables, $R$ and $Y$, and on the random effects of the (correlated) error terms, $e_{tm}$, combined in the $w_{tm}$. If the model is to be used to forecast $P$ and/or $Q$, the analyst must model $R$ and $Y$ to estimate future levels of $R$ and $Y$ as a basis for the forecasts. Thus, even though the exogenous variables are not necessarily explicitly modeled in the structure, some knowledge of their behavior is important. In economics, the exogenous variables are commonly random variables; they are not under the control of an experimenter. To repeat an earlier definition, $R$ and $Y$ are assumed to influence $P$ and $Q$, but not to be influenced by changes in $P$ and $Q$.

As it happens, both structural equations in the previous example are identified, but, if $\beta_2 = 0$ (and hence the supply equation is not identified), it still is possible to solve $Q$ and $P$ for $R$. Thus, one could still use the reduced form to predict $Q$ and $P$, even though it is not technically possible to recover estimates of the parameters of the under-identified supply equation, the $\alpha$'s. Note further that, if the variable $Y$ logically appears in the supply equation but not in the demand equation, the demand equation is still identified; in this example, it is over-identified. The supply equation remains under-identified, but $P$ and $Q$ now depend on $R$ and $Y$ in the reduced forms. The reduced-form parameters are identified.

The reduced-form equations also can be used to emphasize the importance of internal consistency in the model. The endogenous variables are random variables with certain statistical properties; they may or may not have constant means, variances, etc. But whatever the statistical properties of the left-hand-side variables (of the reduced forms), the specification of the right-hand side must be consistent with this behavior. If $P$ is not trending, then trends in the right-hand-side variables are inconsistent with this observed behavior unless two explanatory variables can be shown to have offsetting effects. Or, if $P$ is trending but the observable explanatory variables are not trending, this implies systematic behavior in the reduced-form error term, which is implausible, implying an error in specifying the right-hand-side variables. This discussion reemphasizes the importance of having a descriptive understanding of the time-series behavior of the variables in the model.[8]

## A Recursive System

Because of the time lags in the production process, the supply side of agricultural markets, or at least the domestic production component, is likely to be recursive to the demand side of the market. Current production is determined by past decisions and can be changed little, if at all, by current prices. As noted before, the demand side may have simultaneous components. The current domestic apple crop is fixed in size each year by decisions made in the past. The crop, however, can be allocated among fresh fruit and processing uses.

8. The literature on stochastic trends, integrated series, spurious regressions, and error correction models all relate to the nature of the behavior of the random variables in models. This literature is beyond the scope of this book (e.g., see Hamilton 1994).

Within the year, the fruit can be stored, and inventories of processed products exist. Also, apples and apple products are imported and exported.

For some markets and research problems, such simultaneity may not be important. In a fully recursive model, there is a causal chain of endogenous variables. Current supply depends on past prices, current supply (along with demand) determines the current price, and this price determines future supply, etc.

The beef, pork, and chicken markets have sometimes been modeled as recursive. Current (say, quarterly) domestic production is determined by decisions made many months (even years) in the past, and consumption (domestic disappearance) is computed from a balance sheet. $Q_t = I_{t-1} + A_t + M_t - X_t - I_t$, where $Q$ = consumption, $I$ = inventory, $A$ = domestic production, $M$ = imports, and $X$ = exports. In the United States in 2012, domestic consumption of pork was 78 percent of the predetermined supply (beginning inventory plus production), net exports (exports minus imports) were about 19 percent of supply, and the residual amount was accounted for by ending inventory. Since consumption is a fairly large proportion of pork production and since inventory changes from quarter to quarter (or year to year) are small, analysts sometimes argue that consumption is largely predetermined by the available supply. But, since prices are determined jointly with domestic consumption, exports and imports, and inventory changes, one can also argue that domestic consumption and prices are simultaneously determined. It is partly a matter of judgment as to whether consumption can be treated as determining price in an inverse demand function or whether price and consumption should be treated as simultaneously determined. These judgments will probably vary from commodity to commodity. Statistical tests are also available to evaluate whether a variable is exogenous.

Recursive models, if applicable to the market under analysis, are appealing. The parameters of the equations in the structural model are identified, and each structural equation in the model can be estimated appropriately by the method of OLS. For this to be true, not only do the current endogenous variables have to be determined in a logical causal chain, but the error terms of each equation cannot be correlated with each other. This means that the random events affecting demand do not affect supply. (This is in contrast to the simultaneous system, discussed earlier, in which the error terms are assumed to be correlated with each other.)

If the system of equations is recursive, then a single equation from the system can be estimated separately. For instance, the literature contains price-dependent, single equations that have been interpreted as demand functions. The remainder of the model need not be estimated unless required by the research objective. We now turn to a discussion of selected equations in a model.

## Market Supply Equations

The complex production process for agricultural commodities permits substantial latitude in building supply equations. Careful definition of terms is essential. Total supply in a specific time period may depend not only on current production but also on carryover stocks and imports. The domestic production, inventory, and trade components should be modeled in separate equations. In what follows, we concentrate on models of domestic production.

Production equals yield per unit multiplied by the number of units. Units may be in area or animal numbers. For instance, total milk production equals the number of producing cows multiplied by the average production per cow. Thus, although an analyst may make total production dependent in one equation, it may be more informative for the analyst to model yields and units separately in two equations. For animals, the numbers culled and added to the herd might be modeled separately.

If production is modeled in one equation, the analyst still must distinguish among total production; harvested production; commercial production entering the marketplace, sometimes called "marketings"(as opposed to total output, including that used on farms); and/or sales (allocations) to particular markets, like fresh fruit versus processing. Although total production is not influenced by current price, the quantity harvested and/or allocated to alternative uses can be influenced by current price. Thus, model specification is influenced by the precise definition of the quantity variable being modeled.

The producer usually has more complete control over the area (acres or hectares) planted or the number of female animals in the breeding herd than over yields. Thus, variables like acres planted or sows farrowing are often used as the dependent variable when the research problem involves estimating the supply response of farmers. Crop yields, in particular, are influenced by factors over which the producer has little control, like weather conditions. Of course, the producer can influence the level of soil fertility and other management practices, and for livestock, the producer can alter feeding regimes. But yield equations have proven difficult to specify. With improvements in technology and a producer's managerial ability, yields do trend upward, but aside from trend, other variables often are not statistically important.

We turn, then, to equations that explain units such as area planted or sows farrowing. The general nature of the explanatory variables is suggested by theory and logic. As discussed in Chapter 4, current output is determined by expectations held at the time that production decisions are made. $A_t = S(P^*_{t-1})$. Price expectations, $P^*_{t-1}$, have been modeled in four ways: naive, adaptive, rational, and quasi-rational (discussed in Chapters 4 and 9). Moreover, individual farmers are likely to form expectations in different ways, and market outcomes are an aggregation of individual decisions (Chavas 1999).

Space limitations do not permit us to describe specific measures of expectations in detail; however, naive, adaptive, and quasi-rational expectations all use current (at the time the decision is made) and possibly past prices. The way that past prices enter the model does vary with the alternative definitions.[9] If a futures market exists for the commodity under analysis, then prices for future delivery are one way to measure the rationally expected prices, although futures prices have been rather infrequently used in supply analyses (for an example, see Tronstad and McNeill 1989). The first decision is how to measure expected prices; alternative definitions exist, and the alternative definitions can result in different empirical estimates of the price elasticity of supply.

A potentially important shifter of the supply relation is price risk, since expectations are almost never realized. Empirical measures of price risk are based on historical experience of the differences between expected prices and their realizations. At time $t-1$, the realization of the expectation held at time $t-2$ is known; hence, the difference $P^*_{t-2} - P_{t-1}$ can be computed. A common practice is to construct a measure of price risk, call it $R_{t-1}$, based on such differences. Again, many decisions must be made. First, the definition of risk will vary with the definition of expected price. Second, there is the question of how to use the differences. For example, one might use the squared differences and compute a statistic analogous to a variance, but this in turn requires the decision maker to decide on how many past observations to use in computing the variance, i.e., farmers' current expectation about the level of risk that they face. Also, one can reasonably argue that, for a farmer, only the negative differences involve risk—it is the cases when the realized price is below the expected price that are risky. Alternatively, the researcher might use an average of the absolute differences, and again numerous decisions must be made, such as the length of time over which to compute the moving average. In sum, as in the case of measuring expected prices, the analyst faces many choices that may influence the empirical results.

The researcher must assure, of course, that other important determinants of supply appear in the model. These can include the prices of major inputs, the prices (or unit profits) of alternative enterprises competing for the same resources, and the effects of joint product prices (when this is relevant). In an analysis of the pork supply, for example, it is clear that feed costs are an important determinant of profitability and supply decisions; this cost is approximated by the price of corn in typical pork supply equations. In Tronstad and McNeill (1989), the number of sows farrowing in the current quarter was made a function of the expected prices of hogs and corn in prior quarters (with varying lag

9. The adaptive expectations model assumes that the expected price is a geometric weighted average of current and past prices. If price expectations are truly adaptive, the geometric weight can be estimated by including the lagged dependent variable, $A_{t-1,}$ as a regressor and using its coefficient to compute the weights (Nerlove 1956). Since the lagged dependent variable can be justified as an explanatory variable for other reasons, a significant coefficient for the lagged dependent variable does not necessarily mean that the adaptive expectations model is the correct interpretation (e.g., Hendry 1995: 232).

structures) and of the riskiness of these prices. Their research analyzed the consequences of using alternative measures of risk, based on cash and futures prices as alternative ways of measuring expected prices. The model also contains variables to capture seasonal and trend effects.

Among other things, Tronstad and McNeill concluded that asymmetric measures of price risk—those that consider only downside price risk—perform better than the symmetric measures. The evidence also seems to favor using futures prices, rather than spot prices, as measures of price expectations.

In addition, the analyst must be cognizant of the possible effects on supply of government programs (Lin et al. 2000). These programs probably affect the economic incentives faced by farmers, and if so, the program effects need to be modeled. Moreover, government programs can change, and even if the price-support level is fixed, market prices may vary such that the support level is effective in some years but not others. In the terminology of econometrics, there is a "switching" effect. Although government program effects are potentially important, a detailed description of such models is beyond the scope of this book.

Supply schedules for many agricultural commodities shift because of technological change. In fact, changes in technology often seem to be the dominant factor in explaining supply changes, particularly changes in yields. Unfortunately, there is no direct measure of "changes in technology," and analysts must resort to the rather unsatisfactory device of a proxy variable. The most common proxy is a trend variable, e.g., the numbers 1, 2, . . . , $T$. This specification assumes a smooth change in technology of equal amounts each time period. Sometimes, general measures of changes in productivity are used, such as an index of changes in productivity.

Using trend as a proxy for the concept "technology" illustrates a rather common practice in empirical price analysis. In cases where the variable itself is not directly observable, the analyst must either use a proxy variable or omit the concept from the model. This raises the question, is it better from a statistical point of view to use a proxy variable or to omit the concept? No Golden-Rule answer exists, but the common practice is to use proxy variables. If, for example, the objective is to estimate a price elasticity of supply, the estimated elasticity probably has a smaller bias with a proxy variable included in the model than if it were omitted.[10]

Current production can be correlated with past production for a variety of reasons. One set of reasons can be categorized as the costs of change. A response

10. The probability distribution of an unbiased estimator is centered on the true, but unknown, parameter. Thus, bias refers to the extent that the expected value of the estimator differs from the true value. Introducing a proxy variable is likely to reduce bias. Another property of the estimator, however, relates to the variance of its probability distribution. Introducing the proxy variable can increase the variance (relative to omitting it). Thus, given that the research involves a single estimate from a sample, there is no guarantee that the individual estimate will be closer to the true value with the proxy than without it.

to a change in expected price may be slowed by physical and climatic considerations, resource fixity, crop-rotation patterns, capacity constraints, and habitual decisions by managers. A second set of reasons (as indicated in footnote 9), is that price expectations are formed in a way that introduces lags in production responses. If adaptive expectations are assumed, the consequence is that the lagged dependent variable (say, $A_{t-1}$) is an explanatory variable. This specification often improves the statistical fit of the model, but the interpretation of the coefficient of the lagged variable is often uncertain.

The following function helps summarize some of the major considerations in specifying a supply equation,

$$A_t = S(P^*_{t-1}, R^*_{t-1}, X_{t-1}, G_{t-1}, t)$$

where $A$ is the measure of farmers' decision (say, sows farrowing); $P^*$ is the expected product price; $R^*$ is the expected price risk; $X$ is the input price (which may be more than one major input and also may include the price of other products that the farmer can produce with the same resources); $G$ is the government-program effects; $t$ is time, hence time trend and technology change; and the $t - 1$ subscript indicates a lag between the decision to produce and the realization of output (and the actual lag will depend on the commodity). Depending on the problem, variables other than price and risk may also require that expectations be measured. For example, expected costs may be relevant for some problems, or the costs may be well known at the beginning of the production process. Also, the variable $A_{t-1}$ may be included depending on how expected prices and adjustment costs are modeled. It is our view that the analyst should start with a comprehensive specification of the potential factors influencing supply and then determine if it is possible to simplify the specification. This is in contrast to starting with a simple model and adding variables if the simple specification seems inappropriate.

### Market Demand Equations

If prices and quantities are determined recursively, then price is the logical dependent variable in the demand equation(s), i.e., an inverse demand specification. In any case, even if prices and quantities are simultaneously determined, the own-price variable can be placed on the left-hand side of the equation. With more than one current endogenous variable in an equation, the researcher has a choice about which one to treat as dependent. In analyses of agricultural markets, it is common to use inverse demand functions, but with simultaneity in supply and demand, it would be possible to have an inverse supply equation. That is, in a simple two-equation supply-demand model, the endogenous quantity would be on the left-hand side of one equation and the endogenous price would be on the left-hand side of the other equation. It is mathematically convenient in a system of simultaneous equations, which has one equation for each current endogenous variable, to have each endogenous variable appear on the left-hand side only once.

Recalling that we are discussing recursive models, we think in terms of price-dependent demand equations, but one of the decisions an analyst must make relates to the definition of variables as endogenous and exogenous. The following discussion lists the types of variables that need to be considered. Some illustrations of these points are provided in subsequent chapters.

Following from basic demand concepts, the demand specification should consider the effects of income and of substitutes and complements. For a price-dependent function, the effects of substitutes and complements may be captured by quantity variables. Income is commonly measured by consumers' disposable income. Own-price and income may be deflated by an index of general prices to obtain real income and prices. (Deflating is discussed later in a separate section.) In market demand equations, at the retail level, quantities are typically converted to a per capita basis by dividing by a measure of the relevant population.

Analogous to the supply side, consumers may face costs of adjustment, and an appropriate model may require modeling the lagged effects (Chapter 2). The importance of lags will depend on the unit of measure (years, quarters, or months) and the commodity under analysis.

With the growing complexity of economies, demand specifications should consider a variety of other factors, such as changes in demographic variables (age distribution or ethnicity), generic advertising effects, etc. The particular variables to be included in the model are problem-specific. For example, the demand for fluid milk probably depends on the proportion of the population that is pre-teenagers. If, as in the United States, a generic advertising program exists for milk, it should be modeled (for problems of modeling the advertising effects for milk, see Tomek and Kaiser 1999).

For analyses of derived (farm-level) demand for commodities, the modeling strategy may be different. As noted in Chapters 2 and 6, derived demand can be approached via several conceptual models. In either approach, one is concerned with farm-level prices and quantities. If the analyst takes a fixed-proportions point of view (Chapter 6) and makes farm price a function of farm-level quantities, then theory requires that one of the regressors be a measure of the marketing costs. If the problem relates to regional demand—say, the demand for apples produced in New York State—then it is important to recognize that the principal substitute is apples produced in other regions.

The foregoing discussion does not exhaust all the detailed considerations in modeling demand, but it does illustrate several themes of this chapter. Namely, model specification is problem-specific, and in addition to a strong conceptual base, the analyst must have a deep knowledge of the sector of the economy relevant to the research problem.

## Data Selection: The Sample

We have already noted that the selection of variables and how accurately the series are measured influence results. Among the additional decisions that have to be made in using time-series observations are the following:

1. What specific years should be included in the analysis? That is, what should the length of the sample be?
2. Should annual, quarterly, monthly, weekly, or daily observations be used?

Two conflicting factors enter into the decision as to what historical time period to use. One consideration is the number of observations. Other things being equal, a large number of observations is preferred to estimate a given model with a fixed number of parameters. This implies using a long sample. A second consideration, however, is the desire to have a time period with a relatively constant structure, which will yield results relevant for decisions about the future.

Judgment about the latter factor requires expert knowledge. The analyst should be informed about important events that have influenced price behavior. These may include shocks (events) that have temporary or permanent effects. For example, the added demand occasioned by the entry of the (former) Soviet Union into the international grain markets in 1973 appears to have shifted many commodity prices to a higher level that persisted. Or changes in a government price-support policy can affect supply responses, and over time, a number of policy shifts may occur, resulting in different price regimes. In all such cases, either the effects of these events must be modeled or the sample size must be reduced to avoid the period with the potentially different structure.

In sum, expert knowledge is the starting point for data selection. Then, given knowledge about potential changes in the structural parameters, the trade-off is between having a long sample, while modeling the effects of special events, and having a shorter sample and a simpler model. If uncertainty exists about the importance of particular events, then statistical tests can be helpful. It is possible to test for the effects of particular events or structural change, but this is best done by identifying the potential dates in the time series that are relevant. The selection of the sample period should not be a routine, arbitrary decision. With an improper model relative to the sample period, empirical results will be sensitive to the deletion or addition of a single time period; i.e., the particular results will be sample-specific.

The second question relates to the frequency of observations—the unit of time covered by the observations. Often, the objective of the research dictates the choice. If the objective requires estimation of demand functions by season or forecasting a monthly price, then obviously annual observations are not appropriate. For structural modeling, however, the availability of observations on all the variables in the model is an issue. For example, observations on inventories may be available only on a quarterly basis, and unless the analyst is willing to interpolate between quarters, it will be necessary to use quarterly data.

Models for monthly or quarterly observations typically must be more complex than for annual observations. If nothing else, seasonality may have to be modeled. The increased complexity is perhaps offset by the increased utility of the model. Decisions may very well require information at the monthly level. Again, this emphasizes that modeling and data selection are conditioned by the research problem.

We close this section by reminding readers of the importance of having consistent data series. In a previous section, we mentioned the potential concern about changes in the definition of variables by the agency compiling the data and about understanding what is being measured by the data. Another type of consistency relates to distinguishing between calendar years, marketing years, and fiscal years. It also may be important to distinguish among farm-, wholesale-, and retail-level observations.

## Deflating

Much of economics is about the effects of changes in relative prices, and thus one of the issues in model specification is whether or not to deflate and by what variables. Sometimes these questions can be answered by the theoretical model being used in the analysis. That is, deflating is required by the model. For example, if the research requires the estimation of a system of demand equations and if the researcher is using the Almost Ideal Demand System, the appropriate deflator for the expenditures variable in each of the equations is the Stone Price Index (Deaton and Muellbauer 1980).

If the researcher is not specifying a full demand system but, say, is specifying a single demand equation, then price and income may be deflated by the CPI for all items. As noted in Chapter 10, the CPI is intended to measure the general price level faced by consumers—their cost of living. Thus, if $Y_t$ is personal disposable income and $\text{CPI}_t$ is the CPI observed at time $t$, then $Y_t/\text{CPI}_t$ is a measure of real disposable income. This may very well be a reasonable approximation of the income variable that influences consumers' decisions. Likewise, dividing the nominal price of a single good by the CPI gives a measure of the movement of the good's price relative to all other prices. The CPI is a representation of the aggregate prices faced by consumers.

The CPI is not, however, a logical deflator for every empirical application. For a supply analysis, profitability depends on the product's price relative to the prices of inputs. Opportunity costs depend on the prices of other products. The logical deflator for a supply equation may not be obvious, but it probably involves an index of prices paid by farmers, a related sub-index, and/or an index of prices received by farmers.

In our view, some analysts have not made thoughtful decisions about the selection of deflators. For example, a number of empirical analyses of commodity supply have used the CPI as a price deflator with little apparent theoretical justification. Although various price indexes can be highly correlated, this has not always been true in the United States (Chapter 10). The choice of deflator can cause large differences in the empirical results.

Deflating influences the time-series properties of the resulting variable. In the United States, for example, nominal commodity prices are not necessarily characterized by trends; they tend to fluctuate around a constant mean for considerable periods of time (Chapter 9). If a price with a constant mean is deflated by an index, like the CPI, that has an upward-trending mean, then the deflated price will trend downward. If current and lagged deflated prices are used to represent farmers' expectations, the implication is that farmers expect prices relevant to their production decisions to trend downward. When forecasts are made for a point in time some distance beyond the end of the sample using the downward-trending price series, illogical results may be obtained. A downward-trending real price is potentially inconsistent with a dependent variable (say, acres planted) that is not trending or is trending upward.

By deflating, one seeks to obtain the best possible measure of real prices (or whatever concept is needed). In statistical terms, "best" means minimizing measurement error. The analyst wants to construct the variable that comes closest to measuring the relative prices to which decision makers are responding. The selection of a deflator should be a thoughtful decision based on economic logic.

The use of ratios as explanatory variables is also sometimes justified for statistical reasons. One alleged benefit of using ratios is to deal with the problem of high correlations among explanatory variables. The potential problem—called multicollinearity or high intercorrelations—is that two or more explanatory variables are so highly correlated that it is not possible to obtain precise estimates of their separate effects. If a ratio of prices is used, their separate effects are combined into one variable, which superficially may appear to be a solution to the problem. It is not, however, for reasons analogous to those discussed in the section on first differences (later in the chapter).

The second statistical problem that may justify the use of ratios relates to nature of the variance of the error term in the statistical model. A common assumption of estimators, like OLS, is that this variance is a constant. If it is not and if it is correlated with one of the regressors or the square of the regressor, then a data transformation involving a ratio of the variable, influencing the variance of the error term, or its square root will make the variance of the error term a constant. Readers interested in a discussion of this topic should consult the section on heteroscedasticity in an econometrics textbook.

On balance, deflating will more likely be justified in price-analysis applications by economic logic than by statistical reasons. In this context, we conclude this section by calling attention to a practical problem associated with deflating. Namely, if the price index used for deflating contains the price series being deflated, then the coefficients relating the deflated variable to other variables are likely to be biased. This problem is a variant of the discussion of the effect of deflating on the time-series properties of the deflated variable. An errors-in-variables problem is created.

The amount of bias depends on the relative weight that the price series being deflated has in the index. If it is a large component of the index (e.g., the price of beef in an index of U.S. meat prices), then the bias is large. A linear regression equation with price as a function of quantity illustrates the point. A small quantity implies a high product price, and this also means a larger price index. The larger index, however, means that the ratio is reduced. In other words, since the index changes because of the change in a component's price, deflating by that index tends to cancel the influence of the price change. The ultimate case is when the index is based solely on the commodity price used to construct the index. Then, deflating will result in a real price that is a constant equal to the nominal price in the base year in the index. From the viewpoint of economics, the ratio does not involve the appropriate relative prices. From the viewpoint of statistical estimation, the explanatory variable will have little or, in the ultimate case, no variability, which means that it will not be possible to obtain a precise estimate of the coefficient(s) associated with the deflated variable.

## Other Data Transformations

In transforming the original data, the researcher must keep two questions in mind: what is the purpose the transformation, and what are the statistical consequences of the transformation? We start by reminding readers that transformed variables have different time-series properties than the original series. The mean and variance change, autocorrelation may be increased or reduced, etc. Transformations are presumably intended to be a benefit to the analysis, but they can also create problems. For example, differencing and summing operations can introduce spurious cycles into the transformed data. Harvey (1981: 81–83) illustrates how such transformations have probably led to erroneous conclusions about the existence of 20-year cycles in economic data.

### Alternative Functional Forms

The researcher must decide which specific functional form(s) is (are) appropriate for the behavioral equation(s) being estimated. The criteria used in selecting functional forms can be conveniently divided into three categories. The first is the mathematical properties of the function relative to the logic underlying the economic relation. A consumption function illustrates the point. A household's consumption of an individual product may be zero at and below some level of income, but as income increases above this level, consumption normally increases. This growth is likely to level off and approach a maximum at high levels of income. For some products, consumption may even turn down after reaching a particular level of income. Thus, in specifying a consumption-income relation, the functional form should be flexible enough to intersect the income axis (where consumption equals zero) at a positive level of income and then curve upward (a changing positive slope) approaching a maximum. Since a number of functions have this general shape, logic still does not dictate the precise functional form to be used.[11] The non-linear supply of storage function discussed in Chapter 12 is another example of theory suggesting a curvilinear equation, but not the specific form.

The second category of criteria is based on statistical tests and fits of the equation. The idea is to select the functional form that "best" fits the data, but the definition of *best* varies with the statistical criterion used. One example is the use of the $R^2$ statistic (coefficient of determination) corrected for degrees of freedom (discussed in Chapter 15). Valid comparisons of alternative functional forms can be made, provided that the dependent variable is not transformed from one functional form to another. In this case, the question is merely, which set of right-hand-side variables (which change with the transformations involved in the selected function) gives the largest explanatory power? If the dependent variable is transformed, then $R^2$-type criteria are difficult to use because one is now comparing apples and oranges. For example, the logarithm of the dependent variable has a different variance than the original variable, and thus different levels of variability are being "explained" in the alternative models. A considerable literature exists in econometrics about the choice of functional form, but a full discussion is beyond the scope of this book.

The third category of criteria in selecting a functional form is simplicity, with the simplest form being preferred. Some common functional forms used in empirical price analysis are summarized in Table 14-2. The slope coefficients in the table are the first derivatives of $Y$ with respect to $X$, and when the function contains logarithms of the variables, the derivatives assume natural (base $e$), not common, logarithms. In the curvilinear functions, it is clear that the slope changes with the level of $X$, and in some cases, the slope depends on two parameters rather than one. In this context, *simple* is defined in terms of the number of parameters and whether or not the data require data transformations.

11. The problem of functional form and estimation procedures for consumption functions is more complex for data sets involving individual households. In particular, zero observations on the dependent variable (consumption) potentially lead to statistical models consistent with the limit on the range of the data. This topic is outside the scope of this book (e.g., Johnston and DiNardo 1997: Chapter 13). Another issue is that, for demand systems, it is essential to have the same functional form for each equation in the system.

**Table 14-2.** Some alternative functional forms used in price analysis

| Name | Equation | Slope ($dY/dX$) |
|---|---|---|
| Linear | $Y = \alpha + \beta X$ | $\beta$ |
| Quadratic | $Y = \alpha + \beta X + \gamma X^2$ | $\beta + 2\gamma X$ |
| Hyperbola | $Y = \alpha - \beta/X$ | $\beta/X^2$ |
| Semi-log | $Y = \alpha + \beta \ln X$ | $\beta/X$ |
| Double-log | In $Y = \alpha + \beta \ln X$ | $\beta(Y/X)$ |
| Log-inverse | In $Y = \alpha - \beta/X$ | $\beta(Y/X^2)$ |
| Transcendental | In $Y = \alpha + \beta \ln X + \gamma X$ | $Y(\beta/X + \gamma)$ |
| Log-log-inverse | In $Y = \alpha + \beta \ln X - \gamma/X$ | $Y/X^2(\beta X + \gamma)$ |

*Note:* Derivatives, *dY/dX*, assume natural logarithms (base *e*).

A linear function is the simplest. The semi-log and the hyperbola require only a simple transformation of $X$.

Because of its simplicity and often because of the lack of a specific reason for assuming a curvilinear relationship, the straight-line function is the most frequently used form. It restricts the slope parameter to be a constant over the range of the data; for each unit change in $X$, $Y$ changes by the amount of the slope parameter. The elasticity estimates, however, vary along the function, as $X$ changes.

The second most common functional form is the double-logarithm specification. This model is curvilinear in the original observations.

$$Y_t = \gamma X_t^{\beta}.$$

By taking logarithms, we make the model linear in the logarithmic variables.

$$\ln Y_t = \ln\gamma + \beta \ln X_t = \alpha + \beta \ln X_t, \quad \text{where } \alpha = \ln\gamma.$$

This specification assumes a constant percentage relationship between $Y$ and $X$, but the slope is not a constant. That is,

$$dY/dX = \beta(Y/X), \quad \text{which also gives } \beta = (dY/dX)(X/Y).$$

Since the percentage relationship is equivalent to the elasticity concept, the double-logarithm specification gives a direct estimate of an elasticity, which is a constant over the range of the data.

In the context of a system of equations, the logarithmic functional form is sometimes used for simulations of the effects, say, on prices of a change in an exogenous variable, say, production or advertising intensity. This leads to the topic of equilibrium displacement models, which are typically specified in terms of elasticities and changes in the logarithms of the variables. A simple example is introduced in the final chapter.

The functional form will make a difference in simulation and forecasting applications of models. The magnitude of the difference in results among functions may be especially large when the analysis is being conducted for observations at the extremes of their ranges. In contrast, all the functional forms are likely to give similar, although not identical, results for observations near their respective means. Thus, a simulation based on a double-log form may be realistic for small percentage changes in variables but less so for large changes, especially when they are outside the range of values in the time series.

The standard deviation (error) of forecast error, used to compute confidence intervals for forecasts, can also help demonstrate the effect of variations in the explanatory variables. This standard deviation depends on the level of the explanatory variables relative to their means; forecasts for variables far from the sample mean will have large standard deviations of forecast error (Johnston and DiNardo 1997: section 1.7). And this standard error will vary with the functional form. Both point forecasts and their confidence intervals can vary in important ways from one functional form to another when the forecast is made at the data extremes.

## First Differences

We start by defining and interpreting the first-difference transformation, and then we address the reasons for considering the use of first differences. The transformation can best by understood by starting with a model in terms of the original data:

$$(6) \qquad Y_t = \alpha + \beta X_t + \gamma t + e_t.$$

The dependent variable, $Y$, depends linearly on the variable $X$ and on the time trend $t$ as well as an error term $e$. The model is, of course, assumed to hold for the entire sample period $t = 1, 2, \ldots, T$. Thus, one is justified in writing the same equation with variables lagged one time period.

$$(7) \qquad Y_{t-1} = \alpha + \beta X_{t-1} + \gamma(t-1) + e_{t-1}.$$

The first-difference transformation involves taking the differences between the adjacent observations in a time series. Equation (7) is subtracted from Equation (6) to give

$$(8) \qquad (Y_t - Y_{t-1}) = \beta(X_t - X_{t-1}) + \gamma(1) + (e_t - e_{t-1}) = \gamma + \beta(X_t - X_{t-1}) + v_t.$$

In practice, the data are differenced, and Equation (8) is estimated. As the equations demonstrate, the estimated slope coefficient receives the same interpretation in Equation (8) as in (6).[12] However, the intercept coefficient, $\gamma$, in first differences is a measure of the trend, if any, in $Y$. Equation (8) states that, even if the change in $X$ is zero, $Y$ is changing by $\gamma$ amount each time period. So, the estimated intercept in Equation (8) receives the same interpretation as the coefficient of the trend variable in Equation (6). Note that the error term in Equation (8) depends on the properties of the $e_t$'s in the original model.

One of the historical justifications for using first differences has little validity. Differencing was thought to help address the problem of large correlations among explanatory variables (multicollinearity), which result in imprecise estimates of coefficients. It is true that, if variables are trending, the differenced variables have smaller correlations with each other than do the variables in their original levels. However, another transformation is available (which we do not discuss) that makes the correlations among the regressors zero (*orthogonalizes* the variables). So, if the objective is simply to eliminate the correlations among the explanatory variables, then orthogonal regression can be used.

Data transformations do not solve the collinearity problem; they merely transform it. The intuitive explanation of why this is so is that, while transformations reduce or eliminate the intercorrelations among regressors, they simultaneously reduce their variances. A precise estimate of the effect of an explanatory variable depends on it varying. If it does not vary through time, then its effect cannot be measured. The fundamental problem of high intercorrelations is that the original data contain insufficient independent variability. It is analogous to having too little information, like too small a sample. This problem is not solved by data transformations.

The principal justification for using first differences relates to possible trends in the data, especially *stochastic* trends. An example of a stochastic trend is the random walk model, say, $Y_t = Y_{t-1} + u_t$ and $X_t = X_{t-1} + v_t$. In these equations, it can be shown that, although $u_t$ and $v_t$ have constant, finite variances, $Y_t$ and $X_t$ have infinite variances (Hamilton 1994). This results in problems of statistical inference for a linear regression model that makes $Y_t$ a function of $X_t$. In particular, when no relationship exists between $Y$ and $X$, there is a considerable probability that the analyst will conclude that a statistically significant relation exists. This is the spurious regression problem because the standard $t$ and $F$ test statistics for hypotheses, discussed in Chapter 15, are not valid (for a deeper, more precise discussion, see Hamilton 1994).

A researcher would, of course, like to discriminate between a possibly spurious relation and an actual relationship. One strategy is to use first differences. We can rewrite the random walk equation as $X_t - X_{t-1} = v_t$, and if we assume the variance of $v_t$ is a finite constant, then the same is true of the differenced

12. The estimated slope coefficients for the two equations will, however, differ. One observation is lost in taking differences, and technically Equation (8) is a different model than Equation (6). They are two different specifications relative to estimating the same slope parameter.

variable. Thus, using the first-difference model, Equation (8), is appropriate if the variables have stochastic trends.

Although economic variables may trend, it is not possible to generalize about the importance of trends in price analysis models. Indeed, we have argued in earlier chapters that little theoretical justification exists for nominal cash prices for commodities to follow a random walk.[13] A more plausible model is that cash prices are autocorrelated (with the autocorrelation coefficient not equal to 1). This implies a stable, dampening cycle in the time series, and such variables have finite variances.

Variables with stochastic trends, instead of having no relationship, may have a stable long-run relationship. This is related to the concept of co-integrated variables and the associated error correction model (e.g., Johnston and DiNardo 1997: section 8.3). These concepts are used in applications in agricultural commodity markets, but if commodity prices do not have stochastic trends, their importance has perhaps been exaggerated. In any case, these topics require more discussion than is feasible in this book.

Finally, it is possible, of course, to compute and use second (third, etc.) differences to obtain variables with a constant variance. Most price analysis models use either the levels of the variables or the first differences of the variables or, for some models, the logarithms of the variables or the first differences of the logarithms.

## Concluding Remarks

A major problem in building econometric models is that many alternative specifications are more or less consistent with the research problem and its underlying theory. Either a linear or curvilinear functional form may be justified; if variables should be deflated, the choice of deflator influences the results; lagging of the variables may be required, but the precise lag structure may be uncertain; empirical definitions (proxies) of unobservable concepts may need to be constructed; and so on, almost ad infinitum. Thus, we stress that a precise problem definition is critically important and that decisions about model specification should be made in a manner consistent with achieving the research objective.

The researcher must have expertise about the commodity sector, its data, its unique features, and the existing literature, but since the specification of the best model is nearly impossible without some prior analysis of a fixed data file, the researcher also should have a research philosophy. Pretesting should be done in a logical, rather than ad hoc, fashion. Finally, the tentative final model should

13. It is possible to test for a stochastic trend. These are called unit root tests, but these tests have relatively low statistical power. Thus, although some tests conducted on commodity prices have found unit roots, the results are far from conclusive. Another consideration is that deflated variables may have a stochastic trend because the deflator is trending. If the deflated variable is logically justified in the model, then differencing may be appropriate.

be subjected to an extensive battery of tests. Some of these are discussed in the next chapter.

## Appendix: Modeling the Variance of Prices

This chapter has emphasized models of price levels (or price changes) observed with the passage of time. As discussed in Chapter 12, the variance of the price of an individual commodity futures contract is also likely to have systematic behavior. Likewise, the parameters of the probability distributions of cash prices of commodities behave systematically with the passage of time (Peterson and Tomek 2005). It is possible to model this systematic behavior. The most common specifications involve time-series models, but one can also use structural models.

Streeter and Tomek (1992) specified and estimated a structural model of the monthly variances of soybean futures contracts. The variance of the November contract, for example, was made a function of economic factors (supply and demand measures), market structure factors (e.g., level of speculative trade), and variables to represent time to maturity, seasonality, and the lagged variance. Their analysis suggested that about 70 percent of the variation in the variance of monthly soybean prices is explainable in terms of observable variables.

It is more common to model variances solely in terms of time-series specifications. This is true, in part, because this approach allows one to model high-frequency (say, daily) observations. The simplest model makes the current variance a function of one or more lagged variances. The relevant literature can be found under the headings of autoregressive conditional heteroscedasticity (ARCH) and its extensions (e.g., Johnston and DiNardo 1997: section 6.9; Hamilton 1994: chapter 21). Time-series models of the variances of commodity prices also logically should include variables to accommodate seasonality and, for futures prices, time-to-maturity effects.

A simple example of an ARCH-type model is as follows. If a futures market is known to be efficient in the sense that its price series has a unit root, then $F_{t+1} = F_t + e_{t+1}$, or $e_{t+1} = F_{t+1} - F_t$. But the conditional variance of $e_{t+1}$ is probably not a constant, and one way to describe the non-constant variance is via an AR($p$) process of the squares of $e$. For an AR(1) process,

$$e^2_{t+1} = \alpha_0 + \alpha_1 e^2_t + v_{t+1}.$$

The $v$ error term is assumed to have the properties of a zero mean, constant variance, and zero covariances. If the AR(1) process is appropriate with a positive coefficient (logically between 0 and 1), the implication is that the variance follows a systematic pattern through time.

For most commodity price series, the AR(1) process is probably an oversimplified specification. An AR($p$) process will involve $p$ lags of the squared errors, and as noted previously, the variance of the changes in futures prices may very well have seasonal and time-to-maturity effects that should be modeled. In other words, the variance of the futures price changes is thought to increase as the maturity date approaches and to vary systematically with the season of the year. Moreover, the adjacent variances are probably autocorrelated.

In the foregoing example, the *e*'s are merely the first differences of the futures prices. More commonly, they would be the fitted residuals of a statistical model. If these residuals evidence heteroscedastic behavior, this casts doubt on the validity of the tests of hypotheses about parameters in the model (discussed in the next chapter).

The foregoing discussion barely introduces the considerable literature on the topic. The purpose of this appendix is mostly to make readers realize that it is possible to model the behavior of variances of probability distributions. Such models can be useful for decision making. For example, one of the factors influencing the prices of options-on-futures contracts is the volatility of the price of the underlying futures. Simple options pricing models assume that the price of the underlying asset has a log-normal distribution with a constant variance. If this is not so—and for commodities, it is not—then options could be priced more precisely with a model that better measured the behavior of the variance over the life of the contract.

## References

Bessler, D. A., J. H. Dorfman, M. T. Holt, and J. T. LaFrance. 2010. "Econometric Developments in Agricultural and Resource Economics: The First 100 Years," *Am. J. Ag. Econ.* 92:571–589.

Chavas, J.-P. 1999. "On the Economic Rationality of Market Participants: The Case of Expectations in the U.S. Pork Market," *J. of Ag. & Res. Econ.* 24(1):19–37.

Clyde, W. C., and C. L. Osler. 1997. "Charting: Chaos Theory in Disguise?" *J. Fut. Mkts.* 17:489–514.

Davis, L. G. 1997. "The Logic of Testing Structural Change in Meat Demand: A Methodological Analysis and Approach," *Am. J. Ag. Econ.* 79:1186–1192.

Deaton, A., and J. Muellbauer. 1980. "An Almost Ideal Demand System," *Am. Econ. Rev.* 70:312–326.

Gardner, B. L. 1992. "How the Data We Make Can Unmake Us: Annals of Factology," *Am. J. Ag. Econ.* 74:1066–1075.

Granger, C. W. J., and P. Newbold. 1986. *Forecasting Economic Time Series*, 2nd ed. Orlando, FL: Academic Press. Chapter 9.

Geweke, J., and W. McCausland. 2001. "Bayesian Specification Analysis in Econometrics," *Am. J. Ag. Econ.* 83:1181–1186.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press.

Harvey, A. C. 1981. *Time Series Models*. Oxford: Philip Allan Publishers.

Hendry, D. F. 1995. *Dynamic Econometrics*. New York: Oxford Univ. Press.

Johnston, J., and J. DiNardo. 1997. *Econometric Methods*, 4th ed. New York: McGraw-Hill Co.

Lin, W., P. C. Westcott, R. Skinner, S. Sanford, and D. G. De La Torre Ugarte. 2000. *Supply Response under the 1996 Farm Act and Implications for the U.S. Field Crops Sector*. USDA Tech. Bul. 1888.

McGuirk, A.M., P. Driscoll, and J. Alwang. 1993. "Misspecification Testing: A Comprehensive Approach," *Am. J. Ag. Econ.* 75:1044–1055.

Peterson, H. H., and W. G. Tomek. 2005. "How Much of Commodity Price Behavior Can a Rational Expectations Storage Model Explain?" *Ag. Econ.* 33:289–303.

Nerlove, M. 1956. "Estimates of the Elasticities of Supply of Selected Agricultural Commodities," *J. Farm Econ.* 38:496–509.

Savit, R. 1988. "When Random Is Not Random: An Introduction to Chaos in Market Prices," *J. Fut. Mkts.* 8:271–289.

Sims, C. A. 1980. "Macroeconomics and Reality," *Econometrica* 48:1–48.

Spanos, A., and A. McGuirk. 2001. "The Model Specification Problem from a Probabilistic Reduction Perspective," *Am. J. Ag. Econ.* 83:1168–1176.

Streeter, D. H., and W. G. Tomek. 1992. "Variability in Soybean Futures Prices: An Integrated Framework," *J. Fut. Mkts.* 12:705–728.

Tomek, W. G. 1993. "Confirmation and Replication in Empirical Econometrics: A Step toward Improved Scholarship," *Am. J. Ag. Econ.* 75(Oct., 75th anniversary issue):6–14.

Tomek, W. G., and H. M. Kaiser. 1999. "On Improving Econometric Analyses of Generic Advertising Impacts," *Agribusiness* 15:485–500.

Tronstad, R., and T. J. McNeill. 1989. "Asymmetric Price Risk: An Econometric Analysis of Aggregate Sow Farrowings," *Am. J. Ag. Econ.* 71:630–637.

Wallace, T. D. 1977. "Pretesting Estimation in Regression: A Review," *Am. J. Ag. Econ.* 59:431–443.

Warren, G. F., and F. A. Pearson. 1928. *Interrelationships of Supply and Price*. Cornell Univ. Ag. Exp. Sta. Bul. 466.

Working, E. 1927. "What Do Statistical 'Demand Curves' Show?" *Quart. J. Econ.* 41:212–235.

CHAPTER **15**

# Using and Evaluating Results

This chapter provides a basis for using and evaluating the results of price analyses based on regression techniques. As noted in Chapter 14, an important component of evaluating results is how well the results fulfill the research objectives. The discussion in this chapter must necessarily be rather general, however, because we cannot cover specific research applications in detail. But we do provide general insights into the appraisal of empirical price analyses.

The first section discusses the interpretation of regression results, which includes an implicit evaluation of results. A second section explicitly covers major issues related to the adequacy of models. Emphasis is placed on possible specification errors as well as errors in measuring variables. Subsequent sections discuss computing and interpreting elasticity coefficients and forecasts.

## Interpreting Estimated Parameters

### Net Regression Coefficients

In the previous chapter, we wrote a linear regression model with the unknown intercept and slope parameters of the model represented by β's. These parameters are often estimated by the method of ordinary least squares (OLS), although other estimators exist. The estimated equation may be written

$$Y_t = b_0 + b_1 X_{t1} + b_2 X_{t2} + \cdots + b_K X_{tK} + u_t,$$

where the $b_k$'s are the estimated coefficients and the $u_t$'s are the associated residuals ($t = 1, 2, \ldots, T$). The residuals are estimates of the unknown errors in the model, the $e_t$'s.

The individual $b_1, \ldots, b_K$ are estimates of the relationship between the respective $X_{tk}$ and $Y_t$, net of the effect of the other variables in the model. Each is an estimate of the change in the dependent variable in response to a one-unit change in the particular $X_{tk}$, the other independent variables held constant. This interpretation holds regardless of how the variables are measured. Interpretation is easier if the variables are in comparable units, e.g., quantities in pounds and prices in cents per pound. If the variables are in logarithms, then $b_1$ (say) is an estimate of the change in the logarithm of $Y_t$ given a one-unit change in the logarithm of $X_{t1}$; i.e., it is an estimate of the percentage change. Since the magnitudes of the coefficients vary with the units of measure (say, pounds versus tons or kilograms), the absolute magnitudes of the estimated coefficients are not measures of the relative importance of the variables.[1] The coefficients can be made larger or smaller by the choice of the units of measure.

As noted in Chapter 14 (in Table 14-2), the slope associated with a particular $X_{tk}$ will depend on the functional form. Thus, the coefficients discussed above pertain to the unit change of an individual variable, but in a non-linear function, the slope is defined by the appropriate first derivative. For example, in a quadratic relationship, the slope depends on the coefficients of $X_t$ and $X_t^2$ (Table 14-2).

The estimated intercept coefficient, $b_0$, determines the level of the equation. Technically, it is the estimate of the level of the dependent variable when all the independent variables are zero. Typically, the intercept coefficient has no economic interpretation because it is usually not meaningful to have all of the $X_{tk}$'s equal zero. The regression model is applicable, at best, within the range of the sample observations, and this range usually does not include zero. On occasion, however, the range of the data may include zero. This occurs, for example, when the model is fitted in terms of the first differences of the original data. In this case, as noted in Chapter 14, the intercept is an estimate of the change in $Y_t$ when the changes in the $X_{tk}$'s are zero.

## Short versus Long Run

It is sometimes necessary to make a distinction between the short and the long run (Chapters 2 and 4). For example, the quantity supplied does not adjust instantly to changes in price because of the costs of adjustment. One way to treat lags in empirical analysis is to specify that the current level of $Y$ is a function of the current and past levels of $X$. (For now, we concentrate on the relationship of $Y$ to one explanatory variable while assuming the effects of other variables are constant.)

A problem in empirical applications is the specification of the *length* of the lag. A distributed lag specification may be written $Y_t = f(X_t, X_{t-1}, \ldots, X_{t-n})$, and the problem is to define $n$ correctly. A practical problem arises if $n$ is large. For a fixed number of observations, $T$, the addition of each lag comes at the cost of two degrees of freedom; each lagged variable adds a parameter to be estimated

1. Three measures of the relative importance of independent variables in regression models are the partial-$r^2$ coefficients, the $t$-ratios, and the standardized regression coefficients (sometimes called beta coefficients, but not to be confused with the unknown parameters defined as $\beta$). We discuss $t$-ratios in a subsequent sub-section.

and results in the loss of one row of data. This problem has led analysts to consider restrictions on the *form* of the lag. In other words, rather than allowing all the parameters of the lagged $X$'s to be different, restrictions are placed on them, and the restrictions reduce the number of parameters to be estimated. The specification of models to capture distributed lag effects, therefore, requires the analyst to address the issues of the length and form of the lag.

The *geometric form distributed lag model* was popularized by Nerlove (1956, 1958) and is still used today. This model assumes that the effect of $X$ is largest in the initial time $t$ and that the effect decays geometrically with the passage of time. Implicit in this specification is that the current level of $Y$ depends on all past levels of $X$. We give an example in a later section.

The literature on distributed lag specifications has expanded enormously (for one typology of dynamic equations, see Hendry 1995: Chapter 7), and space limitations do not permit a discussion of the numerous alternatives. But it is possible to illustrate the distinction between estimates of short- and long-run coefficients using the family of *autoregressive distributed lag* (ADL) models. For a one-period lag in $Y$ and in $X$, an ADL(1,1) model, we can write

$$Y_t = \pi_0 + \pi_1 X_t + \pi_2 X_{t-1} + \pi_3 Y_{t-1} + v_t.$$

Obvious generalizations involve additional lags in $X$ and/or $Y$, but this equation seems to be adequate for many applications. Many model specifications are special cases of this equation. In particular, the geometric form distributed lag model is the case when $\pi_2 = 0$. The first-difference transformation is the special case of $\pi_3 = 1$ and $\pi_1 = -\pi_2$. In other words, some special cases involve restrictions on the ADL(1,1) model, but the ADL(1,1) model is itself a special case in the sense that lag parameters of length two or longer are zero. With additional algebra, the ADL(1,1) model can also be rewritten in error-correction form (for details see Hendry 1995: Chapter 7).

In the ADL(1,1) model, the short-run effect of X on Y is given by the parameter $\pi_1$. As noted in the previous subsection, this coefficient is a measure of the change in $Y_t$ given a unit change in $X_t$, which in this case, is the contemporaneous effect of $X$. In long-run equilibrium, the levels of $Y$ and $X$ will be unchanged at their long-run levels, e.g., $X_{t-1} = X_t = X^*$. Thus,

$$Y^* = \pi_0 + \pi_1 X^* + \pi_2 X^* + \pi_3 Y^*.$$

Solving for $Y^*$, we have

$$Y^* = \pi_0/(1-\pi_3) + [(\pi_1 + \pi_2)/(1 - \pi_3)]X^*.$$

If this specification is a correct representation of the long-run effect, then the total or long-run effect of a unit change in $X$ on $Y$ is $(\pi_1 + \pi_2)/(1 - \pi_3)$.

The foregoing idea can be generalized to other specifications. For example, if the model also contained $X_{t-2}$, then the numerator of the long-run effect would be the sum of the three parameters associated with the three $X$ variables, and the denominator would be 1 minus the parameter of the one-period-lagged $Y$. The geometric form distributed lag model makes $\pi_2 = 0$, and the long-run effect in this model is $\pi_1/(1 - \pi_3)$.

### Residuals

The residuals, $u_t$, are estimates of the unobservable errors, $e_t$. The residuals obtained using the OLS estimator have three properties. First, the sum of the squared residuals, $\sum u_t^2$, is the minimum for any estimator. Other methods of estimation exist, but OLS is the method that, by definition, minimizes the sum of the squared residuals.

Second, the sum of the residuals is zero. $\sum u_t = 0$. This is consistent with the assumption that the mean of $e_t$ is zero. It also implies that the OLS-fitted equation passes through the point of means of the variables in the equation. This is important to know because the analyst can use this information to compute an elasticity coefficient at the point of means of the data, thus assuring that the elasticity has been computed for a point on the regression plane.

Third, the correlation between each of the explanatory variables, $X_{tk}$, and the residuals, $u_t$, is zero. In this sense, the explanatory variables are "independent" of—not correlated with—the residuals. The OLS estimator forces this to be true, even if it is not true for the data. In other words, for OLS to be a desirable estimator, the right-hand-side variables should be at least weakly exogenous, which is to say, the contemporaneous (time $t$) observations for each regressor should have a zero covariance with the time $t$ error term.[2]

### Standard Errors

Courses in introductory statistics discuss the concept of the variance and standard deviation of a single random variable. These statistics are based on the variability of the individual observations about their mean. For the variance, the individual deviations from the mean are computed, squared, summed, and divided by degrees of freedom.

The variance of regression is simply an extension of this concept. Recall that the linear regression model assumes that the variance of the error term is a constant. This parameter represents the variability of the $Y_t$ around the true regression line or plane. The estimate of this unknown variance is computed from the residuals, using the assumption that the variance is a constant. Namely, the estimate is

2. The term *independent variable* does not mean that the right-hand-side variables are independent of each other. In general, the explanatory variables in econometric models have some correlation with each other; they are intercorrelated. Rather, the classical linear regression model assumes that the explanatory variables are uncorrelated with the true error term.

$$s_u^{\,2} = \sum u_t^{\,2}/(T - K - 1),$$

where $T$ is the number of observations, $K + 1$ is the number of coefficients ($b_k$'s) in the equation, and hence there are $T - K - 1$ degrees of freedom.

The standard error of estimate is the square root of the variance of the regression. The standard error is a measure of the variability in the same units of measure as the dependent variable. If $Y_t$ is in cents per pound, then $s_u$ is also in cents per pound. Naturally, other things being equal, the analyst would like to have a relatively small standard error of estimate, but it is possible for a particular economic sector to have large random variations that are not susceptible to systematic explanation.

As indicated, the regression coefficients, $b_k$, are estimates of the respective unknown $\beta_k$. Thus, even if the model is correct, the estimate of each parameter is subject to sampling error. The $b_k$ are random variables, and under the traditional assumptions of the linear regression model, including the assumption that the $e_t$ are normally distributed, the $b_k$ have a multivariate normal distribution. Thus, a single $b_k$ has a normal probability distribution with a mean and a variance. The variance can be estimated, and standard errors of the regression coefficients, $s_b$, can be computed and interpreted by analogy with the standard deviation of a variable, like $Y_t$.

The analyst would like to have a small standard error of each regression coefficient relative to the size of the coefficient. As emphasized in a later section on appraising results, the quality of the estimated standard errors depends on the correctness of the model. If, for example, the assumptions underlying the use of the OLS estimator are not met, the estimated standard errors and the associated statistical tests may be seriously wrong.

## Coefficients of Determination

The coefficient of multiple determination, $R^2$, is a measure of the degree of linear association between the dependent variable and the collective set of independent variables. Strictly speaking, the coefficient is applicable only for the OLS estimator, since it forces the correlation between the explanatory variables and the error term to be zero. Thus, the OLS estimator divides the total variability of $Y_t$ unambiguously between the variability of the $X_{tk}$'s and $u_t$.

$$R^2 = [\text{variability in } Y \text{ associated with } X\text{'s}]/[\text{total variability in } Y].$$

$R^2$ is a ratio with a range from 0 to 1. For example, if $R^2 = 0.75$, then 75 percent of the variability in $Y_t$ is estimated as being associated with the variability of the $X_{tk}$.

Adding independent variables to the model increases $R^2$ at least slightly even if no "true" relationship exists between the added variable and the dependent variable (Greene 2012: 42). Also, as the number of independent variables increases, with the number of observations held constant, the degrees of freedom decrease. For instance, two observations exactly determine the regression line for a simple two-variable equation, the degrees of freedom are 0, and $r^2 = 1$. Intuitively, an $R^2$ near 1 with few degrees of freedom suggests a misleading over-estimate of the actual degree of association. This idea has led to the use of a coefficient of determination adjusted for degrees of freedom.

$$\text{Adjusted-}R^2 = 1-(1-R^2)[(T-1)/(T-K-1)] = 1-s_u^2/s_y^2,$$

where $T$ is the number of observations, $K$ is the number of independent variables, $s_u^2$ is the variance of regression (see above), and $s_y^2$ is the variance of $Y_t$.

The interpretation of the adjusted-$R^2$ is the same as the interpretation of $R^2$. It is mathematically possible, however, for the adjusted-$R^2$ to be negative, i.e., to range from $-\infty$ to 1. A negative coefficient should be interpreted as a zero level of association.

Price analysts have used $R^2$ coefficients as indicators of the quality of the empirical model, but correctly specified models can have small $R^2$'s. Also, as the formula for the adjusted-$R^2$ makes clear, its magnitude depends on the variance of $Y_t$, which is computed from deviations around $Y_t$'s (constant) mean. If $Y_t$ is trending—the mean is systematically changing—then $R^2$ is a potentially misleading statistic. As McGuirk and Driscoll (1995) point out, the traditional $R^2$ statistic can vastly exaggerate the true degree of association.

## *t*-Tests and Tests of Logic

Since the results of regression analysis are often appraised via statistical tests, it is important to be able to interpret them. The traditional approach assumes that the linear regression model, with normally distributed errors, is appropriate.

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \cdots + e_t.$$

A common null hypothesis is that no relationship exists between a regressor and the dependent variable, say, $\beta_1 = 0$.

Two kinds of alternative hypotheses are used. One is that $\beta_1$ is not zero. This requires what is called a two-tail test. If a non-zero relationship exists, no assumption is made about the sign of the relationship. If, however, the researcher has a strong logical reason to believe that the true parameter must be negative (or positive), then a one-tail test is appropriate. If the true parameter must be negative, if it is not zero, then the alternative hypothesis is $\beta_1 < 0$. This, for example, is the appropriate alternative hypothesis for the price-quantity coefficient in a demand relation.

Given (say) the OLS estimates of the coefficients and their standard errors, it is possible to compute the statistic

$$t_1 = (b_1 - \beta_1)/s_{b1},$$

and given the null hypothesis that the true parameter is zero, the statistic becomes $t_1 = b_1/s_{b1}$. If the assumptions of the linear regression model with normally distributed errors are true, the test statistic has a *t* distribution. This probability distribution is symmetric about its mean of zero, but its variance changes as the degrees of freedom change. The *t* distribution approaches the normal distribution, with mean equals 0 and variance (standard deviation) equals 1, as the number of observations (degrees of freedom) gets larger. These statistical concepts provide the basis for obtaining the probability that the computed test statistic could have come from a given sample (fixed degrees of freedom) from a population in which the true parameter is zero.

The mechanics of the test thus involve computing the *t* statistic and then comparing it with the critical value of *t* available in tables of the *t* distribution. The critical value will depend on the degrees of freedom, which is the number of observations minus the number of $b_k$ coefficients ($T - K - 1$ in our notation) and on the *level of significance* selected by the researcher, which is the probability of type I error that the researcher is willing to accept. Type I error occurs when a true null hypothesis is rejected. Commonly used levels of significance are 0.01, 0.05, and 0.10, and the particular level is often arbitrarily selected by the analyst. The researcher must decide whether or not the result is "statistically significant."

If the absolute value of the computed *t* statistic is larger than the critical value, then the null hypothesis is rejected. A sample with $T - K - 1$ degrees of freedom from a population with the true parameter equal to zero might give the computed *t*, but the probability of this happening is sufficiently small that the alternative hypothesis is accepted. If the computed *t* is less than the table value, then the null hypothesis cannot be rejected. Outputs of computer programs usually provide an exact level of type I error associated with the particular coefficient and its standard error, and in this case, the analyst must decide whether a level like 0.06 is going to be treated as significant or not.

The difference between a one-tail and two-tail test relates to whether one or two tails of the *t* distribution are used to reach a conclusion. For a two-tail test at the 5 percent level of significance, both tails are considered, with 2.5 percent of the area in each tail. For a one-tail test, the entire 5 percent of the distribution is in one tail.

A one-tail test is illustrated for the following equation. The estimates are based on 18 observations, and since the equation contains 4 coefficients, there are 14 degrees of freedom. The standard error of each slope coefficient is given in parentheses below the coefficient.

$$\begin{aligned} Y_t = 184.3 - &\underset{(0.232)}{1.057} X_{t1} - \underset{(0.164)}{0.384} X_{t2} + \underset{(0.010)}{0.013} X_{t3} + \mathrm{u}_t \end{aligned}$$

Suppose the focus of the research is on the parameter of $X_{t1}$ and we are certain that, if its parameter is not zero, then it must be negative. A one-tail test is then appropriate. The computed $t_1$ in this example, for the hypothesis $\beta_1 = 0$, is

$$t_1 = -1.057/0.232 = -4.556.$$

For a one-tail test at the 1 percent level of significance with 14 degrees of freedom, the table value of $t = |2.624|$. Clearly the probability is tiny that the sample statistic, –4.556, could have been generated by a population in which the true parameter is zero. The researcher concludes that the parameter is not zero, rejecting the null hypothesis. Note, this is not equivalent to saying that the estimate, −1.057, equals the true parameter value.

If the coefficient had been positive, but the research logic required that it be negative, the null hypothesis would not have been rejected for large positive $t$ values. This is an implication of the one-tail test. Of course, if logic requires a negative sign, but the result is a positive sign, then this is symptom of a problem in the model. One should not be using statistical tests mechanically and concluding that results are "significant" when they are in fact illogical.

Historically, $t$-ratios have been used as a criterion for adding or dropping variables from a model. One practice was to drop variables that had coefficients less than twice the size of their standard errors (ratios less than 2 in absolute value). This rule of thumb arose from the observation that the critical (table) value of $t$ is approximately 2 for the 5 percent level of significance in two tails with 20 degrees of freedom. A less stringent rule is to drop variables with $t$-ratios of less than 1. This is equivalent to dropping variables that reduce the $R^2$ coefficient corrected for degrees of freedom. As noted previously, $R^2$ increases as variables are added to the model, and $R^2 = 1$ when the degrees of freedom equal 0. In contrast, the adjusted-$R^2$ increases only if the added variable has a coefficient with a $t$-ratio large than 1 in absolute value (Greene 2012: 43).

The foregoing discussion of the probability of type I error has been based on the assumption that the researcher started with the correct model; obtained the relevant, error-free data; fit the model; and conducted a single $t$-test. This never happens in economic applications. The model may be mis-specified, the data may contain errors, and the researcher has probably searched for a better model by using a sequence of statistical tests. All these factors cast doubt on the interpretation of the level of significance of a $t$-test.

A model in which relevant variables are omitted can still have large $t$-ratios for the coefficients of the included variables. But such results are misleading; they are typically seriously biased (e.g., Spanos and McGuirk 2001). Likewise, if variables are measured with error, the results are biased. Although the $t$-ratio is large, the true parameter probably is not within a (say) 95 percent confidence interval computed from the $t$ statistic.

Moreover, if the researcher has searched for the preferred model via a series of statistical tests (pretested), the level of significance has been eroded. Since pretesting is almost universal in applied econometrics, the true level of significance for tests conducted on the final results is unknown (Wallace 1977). For example, a test conducted at the nominal 5 percent level of significance may have an actual level of 40 percent. Intuitively, if the researcher searches for a preferred model over many alternative specifications, using a fixed data file, he or she will ultimately find something that fits the particular data set well, but the result has a high probability of being spurious.

A second issue in interpretation is the routine testing of the hypotheses $\beta_k = 0$ ($k = 1, 2, \ldots, K$) seriatim. The $t$-test was designed for a single simple hypothesis, not for a sequence of tests.[3] For example, if 10 tests were done for 10 different parameters at the 10 percent level of significance, then a large probability (0.65) exists that one of the hypotheses will be rejected, even though all 10 are true. This is just a manifestation of the concept of type I error and the laws of probability.

Of course, a researcher wants regression coefficients that are large relative to their standard errors, i.e., have large $t$-ratios. But given the model experimentation that lies behind final results and other potential problems, we believe that $t$-ratios should be interpreted qualitatively rather than in terms of formal probability statements. Basically, the larger the $t$-ratio, the larger is the statistical importance of the variable in the particular model. This conclusion, however, is far different than statements about the correctness of the model or about the precise levels of significance and confidence intervals.

The researcher should ask, why am I doing the test? All too often, tests of null hypotheses are unrelated to the research problem, and the common practice of using small levels of type I error in routine $t$-tests has a corresponding consequence of making the type II error large, *ceteris paribus*. Type II error is the case of concluding that a variable has no effect when in fact it does have a nonzero effect. Variables are presumably included in the model precisely because the analyst believes that their parameters are *not* zero. If the researcher has reason to believe that a variable is important, then it should not be discarded lightly. Yet this can be the consequence of pretesting methods conducted at tiny levels of significance.

A test of a hypothesis for an individual parameter should follow from the research focus. If the focus is, say, on the parameter of advertising expenditure (i.e., Does advertising significantly influence the level of demand?), then a $t$-test is appropriate.

3. A test for a null hypothesis about a set of parameters should be based on the $F$ statistic. Regression programs routinely provide the $F$ value for the null hypothesis that all of the slope parameters in the model are collectively zero. This is usually not an interesting hypothesis because the researcher has built the model on the premise that the variables have non-zero parameters. However, the $F$ test is important for hypotheses about sub-sets of parameters.

## Model Adequacy

The interpretation of results, as discussed in the prior section, has an implied evaluation component. Indeed, models were often judged to be "good" if the coefficients had logical signs with large *t*-ratios and if the $R^2$ was large. Modern practice requires a more comprehensive battery of tests. The tests relate to the assumptions about the error term (zero autocorrelation, constant variance, and normal distribution) and to the assumptions about the explanatory variables (no omitted variables, correct functional form, and explanatory variables are exogenous). The details of these tests are covered in a variety of sources (e.g., McGuirk, Driscoll, and Alwang 1993; Greene 2012). Here, we provide a sense of the issues and their importance.

### Specification Error

The two major categories of problems are *specification error* and *errors in variables* (measurement errors). Specification error, in its broadest sense, is any error in model specification. Errors may occur in the assumptions and treatment of the error term or in the explanatory variables. In practice, the term usually refers to the omission of relevant variables or to the inclusion of irrelevant variables, which also relates to the functional form.[4]

A relevant variable may be omitted from a model because the analyst is unaware of its importance or is unable to obtain observations on it. The omission of a relevant variable is also possibly a deliberate response to the problem of multicollinearity. Multicollinearity, as mentioned in Chapter 14, is the problem that the correlations among explanatory variables are so large that the least squares estimator cannot disentangle the separate effects of the variables. One way to deal with the problem is to drop one of the intercorrelated variables, but this creates the problem of omitting a relevant variable. In this case, the analyst is confronted with a trade-off between two problems (which is further discussed below).

If a relevant variable is omitted and if it is related to the explanatory variables included in the equation, which is common, the estimates of the parameters of the included variables are biased. The sign and the magnitude of the bias depend on the sign and magnitude of the (unknown) parameter of the excluded variable and on the relationship of the excluded variable to the included variables (Greene 2012: 56–58).

For example, in a demand equation, if income is excluded, if it has a positive parameter, and if it is positively related to the other regressor, then its omission results in a positive bias. This is the case in the following equation in which price is a function of quantity:

4. For example, if a linear function is specified but a quadratic function is correct, then the quadratic term is an omitted variables. Or, vice versa, if a quadratic term is included but the relationship is linear, the model contains an irrelevant variable.

$$P_t = 104.2 - 0.318Q_t, r^2 = 0.84.$$
$$(0.043)$$

If income, $Y_t$, is then added to the model,

$$P_t = 102.5 - 0.814Q_t + 0.024Y_t, \quad R^2 = 0.95.$$
$$(0.115) \quad (0.005)$$

Excluding income results in the coefficient of $Q_t$ changing from –0.814 to –0.318, i.e., a positive bias.

The reader should also note that the coefficient –0.318 has a small standard error, and therefore a large *t*-ratio, but clearly this result is misleading. Of course, in practice, the omission of a relevant variable is not as obvious as in this example. It is precisely when the error is unknown that the concern exists.

One of the potential symptoms of an erroneous functional form and/or omitted variables is that the estimated residuals of the fitted equation are autocorrelated (in the context of time-series observations). The systematic effect of the omitted variable, while biasing the coefficients of the included variables, also influences the residuals; a part of the effect is in the error term. Thus, if a statistical test finds the existence of autocorrelation in the residuals, this could very well be evidence of model mis-specification rather than evidence that the true errors are autocorrelated. Similarly, tests that find heteroscedasticity (non-constant variance of the error term) may be evidence of model mis-specification. In other words, when the null hypothesis of zero autocorrelations (or of homoscedasticity) is rejected, the analyst must be careful about the conclusion reached about the alternative. Although the true errors might be autocorrelated or heteroscedastic, it is perhaps more likely that the model is mis-specified.

Instability in the estimates of parameters, as the sample varies, is another sign of possible specification error. Of course, if a logical reason exists for one or more parameters to change within the sample period, this change should be modeled. Otherwise, the standard statistical model assumes that the parameters—the $\beta$'s—are constants over the sample period. With modern computer software, it is relatively easy to check whether the estimated coefficients vary significantly as the sample is altered. Tomek and Kaiser (1999) provided an example of how the coefficients of an advertising variable in a fluid milk demand equation changed for a given model specification as data points were added. In a revised model, the estimated effect of advertising on the demand for milk was relatively constant over the sample period.

It can be difficult, however, to determine the cause of instability in the estimated coefficients; the specification error may not be obvious. If the analyst has a logical reason to believe that a structural change has occurred within the sample period, the change should be modeled. It is also possible to search for possible points of structural change, but finding a change empirically without a good reason for expecting one is unsatisfying (Tomek and Mount 2009). In this situation, an alternative, improved specification may not be clear.

The inclusion of an irrelevant variable in an equation does not cause bias. This is true because an irrelevant variable, by definition, has a true parameter of zero. Since there is no bias, the inclusion of an irrelevant variable may appear to be less serious than the exclusion of a relevant variable. But this is not necessarily true. If the included irrelevant variable is correlated with one or more of the other explanatory variables, the variances of the regression coefficients are inflated unnecessarily.

The analyst would like to have a model that includes all the relevant variables but excludes irrelevant variables. This is easy to say but difficult to do. With a modern battery of statistical tests, it is possible determine if a model is statistically adequate or not. But, if the fitted model "fails" one or more tests, it is often difficult, as previously noted, to decide on how to improve the model. Ultimately, the expertise of the analyst is important in selecting the final model, and as noted before, such decisions are made in the context of the research objectives. For example, in the Tomek and Kaiser (1999) paper, the focus was on the effect of advertising expenditures on milk demand. It was important that this parameter be stable if the model was to be used for decisions about funding the advertising of milk.

Analysts sometimes exclude a variable from a model when multicollinearity exists even though the omitted variable is logically relevant in the model. The omission, of course, restricts the parameter of the omitted variable to zero, and since this is likely to be an erroneous restriction, it biases the estimates of the remaining parameters. In this situation, the analyst is making the judgment that the gain from reducing the variances of the coefficients of the retained variables exceeds the loss resulting from an increase in bias. Bias and variance can be combined in a concept called the mean of the squared errors (MSE), and in principle, the decision could be made by selecting the alternative model that minimizes the MSE. But the true MSE is unknown, and the analyst must make a qualitative judgment.

More generally, any solution to the multicollinearity problem requires additional information. This information can come in the form of restrictions on the parameters, only one of which is to treat a parameter as zero. The potential problem is that the restrictions are wrong, and as noted in Chapter 14, data transformations that reduce the correlation among regressors do not solve the problem. Alternatively, the analyst may try to increase the sample size to obtain a sample in which the variables have greater independent variability. But this solution too may be impractical or costly. For all of these reasons, multicollinearity is an exceedingly difficult problem to deal with in empirical analyses.

Since so many alternatives can be considered, both in terms of model specifications and sample periods, it is important to understand and report the sensitivity of the empirical results to the alternatives considered. Potential users of research are unlikely to trust the results if only the "best" are reported without any information about the range of alternatives considered and the corresponding sensitivity of results. Thus, unless analysts are precise about the data, models, and computational methods, their results will be treated skeptically. If the empirical price analysis results are to be used for important decisions, it is essential that the results be replicated before being used.

### Errors in Variables

The variables used in a model may be measured with error. Mistakes can be made by the agency compiling and publishing the data and by the analyst in entering and transforming the data, but the most important source of measurement error is perhaps that the observed variable does not correctly represent the economic concept relevant to the model. In specifying a consumption function from household surveys, for example, the analyst may use the income reported by the household, but households may be making consumption decisions based on what they perceive to be their permanent income, which is not reported. Or, in a supply analysis, the empirical measure of farmers' expectations may not correspond with their true expectations.

If one or more explanatory variables are measured with error, the least squares estimates of the parameters in the equation are biased (Greene 2012: 97–99). The measurement-error problem can be viewed as a variant of the omitted-variable problem. For example, suppose the true model is

$$Y_t = \beta_0 + \beta_1 X_t^* + e_t,$$

where $X_t = X_t^* + v_t$ is the observed variable. In other words, the analyst observes the true variable plus or minus errors. Many possible assumptions can be made about the nature of the errors. The simplest set of assumptions is that $v_t$ is normally distributed with mean zero, constant variance, and zero covariances. It can also be assumed that the $v_t$ and the $e_t$ are uncorrelated.

Substituting for $X_t^*$ in the regression equation, we have

$$Y_t = \beta_0 + \beta_1(X_t - v_t) + e_t = \beta_0 + \beta_1 X_t - \beta_1 v_t + e_t,$$

and hence, fitting the model $Y_t = \beta_0 + \beta_1 X_t + w_t$, where $w_t = e_t - \beta_1 v_t$, is equivalent to omitting the random variable $v_t$ from the model. Of course, $v_t$ is unknown, but by definition it is related to $X_t$. Thus, analogous to the omitted-variable problem, measurement error results in a biased estimate of the focus parameter, which in this example is $\beta_1$.

Obviously, it is incumbent on an analyst to minimize measurement errors. In some cases, the problem is that observations on an underlying concept are not available, so proxy variables are substituted for the unmeasurable concept. The use of a proxy for the unobservable concept assumes that, from a statistical viewpoint, it is better to include a variable to represent the concept, even though it is measured with error, than to omit the concept from the model. It is not possible to say unequivocally that it is better to use a proxy variable than to ignore the concept, but the evidence generally favors the use of proxy variables (Greene 2012: 87f).

Intuitively, the better the proxy variable mimics the behavior of the unobservable concept, the better the quality of the statistical results. Alternative proxy variables can have large effects on the estimated coefficients. In supply equations, it is rather common to use a trend variable to capture the possible omitted effects shifting the level of supply. This was done in a supply model for cotton in the classic paper by Nerlove (1956) and later refitted by Tomek (1972). The model is based on the assumption of a geometric form distributed lag in the supply response, and the fitted equation, with standard errors in parentheses, is

$$\begin{aligned} A_t = {} & \underset{(5.40)}{6.319} + \underset{(0.15)}{0.566}P_{t-1} + \underset{(0.15)}{0.565}A_{t-1} + \underset{(0.10)}{0.271}t, \quad \text{adjusted-}R^2 = 0.71, \end{aligned}$$

where $A_t$ is the cotton acreage planted in year $t$ in the United States, in millions of acres; $P_{t-1}$ is the season average price of cotton, deflated by the Index of Prices Paid by Farmers, in cents per pound; and $t$ is the time trend (= 1, 2, . . . ) in the sample period, crop years 1910–1933.

The results seem logical, and the slope coefficients are large relative to their standard errors. The coefficient of the trend variable suggests that the cotton acreage planted is growing 271,000 acres per year, net of the other effects of the other regressors. The short-run effect of price on acres planted is 0.566, and the long-run effect is 0.566/(1 − 0.565) = 1.301. The initial impact of a 1-cent per pound increase in the real price of cotton is an increase in planted acres of 566,000. The long-run effect of the 1-cent change is, however, an increase of 1.3 million acres, *ceteris paribus*.

An inspection of a scatter diagram between lagged prices and planted acres, however, suggests a one-time shift in the level of supply from the 1910–24 period to the 1925–33 period (Figure 15-1). This type of shift can be accommodated by using a zero-one variable.[5] The variable $D_t$ is defined to equal 0 in the years 1910–24 and to equal 1 in the years 1925–33, and $D_t$ is substituted for $t$ in the equation:

5. Zero-one variables are also called *dummy variables* or *binary variables*. They are especially useful in modeling concepts involving categories. If, for example, the concept to be measured is the educational level of individuals in a sample, the individuals might be categorized by levels of education, such as high (secondary) school graduate and college graduate. The categories are mutually exclusive, and if (say) four categories of education are appropriate, they are completely defined by three dummy variables. Issues related to the use and interpretation of dummy variables are discussed in standard econometrics textbooks (e.g., Greene 2012: 149–158).

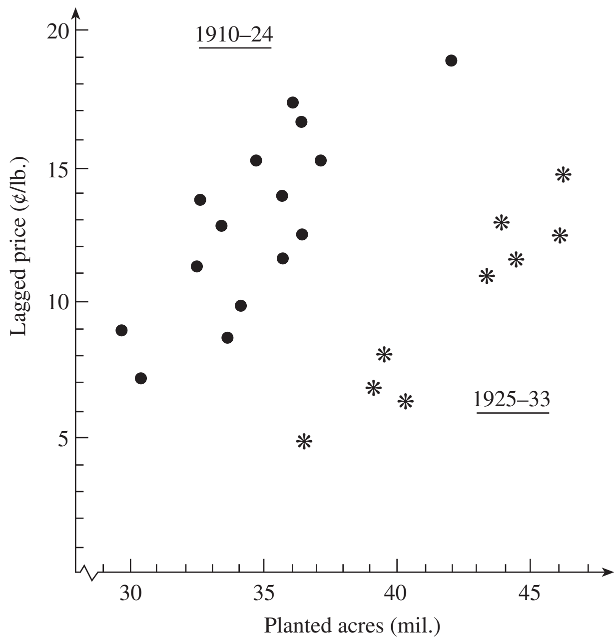



**Figure 15-1.** Relationship between cotton acreage and lagged deflated price, United States, 1910–33

$$A_t = 22.417 + 0.749P_{t-1} + 0.068A_{t-1} + 9.352D_t, \quad \text{adjusted-}R^2 = 0.89.$$
$$(4.31) \quad (0.10) \quad (0.13) \quad (1.32)$$

The coefficient of lagged price is more than 30 percent larger in the second specification, while the coefficient of lagged acreage declines from 0.565 to 0.068. The latter coefficient is not statistically different from zero. Thus, no long-run effect exists in the second specification. Clearly, the alternative proxy variables in the foregoing model cause rather large changes in the coefficients of the other variables in the model.

The zero-one variable is capturing a shift in the intercept (level) of the equation. In 1910–24, $D_t = 0$, and the intercept coefficient is 22.417. In 1925–33, $D_t = 1$, and the intercept is 22.417 + 9.352 = 31.769. The $D$ variable, like the $t$ variable, does not tell why supply changed, but in the example, the $D$ variable appears to be a better proxy for the unknown effect. That is, the supply shift appears to have been a one-time event, not a gradual change over the sample period. The one-time event resulted in a 9.352 million increase in planted acres, *ceteris paribus*. Such a large change is attributed, in part, to the control of the boll weevil, a pest in cotton production (Walsh, as quoted in Tomek 1972).

An obvious conclusion of this section is that analysts should strive to minimize measurement errors. The definition of proxy variables should not involve the unthinking use of trend or code variables. If possible, the proxy should be based on a fundamental knowledge of the underlying concept and how it might best be measured.

Analysts should also have a sense of humility about the quality of their results. Computer output can look extremely precise but be precisely wrong. In reporting results, the analyst should make the limitations of the analysis clear. Also, the coefficients can be no more accurate than the original data, and analysts should not convey a false sense of accuracy by reporting a large number of digits for the regression coefficients.

## Computing and Appraising Elasticities and Flexibilities

Since elasticities play an important role in economics, analysts frequently compute them from the estimates of structural equations. Elasticities (or flexibilities) are given directly by the coefficients of equations that are linear in the logarithms of the variables, but for most other functional forms, they must be calculated.

A linear demand function illustrates how elasticities are obtained. Assume the analyst has fitted

$$Q_t = b_0 - b_1 P_t + \text{other relevant variables.}$$

The coefficient $b_1$ is an estimate of the relationship between own-price ($P$) and quantity ($Q$), net of the effects of the other explanatory variables. The price elasticity of demand is defined (Chapter 3) as

$$E = (dQ/dP)(P/Q),$$

holding other factors constant. Since $b_1$ can be viewed as an estimate of the change in $Q$ given a one-unit change in $P$, the ratio $dQ/dP$ equals $b_1$, where $b_1$ is negative in a demand equation. Thus, the elasticity is computed as

$$E = b_1(P/Q).$$

The ratio $P/Q$ must represent a point on the estimated demand function; actual observations of $P$ and $Q$ cannot be used because such observations typically are not on the regression line. The usual practice is to use the arithmetic means of $P$ and $Q$, since the least squares fit is known to go through this point and it is a point in the "middle" of the data. With a demand function that is linear in $P$ and $Q$, the elasticity varies as the ratio $P/Q$ changes. To compute an elasticity at any point on a linear function, the relevant values of the independent variables are inserted into the equation and the corresponding value of $Q$ is computed. The selected level of $P$ used to compute $Q$ and the computed value of Q are then used in the ratio to obtain the elasticity.

The foregoing discussion also implies that the elasticity will change if the level of the demand function shifts. In other words, if income, $Y_t$, increases, the level of demand presumably increases, and, *ceteris paribus*, $Q_t$ will be larger for any given price level, $P_t$. This makes comparing elasticities for different points in time difficult. With the passage of time, the factors influencing demand will have changed, and even if the slope coefficient relating price to quantity has not changed, the elasticity will be computed for $P$'s and $Q$'s on different demand functions. That is, the ratio $P/Q$ can change not only because of movements along a linear function but because the function itself has shifted with the passage of time. Thus, comparisons of elasticities at different points in time require a precise statement about the conditions of the comparison.

As previously noted, inverse demand equations are frequently estimated for models of agricultural products,

$$P_t = a_0 - a_1 Q_t + \text{other conditioning variables.}$$

In this case, it is natural to estimate flexibility coefficients. The own-price flexibility is

$$F = (dP/dQ)(Q/P),$$

holding other variables constant. Thus, the estimated flexibility is computed as $F = a_1(Q/P)$, and as before, the mean values of $P$ and $Q$ are commonly used to obtain the ratio.

The price elasticity of demand is sometimes approximated by using the reciprocal of the flexibility coefficient. As mentioned in Chapter 3, however, the reciprocal of $a_1$ obtained from a price-dependent equation is not equal to the statistical estimate, $b_1$, from a quantity-dependent equation. The reciprocal of the flexibility sets the lower limit of the elasticity (in absolute value).

Estimates of cross-price elasticities, income elasticities, and so forth can be computed from linear equations using procedures analogous to those discussed above. For instance, an acreage response equation for cotton was presented in the previous section. The price elasticity of supply (acreage response) can be computed as

$$E_s = 0.749(11.87/37.32) = 0.24,$$

where 0.749 is the slope coefficient, the mean of real price is 11.87 cents per pound, and the mean of acreage is 37.32 million acres. Formulas for computing elasticities for other functional forms are provided in Table 14-2.

Elasticity estimates should not be accepted uncritically. Some type of statistical analysis usually is the basis for the computed elasticity, and the discussions in Chapter 14 and in this chapter highlight the difficulties of obtaining high-quality empirical results. Thus, at best, the estimated coefficient is an approximation of the theoretical concept; nevertheless, these approximations can be useful. (We illustrate a use in Chapter 16.) As discussed in Chapter 3, however, elasticity estimates will be influenced by the degree of aggregation (T-bone steak versus all beef) and by the market (say, retail versus farm) demand. Thus, it is important to assure that the elasticity estimate is appropriate for the problem being addressed.

In addition, as the reader may recall, different samples from the same population will not give exactly the same estimates of the unknown parameters. Sampling error arises from the random variability not accounted for by the observed variables. Even if the model is correctly specified and the variables are measured without error, the random errors will vary from sample to sample. In economics, we may have just one set of time-series observations, but the analyst must remember that even under "perfect" conditions, as time passes and new observations are obtained (a different sample), some changes in the estimated coefficients should be expected. One does expect, however, that the changes are a result of sampling error, not specification error. Ultimately, the analyst must make a judgment about whether the quality of the available elasticity estimates is sufficient to give a relatively precise answer to the research question.

## Forecasting from Regression Equations

### Mechanics

The computations for making a point forecast of the dependent variable in a regression equation are simple. The analyst has estimated an appropriate equation from the sample period observations, $t = 1, 2, \ldots, T$. Then, to make a forecast, the analyst must obtain estimates of the relevant values of the explanatory variables for the forecast period, defined as time $T + 1$. These estimates are inserted into the estimated equation, and the dependent variable is computed. Letting $T + 1 = *$, the forecast is computed as

$$\hat{Y}_* = b_0 + b_1 X_{*1} + b_2 X_{*2} + \cdots + b_K X_{*K},$$

where the $b_k$'s are estimated from the sample data, and the $X_{*k}$'s are estimates for the forecast period. This procedure assumes, of course, that the model's parameters do not change from the sample period to the forecast period.

Given the estimated equation, the major problem is obtaining the ancillary forecasts of the $X_{tk}$'s for the forecast period. The explicit statistical model makes $Y$ a function of a set of $X_k$'s; the $X_k$'s are themselves not modeled. If the explanatory variables are truly predetermined, then their values are determined by outside forces or by prior events, but this does not necessarily make it easy to obtain the appropriate $X_{*k}$'s to make a forecast. The general point is that the forecast is *conditional* on the values of the right-hand-side variables. If the conditioning information changes, the forecast will change. A correct model gives a correct conditional estimate of the dependent variable, but if the conditioning information is wrong, then the forecast will be in error.[6]

The ancillary forecasts of the right-hand-side variables are obtained in a variety of ways. Sometimes the explanatory variables enter the model in lagged form. Current production is a function of lagged prices. Thus, it is possible to observe, say, the planting-time price (and previous prices, if appropriate) at the time the forecast is being made. Or, in time-series models, the current value of the variable to be forecast is a function of lagged values of the same variable (and perhaps other variables). Again, the analyst has observations on the current and past values of the variable to forecast future values.

Short-run forecasts can also take advantage of known, fixed biological relationships in the production process, particularly for animals and animal products. Government agencies, like the USDA, provide data on the size of the breeding herd or flock, the distribution and size of animals by weight and age categories, etc. Given estimates of growth rates, the current information can be used to estimate future output. Thus, the hatch of broiler-type chicks can be used to forecast the output of broiler (chicken) meat 2 or 3 months later. Forecasts of broiler output somewhat further into the future might be made by using the number of eggs placed in incubators and their hatching rate, which determine the number of chicks hatched.

Some national governments provide data on inventories, area planted, etc. Indeed, in the United States, the USDA provides monthly estimates of expected production for the major crops in the months prior to harvest. Thus, it is possible to use such data in a model to forecast prices.

Sometimes, the ancillary forecasts are based on formal models of past behavior. It may be possible, for example, to develop a relatively simple time-series model of a variable like personal disposable income and to use the model to make short-term projections. Also, economists for banks, government agencies, and consulting firms often make forecasts of economic variables that can be used for other forecasts. Likewise, demographers provide estimates of growth rates in the population.

In sum, the price analyst can obtain ancillary forecasts of explanatory variables by taking advantage of the possible lag structure in the model, from

6. As discussed in Chapters 12 and 13, price quotations for contracts for future delivery can be viewed as conditional forecasts. In futures markets, the conditioning information changes frequently; hence, futures-contract prices fluctuate more or less continuously. Thus, in an efficient market, the current price reflects known information, but because new information enters the market, prices change. If a formal econometric model is to be used for forecasting and decision making, then forecasts should be updated as new information becomes available.

secondary sources of data or forecasts or from one's own models of the variables. It needs to be recognized, however, that these ancillary forecasts are subject to error and that the forecasts of the dependent variable are conditional on the estimated values of the explanatory variables.

## Interpreting Forecasts

The computed dependent variable ($\hat{Y}_*$ in the forecast equation) is an estimate. A point of confusion is that this estimate can be given two interpretations: as an estimate of a conditional mean of a probability distribution of forecasts or as an estimate of an individual $Y_*$ from the distribution, $* = T + 1$.

We have already emphasized that the computed value of $Y$ is conditional on the values of the $X_k$'s used. A single estimate is computed for a particular future period, but conceptually, a conditional distribution of $Y_*$'s exists for a given set of $X_{*k}$'s. That is, if the conditions of period $T + 1 = *$ were replicated many times, the observed $Y_*$'s would differ because the errors, $e_*$, would differ for each replication. Thus, under the usual assumptions, the different individual $Y_*$'s are normally distributed around a mean and have a given variance. The mean is conditional on the particular levels of the $X_{*k}$. In this context, the forecast can be interpreted as an estimate of the mean for the given levels of the regressors. This interpretation treats the errors as having a zero mean, which is an assumption of the model.

The more common interpretation is, however, that the forecast is an estimate of the particular (individual) $Y_*$ that will happen in time $T + 1$. This individual value will differ from the mean by a random error. The actual outcome in the economy presumably depends on the observable explanatory variables but also on an unknown random component. Under this interpretation, the forecast is subject to two types of errors. Like the interpretation of the forecast as an estimate of a conditional mean, the forecast is subject to sampling error. The forecast is computed from estimated regression coefficients, not the true parameters. But, in addition, since the forecast is for an individual $Y_*$, the error term will not be zero. These two sources of error are taken into account in the derivation of the variance of forecast error (or its square root).[7]

These sources of forecast error always exist, but they do not exhaust the reasons for mistaken forecasts. In other words, two sources of error exist even though the model and the conditioning information are correct. But the conditioning information may be wrong and/or the model may be mis-specified. We have already noted that a forecast of $Y_*$ requires ancillary estimates of the right-hand-side variables. Obviously, these ancillary forecasts may be mistaken. At time $T$, when the forecast is made, the analyst attempts to use the best available information, but by the time the forecast period, $T + 1$, arrives, this information

7. A distinction must be made be made between the variance that pertains to a forecast as an estimate of the condition mean and the variance that pertains to an estimate interpreted as a forecast of an individual outcome.

may turn out to be wrong. A different variance of forecast error has been developed to take account of this third source of error (Feldstein 1971), but the confidence intervals obtained from this variance are likely to be very large and hence not useful.

The consequences of possible errors in model specification are even more difficult to discuss precisely in a non-technical way. If the true errors of the model are autocorrelated (or heteroscedastic), then more precise forecasts (forecasts with smaller variances) can be obtained by taking account of the autocorrelation in the modeling, estimation, and forecasting procedures. The mistake "merely" involves not taking account of the autocorrelation, and a more general model specification, which allows the error term to be autocorrelated, will increase the precision of the forecast.

But, as noted earlier, the estimated residuals—not the true errors—may be autocorrelated because of specification error. The specification error results in unnecessarily large variances of forecasting error. Specifically, if the model has omitted a relevant variable, the forecast may be unbiased, *conditional on* the information that was used in the model, but because relevant information has been omitted, the variance of forecast error will be larger than it would have been for the correct model. The larger variance implies that forecasting errors of a particular magnitude have a higher probability of occurring. The fact that the distribution of forecasts is centered on the true mean (are unbiased) is not very useful if the distribution has a large variance. In this case, the preferred approach is to improve the model specification by including the omitted variable, but it is likely to be difficult to determine what has been omitted.

A possibly erroneous functional form can cause a similar problem. Many different functional forms can fit the middle of the data reasonably well, but when forecasts are made for observations at the data extremes, the functional form makes a large difference in the forecast.

The problems discussed in the previous paragraphs may not be obvious from an evaluation of the estimated equation based on the sample data; problems may become obvious only after a forecast has been made. In the next section, we outline some methods of evaluating forecasts, but in all cases, it should be understood that a number computed from an estimated equation is simply a summary of the information formalized by the model specification. Analysts typically modify the estimate using informed judgment. If an apple crop, for example, is thought to have below-average storage quality, then proportionately more apples are likely to be sold at harvest rather than stored. Hence, a forecast of harvest-time price, based on a price-production equation, may be judgmentally lowered to accommodate the additional information about storage quality. The judgment can be viewed as adjusting the error term of the equation in light of information about a random event that is thought to influence price.

If the user of forecasts has more than one source of forecasts available, it may be possible to obtain more precise forecasts by using an average of the individual forecasts. This assumes that the different forecasts have some different information content. A simple or weighted average might be used. In principle, it is possible to find optimal weights (see the literature on composite forecasts, e.g., Granger and Newbold 1986).

## Evaluating Forecasts

In appraising an equation's forecasting ability, four alternative series can, in principle, be considered: (1) ex post simulations, (2) ex post forecasts, (3) unadjusted ex ante forecasts, and (4) adjusted ex ante forecasts. An ex post simulation merely compares the computed $Y_t$'s with the actual $Y_t$'s for the sample period, $t = 1, 2, \ldots, T$. In this sub-section, we use the notation that the computed $Y_t$'s are $P_t$'s, predicted values, and that the actual $Y_t$'s are $A_t$'s, actual values. Many of the tests of model adequacy, mentioned earlier, are based on the residuals of the fitted equation, namely, $u_t = A_t - P_t$, for the sample period.

Forecasts refer to comparisons beyond the sample period, say $t = T + 1, T + 2, \ldots, I$. In an ex post forecast, the analyst has withheld data from the sample (not used it for estimation). Then, the $P_t$'s and $A_t$'s are compared to known values of the explanatory variables but for the observations that were not used to estimate the equation. This approach abstracts from the problem of having erroneous conditioning information, since the "forecasts" use observable values of the right-hand-side variables. Also, this approach is related to tests of the stability of parameter estimates and allows for the immediate evaluation of the model.

Ex ante forecasts are truly predictions for a time beyond the present. Typically, ancillary forecasts must be made for the explanatory variables; they are not known with certainty, except in the case where the explanatory variables appear in the model as lagged values ($X_{t-1}$, etc.). Thus, the computed dependent variable is based on estimated coefficients and on estimated values of the regressors. The unadjusted ex ante forecast is just this computed dependent variable. Since analysts often make judgmental modifications of the computed value, it is the judgmentally modified forecast that is published. We call this type of forecast a modified ex ante forecast.

The foregoing distinctions are important because the forecasting performance probably differs in the four cases. In particular, since the criteria used to select a final model typically require that the model simulate the sample period well, it is not surprising that equations should have outstanding performance as measured by their ability to simulate the sample. At the other extreme, unmodified ex ante forecasts are often poor. Such forecasts are based on estimated coefficients and on ancillary estimates of the regressors. Moreover, the computed result is not modified by any other information. The estimates of the coefficients and/or of the regressors may be poor, and other information, if available, is ignored.

Analysts want to appraise the true ex ante forecasting ability of a model; however, the passage of time is required for evaluation. Actual $Y_t$'s must be observed before an appraisal can be conducted, and more than one comparison is needed for an appraisal ($t = T + 1$, $T + 2$, etc.). Since considerable time must pass before an evaluation can be completed, ex post simulations and ex post forecasts continue to be used in appraisals even though they tend to exaggerate an equation's forecasting performance.

Evaluation criteria fall into two categories: goodness-of-fit and tracking measures. Within each category, a variety of specific performance measures exist, and in practice, a particular measure can be computed in different ways, thus creating confusion. For example, a goodness-of-fit measure might be based on the level of the variable or on the period-to-period change. The fit usually appears much better when the level of the variable is used. We outline some of the performance measures available, indicating the nature of the alternatives and the precision that is needed in computing them.

Tracking measures are designed to test the ability of an equation to identify turning points (changes of direction) in the series at the time they occur, and tracking ability can be measured both for within-sample simulations and post-sample forecasts. The analyst naturally wants an equation that correctly predicts turning points when they occur but that does not falsely predict turning points. A rigorous discussion of turning-point errors has been available for many years (e.g., Theil 1965: Chapter 2), but many confusing counts of turning-point errors exist in the literature. We provide a review of turning-point analysis in the appendix to this chapter.

With respect to goodness-of-fit measures, $R^2$ and $s_u^{\,2}$ are the natural and appropriate measures of performance for the sample period. The $R^2$ of the sample-period multiple regression is identical to the $r^2$ in the simple relationship between $P_t$ and $A_t$, $t = 1, 2, \ldots, T$. Thus, a logical extension for appraising ex post or ex ante forecasts is to compute $r^2$ for $P_t$ and $A_t$ over the total number of comparisons, $t = T + 1, T + 2, \ldots, I$. A more informative analysis is to fit the simple regression:

$$A_t = a + bP_t + e_t.$$

For a perfect forecast, $a = 0$, $b = 1$, and $r^2 = 1$. For the sample-period simulations, the OLS estimator results in $a = 0$, $b = 1$, and $r^2 = R^2$. Thus, the foregoing regression is of interest when applied to ex post or ex ante forecasts. If $a$ is not 0 and/or $b$ is not 1, then the model is producing a systematic error. It is preferable to try to determine the source of the error, but if the problem cannot be found, an expedient is to use the simple regression to correct the forecast. In other words, having detected an error for the $T + 1$ through $I$th period, the model could be used to adjust the forecast for time $I + 1$. This would involve inserting $P_{I+1}$ into the equation and computing $A_{I+1}$.

A commonly used measure of goodness-of-fit is the MSE or its square root (RMSE). For the sample period, MSE is analogous to $s_u^{\,2}$, except that, for $s_u^{\,2}$, the sum of the squared errors is divided by the degrees of freedom rather than by the number of items in the sum. For the forecast period, the sum of squared errors is merely divided by the number of comparisons:

$$\mathrm{MSE} = (1/I)[\sum (P_t - A_t)^2],$$

where $t = T + 1, T + 2, \ldots, I$. The RMSE is in the units of measure of the variable $A$. Thus, the ratio of RSME to the mean of $A_t$ is often computed to obtain a percentage measure of the goodness-of-fit. This kind of measure permits comparisons of forecasts from different models or for different variables. Whether one is using RMSE or the percentage measure, the smaller the measure, the closer the forecast is to the actual value.

The MSE can be decomposed into components analogous to those of the regression above—basically, into a systematic and a random component. Theil (1965) also proposed a decomposition of the MSE into bias (unequal means of $P$ and $A$), unequal variability (different standard deviations), and unequal covariation. Theil's approach may help the analyst understand the reasons for the forecasting errors.

An evaluation of forecasting performance of a particular equation is not very informative unless one has a standard of comparison. An equation could have a mediocre forecasting ability in an absolute sense but still be relatively good compared with the alternative forecasts. Thus, the forecasting performance of one model should be compared with alternatives, which might be other models or perhaps judgmental forecasts.

One basis for comparison is called the naive forecast. The simplest naive forecast is $P_t = A_{t-1}$. This standard is implicit in a coefficient of inequality proposed by Theil (1965). One version of the *coefficient of inequality* is

$$[(1/I)\sum (P_i - A_i)^2]^{1/2} / [(1/I)\sum A_i^2]^{1/2}.$$

In effect, the RMSE is standardized by a component using the actual values. Theil apparently intended that $P_i$ and $A_i$ be defined as the following differences:

$$P_i = P_t - A_{t-1} \quad \text{and} \quad A_i = A_t - A_{t-1}.$$

Note, the definition of $P_i$ involves the change of the forecast from the previous actual level; also, $I$ is the number of $P_i$'s and $A_i$'s in the comparison. If the value of the coefficient of inequality falls between 0 and 1, the forecast is better than the one obtained from a naive model, but if the coefficient is larger than 1, the forecasting equation is performing worse than the naive model. The larger the coefficient of inequality, the poorer is the equation's ability to forecast, in terms of the closeness of fit of the forecast to the actual outcome.

### Economic Evaluation

Forecasts are not made in an abstract context. Rather, they are intended to help decision makers and improve social welfare. The "ultimate" evaluation of a forecast is its economic value, including the economic consequences of forecasting errors. As noted in the discussion of whether speculators can forecast the prices of futures contracts, the question is not solely one of statistical significance but, rather, one of economic significance. In principle, an econometric model may have statistically significant variables, but the cost of developing and using the model may exceed the benefits. The cost includes, of course, the risk associated with the fact that forecasts are not perfect.

More generally, the question is, do forecasts improve social welfare through improved resource allocations? The evidence about the benefits of forecasts of agricultural prices is sparse (Allen 1994; Brorsen and Irwin 1996). Large private firms have research departments or consultants that support the (internal) decisions of the firm's managers, and therefore presumably benefit the firm. Although it is unclear whether forecasts made in the public sector (by government or university economists) have had net benefits (Brorsen and Irwin 1996), it is clear that estimates of expected crop production, inventories, and uses provided by the USDA are used and influence prices. There continues to be a demand for outlook work.

## Concluding Remarks

A major theme of this chapter is that obtaining high-quality, useful empirical results is difficult. Necessary conditions for obtaining high-quality results are the precise definition of the problem and in-depth research, and this chapter provides only a glimpse of the depth of effort that is required. In some instances, the quality of data available to researchers is insufficient to answer the research question. Consumers of the output of price analysts should be skeptical of its quality. It is important to understand the limitations of the empirical results.

## Appendix: Appraising Turning-Point Errors

To be precise in discussing turning points, specific conventions (definitions) must be adopted. Different conventions can result in different conclusions. Thus, letting $P_t$ be the forecast (predicted) value and $A_t$ be the actual value, a correct forecast is defined as

$$\operatorname{sign}(P_t - A_{t-1}) = \operatorname{sign}(A_t - A_{t-1}).$$

An incorrect forecast occurs when the two signs are not equal. This convention assumes that one-step-ahead forecasts are being evaluated and uses the current predicted value *relative to the previous actual value.*

The logic of this convention is that $A_{t-1}$ is assumed to be known at the time the forecast is being made and, hence, that the user is interested in a forecast relative to the known level (and not relative to the previous forecast level). Theil introduced the convention in 1958 (see Theil 1965), but some applied work has used the sign of $P_t - P_{t-1}$ to evaluate turning points. The latter approach does measure the notion of $P$'s ability to move up and down with $A$, which may be of interest in simulations. But if a user's objective is to appraise one-step-ahead forecasting ability, then Theil's criterion seems best, and it is the criterion used here. In any case, the user must recognize that the two rules will lead to different counts of errors.

To determine whether or not a turning point has occurred, the sequence of the signs of $A_t - A_{t-1}$ is examined. If the sign changes, a turning point has taken place. For example,

| $t$ | Sign($A_T - A_{T-1}$) | Comment |
|---|---|---|
| 2 | + | cannot be evaluated |
| 3 | + | no turning point |
| 4 | − | turning point |
| 5 | + | turning point |
| 6 | + | no turning point |
| 7 | − | turning point |

Since one observation is lost in determining the sign of the change and since the initial change has no prior standard for a comparison, the number of evaluations will be two less than the total number of simulations or forecasts. That is, evaluating $T$ simulated values for the sample period, $T - 2$ comparisons can be made. One of the practical problems of evaluating forecasts is that the number of one-step-ahead forecasts available for evaluation, $I$, may be small.

Four possibilites exist when the forecast change is compared with the actual change.

1. No turning point (no change in sign) occurred in the actual value, and none was forecast, a correct decision.
2. No turning point took place, but one was forecast; this is analogous to a type I error in statistical hypothesis tests.
3. A turning point occurred, but none was forecast; this is analogous to a type II error.
4. A turning point took place, and it was correctly forecast.

This information can be summarized in a $2 \times 2$ contingency table, where the main diagonal gives the counts of the two correct cases and the off-diagonals count the two types of error.

A straightforward procedure for doing the counting is to list the actual and predicted values. Then, the signs of the changes can be listed in two columns, remembering for the forecasts to use the sign of the forecast change from the previous actual level. When the signs in the row are not the same, then an error has been made. The type of error depends on the sequence of sign changes in the actual column. For example,

| $t$ | $A_t$ | $P_t$ | $A_t - A_{t-1}$ | $P_t - A_{t-1}$ |
|---|---|---|---|---|
| 1 | 75.5 | 75.5 | * | * |
| 2 | 66.6 | 67.5 | – | – |
| 3 | 64.1 | 69.2 | – | + |
| 4 | 67.3 | 67.5 | + | + |
| 5 | 66.6 | 64.9 | – | – |

In time 3, a turning point was predicted, but none occurred. In times 4 and 5, correct forecasts were made. The actual series increased and then decreased, and this is captured by the forecasts. In time 4, the forecast value is below the time 3 forecast value but above the previous actual value. Hence, following the convention used here, the sign of the forecast change is positive and is a correct forecast of an increase. If the analyst had used a rule based on the change in forecast series, then the time 4 change would have been counted as an error.

It is possible to perform formal statistical tests of the data compiled in the contingency tables. Readers should consult a standard statistics textbook for the details.

## References

Allen, P. G. 1994. "Economic Forecasting in Agriculture," *Internat. J. Forecast.* 10:82–135.

Brorsen, B. W., and S. H. Irwin. 1996. "Improving the Relevance of Research on Price Forecasting and Marketing Strategies," *Ag. & Res. Econ. Rev.* 25:68–75.

Feldstein, M. S. 1971. "The Error of Forecast in Econometric Models When the Forecast-Period Exogenous Variables Are Stochastic," *Econometrica* 39:55–60.

Granger, C. W. J., and P. Newbold. 1986. *Forecasting Economic Time Series*, 2nd ed. Orlando, FL: Academic Press.

Greene, W. H. 2012. *Econometric Analysis*, 7th ed. Upper Saddle River, NJ: Prentice Hall.

Hendry, D. F. 1995. *Dynamic Econometrics*. Oxford: Oxford Univ. Press.

McGuirk, A. M., and P. Driscoll. 1995. "The Hot Air in $R^2$ and Consistent Measures of Explained Variation," *Am. J. Ag. Econ.* 77:319–328.

McGuirk, A. M., P. Driscoll, and J. Alwang. 1993. "Misspecification Testing: A Comprehensive Approach," *Am. J. Ag. Econ.* 75:1044–1055.

Nerlove, M. 1956. "Estimates of the Elasticities of Supply of Selected Agricultural Commodities," *J. Farm Econ.* 38:496–509.

———. 1958. *Distributed Lags and Demand Analysis for Agricultural and Other Commodities*. USDA Ag. Hdbk. 141.

Spanos, A., and A. McGuirk. 2001. "The Model Specification Problem from a Probabilistic Reduction Perspective," *Am. J. Ag. Econ.* 83:1168–1176.

Theil, H. 1965. *Economic Forecasts and Policy*, 2nd rev. ed. Amsterdam: North-Holland.

Tomek, W. G. 1972. "Distributed Lag Models of Cotton Acreage Response: A Further Result," *Am. J. Ag. Econ.* 54:108–110.

Tomek, W. G., and H. M. Kaiser. 1999. "On Improving Econometric Analyses of Generic Advertising Impacts," *Agribusiness* 15:485–500.

Tomek, W. G., and T. D. Mount. 2009. "New Methods for Structural Change Analysis of Agricultural and Food Markets: Discussion," *Am. J. Ag. Econ.* 91:1448–1449.

Wallace, T. D. 1977. "Pretest Estimation in Regression: A Survey," *Am. J. Ag. Econ.* 59:431–443.

CHAPTER 16

# Applications

This chapter illustrates four applications of price analysis that follow naturally from the material covered in this book. The examples do not, of course, cover the full range of possible applications, but we hope that they provide some additional insights that will be useful to the reader. Although this book has stressed the complexities of agricultural product markets, the first two examples show that relatively simple tools can be useful. The last two examples show how analysts have dealt with some of the complexities of agricultural markets.

## Forecasting Basis Convergence

The profitability of storing grain and hedging in futures depends, in part, on expected basis convergence over the storage period (Working 1953). The mechanics involve selling futures contracts at the time the decision to store is made; when the grain is sold out of storage, the futures position is offset by buying the equivalent contracts (Chapter 13). Thus, the decision to store and hedge can be aided by forecasting the expected basis convergence based on the relationship of the change in the basis to its initial size, i.e., the relative price change. A simple linear regression model can be the foundation for the forecast (Working 1953); namely, the change in the basis over a specified storage interval is made a linear function of the initial basis.

$$B_{t+j} - B_t = \alpha_0 + \alpha_1 B_t + e_{t+j},$$

where $B$ is the basis (futures minus cash prices), $t$ is the date of hedge/storage decision, and $t + j$ is the date for completing the hedge. The $\alpha$'s are the parameters of the model to be estimated, and $e_{t+j}$ represents an error term. Equations can be estimated for varying storage intervals (i.e., alternative definitions of $j$) and for different locations (hence, different cash prices). Obviously, the analyst should use the definitions of location and storage length that apply to the particular decision. Heifner (1966) provided a comprehensive set of estimates for the storage of corn over varying time intervals using cash prices reported in Lansing, Michigan, and prices for futures contracts traded on the Chicago Board of Trade (CBOT). The statistical fits, based on the harvest-time basis, are large for storage intervals varying from 2 to 8 months. The $r^2$ coefficient does decrease slightly as the length of the storage interval increases. This is consistent with spot and futures prices having increased variability as the storage season progresses. The data reported in Table 16-1 are used to fit a basis-change model for corn stored in Toledo, Ohio. The cash prices are those reported by the Agricultural Marketing Service of the USDA for Toledo, and the futures prices are for the May contract traded on the CBOT. The assumed storage interval is from early November (at the completion of harvest) to late April (before the expiration of the May contract). This is the type of equation that a farmer or merchant in the Toledo area could use to assist a decision about whether or not to store and hedge in May futures.

The estimated equation is

$$B_{t+j} - B_t = \underset{(2.05)}{-6.587} - \underset{(7.41)}{0.611}B_t + u_{t+j}, \quad \text{adjusted-}r^2 = 0.831,$$

where the $t$-ratios are shown in parentheses. The results suggest that, for each cent of basis level at harvest time, the basis will narrow by 0.61 cent by late April plus the magnitude of the intercept, –6.59 cents. That is, if the basis was 40 cents at harvest, the forecast convergence is approximately 31 cents: $-6.587 - 0.611(40) = 31.027$.

If the cost of storing corn is 4 cents per bushel per month and if the storage interval is 5.5 months, then the cost of storage is 22 cents per bushel (disregarding the opportunity cost). In this context, a forecast convergence greater than 22 cents suggests that storing *with* hedging is potentially profitable. As the data in Table 16-1 indicate, unhedged storage does not assure the return implied by a large initial basis. For example, in 2009–10, the initial basis in Toledo, Ohio, was 50 cents per bushel, but the spot price declined 23.5 cents over the storage interval. Without a hedge, the inventory-owner would have lost 23.5 cents on the value of the commodity plus the cost of storage. In contrast, if the inventory had been hedged, the gross return would have been 38.5 cents per bushel, the actual narrowing of the basis in that year. (The example ignores transactions costs, including possible margin calls over the life of the futures position.)

A forecasting application assumes that the model is appropriate for the sample period and continues to be appropriate for the forecast period. For this example, several institutional features of the corn market may need to be considered in analyzing the appropriateness of the model. One is that the delivery provisions of the CBOT corn futures contract were changed, effective with the contracts

**Table 16-1.** Selected cash and futures prices for corn (¢/bu.), 2000–01 to 2011–12

| Crop year | Price[a] | November[b] | April[c] | Change |
|---|---|---|---|---|
| 2000–01 | May futures | 235.00 | 196.00 | –39.00 |
| | Toledo cash | 197.00 | 180.00 | –17.00 |
| | Basis | 38.00 | 16.00 | –22.00 |
| 2001–02 | May futures | 222.25 | 197.75 | –24.50 |
| | Toledo cash | 190.00 | 191.25 | 1.25 |
| | Basis | 32.25 | 6.50 | –25.75 |
| 2002–03 | May futures | 250.25 | 233.25 | –17.00 |
| | Toledo cash | 246.50 | 243.75 | –2.75 |
| | Basis | 3.75 | –10.50 | –14.25 |
| 2003–04 | May futures | 250.50 | 308.25 | 57.75 |
| | Toledo cash | 225.25 | 302.25 | 77.00 |
| | Basis | 25.25 | 6.00 | –19.25 |
| 2004–05 | May futures | 217.50 | 208.25 | –9.25 |
| | Toledo cash | 171.25 | 198.25 | 27.00 |
| | Basis | 46.25 | 10.00 | –36.25 |
| 2005–06 | May futures | 217.75 | 231.25 | 13.50 |
| | Toledo cash | 157.75 | 218.50 | 60.75 |
| | Basis | 60.00 | 12.75 | –47.25 |
| 2006–07 | May futures | 365.25 | 364.75 | –0.50 |
| | Toledo cash | 326.25 | 355.25 | 29.00 |
| | Basis | 39.00 | 9.50 | –29.50 |
| 2007–08 | May futures | 416.25 | 591.25 | 175.00 |
| | Toledo cash | 372.50 | 574.75 | 202.25 |
| | Basis | 43.75 | 16.50 | –27.25 |
| 2008–09 | May futures | 407.50 | 375.00 | –32.50 |
| | Toledo cash | 365.00 | 371.50 | 6.50 |
| | Basis | 42.50 | 3.50 | –39.00 |
| 2009–2010 | May futures | 419.00 | 357.00 | –62.00 |
| | Toledo cash | 369.00 | 345.50 | –23.50 |
| | Basis | 50.00 | 11.50 | –38.50 |
| 2010–2011 | May futures | 606.75 | 752.25 | 145.50 |
| | Toledo cash | 566.25 | 743.75 | 177.50 |
| | Basis | 40.50 | 8.50 | –32.00 |
| 2011–12 | May futures | 661.75 | 624.00 | –37.75 |
| | Toledo cash | 653.00 | 626.00 | –27.00 |
| | Basis | 8.75 | –2.00 | –10.75 |

[a] May futures price is the settlement price, Chicago Board of Trade May contract; the Toledo cash point is the midpoint of the range for number 2 yellow corn in Toledo, Ohio (USDA 2013b).
[b] Selected days, Nov. 6–12 window.
[c] Selected days, Apr. 24–30 window.



maturing in calendar year 2000. From 1974 through 1999, corn could be delivered at approved elevators in Chicago, St. Louis, and Toledo; those short futures contracts had the option of delivering in Toledo (at a 3-cent per bushel discount from Chicago). Effective in 2000, delivery was required at specified locations

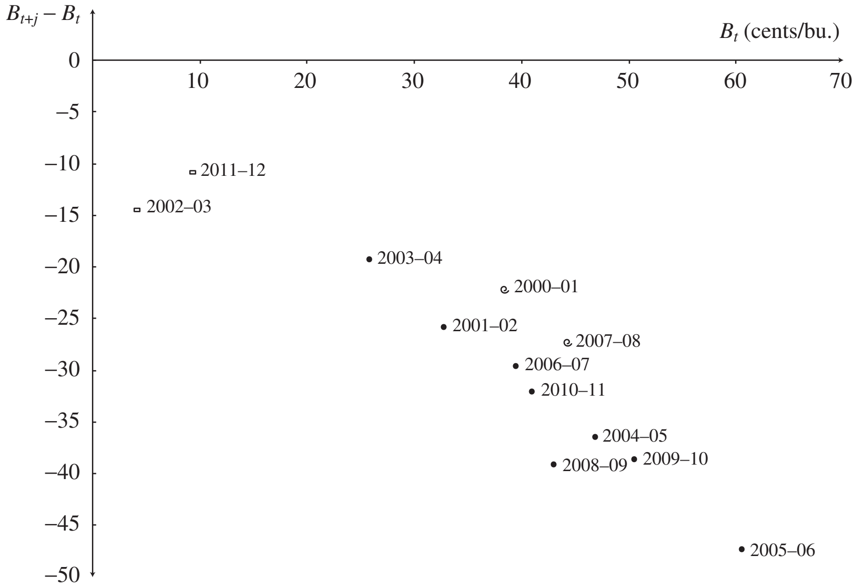



**Figure 16-1.** Relationship of initial basis to the basis change over a fixed storage interval for corn, 2000–01 through 2011–12. See Table 16-1

on the Illinois waterway; hence, our sample starts with observations for the May 2001 contract observed in November 2000. Other contract provisions, however, may have influenced convergence in the sample period (see Irwin et al. 2011).

In any case, an inspection of a plot of the observations, in Figure 16-1, suggests that some influential observations existed; two appear to have a negative effect and two a positive effect on the level of the equation. Thus, we refit the simple model adding two zero-one variables: $D1$ equals 1 in 2000–01 and 2007–08 (and zero otherwise) and $D2$ equals 1 in 2002–03 and 2011–12 (and zero otherwise). The results are as follows:

$$\begin{aligned} B_{t+j} - B_t &= \underset{(0.23)}{-0.935} + \underset{(3.68)}{7.965}D1_{t+j} - \underset{(1.66)}{6.725}D2_{t+j} - \underset{(8.11)}{0.774}B_t + u_{t+j}, \\ \text{adjusted-}R^2 &= 0.936, \end{aligned}$$

with the $t$-ratios shown in parentheses. Adding the dummy variables causes the slope coefficient of $B$ to change from −0.611 to −0.774.

The dummy variables are mainly an expedient to estimate the effects of the outliers; to use the equation forecasting, the analyst would have to decide whether either of dummy variables should be set to 1 for the forecast period. Without a logical reason for defining the dummy variables, the foundation for a decision is unclear. But it should be noted that $D2$ pertains to the 2 years in the sample when the initial basis was small (3.75 and 8.75 cents), and this suggests that $D2$ should be set to 1 when the initial basis is small. That is, convergence (implying a negative sign) is estimated to be 6.7 cents larger than expected for "normal-size" bases. The general point is that the analyst must be alert to possible reasons for changes in the model.

If both dummy variables are set equal to zero to make a forecast, then for an initial 40-cent basis, the estimated convergence is similar to the simple model presented earlier: $-0.935 - 0.774(40) = -31.895$ cents. Note, too, that the intercept coefficient is not different than zero statistically and that $-0.774(40) = 30.97$. If the initial basis was small, say 5 cents, then the forecast convergence will be $-0.935 - 6.725 - 0.774(5) = -11.52$ cents.

Given the seeming usefulness of a rather simple procedure, there are surprisingly few published results for this model in the context of assisting decision makers. Peck and Williams (1991) considered such regressions in analyzing possible delivery problems on the CBOT grain contracts in the years up to 1988–89, and Hranaiova and Tomek (2002) explored whether the simple linear model might be improved by adding a variable that explicitly measures the value of the delivery options embedded in the corn contract. Neither of these applications explicitly evaluates the forecasting model in terms of decisions to store and hedge.

An additional important point to be gained from the foregoing example is that the degree of convergence varies randomly from year to year. Although large initial bases result in large returns to hedged storage, there is nonetheless basis risk associated with the hedge. This basis risk can be measured via the use of the standard deviation of forecast error. Unlike some problems in price analysis, this forecasting application is a case where the level of the right-hand-side variable is known at the time the forecast is made. Having fit the equation to a historical sample, the analyst can observe the initial basis—in the example, the spot and May futures prices on (say) November 5—and insert this basis into the equation to forecast the magnitude of convergence from November until late April. With the standard error of forecast, it is possible to compute a confidence interval and to estimate the probable range of outcomes.[1]

1. If the merchant decides to hedge, another question is the size of the futures position relative to the size of the inventory. A large literature exists on this topic, as noted in Chapter 13. The simplest model assumes that the objective is to minimize the variance of returns, and the magnitude of a variance-minimizing futures position can be estimated from the simple linear regression that makes the change in the cash price a function of the change in the futures price over the storage interval. For the November to April interval considered in our example, the estimated slope coefficient is 1.003 with a $t$-ratio of 21.96. This implies that the hedger should take an equal—but opposite—position in futures to minimize the variance of returns from the hedge. Of course, if the forecast basis convergence is less than storage costs, the manager may decide not to store and hedge, while the estimated model assumed that a hedge was placed routinely every year (i.e., used all 12 observations).

As discussed in Chapter 9, the cash price of a commodity like corn has a systematic seasonal pattern. Thus, an alternative to forecasting basis convergence is to forecast the seasonal increase in the cash price. In principle, such a model could guide storage decisions without hedging in futures; i.e., the firm would store grain if the forecast price increase is expected to cover the costs of storage. Forecasting the seasonal price increase is more difficult, however, than forecasting basis convergence, the change in relative prices (e.g., Heifner 1966).

In the 12-year sample shown in Table 16-1, the cash price increased 42.6 cents from November to April on average, but in 6 of the 12 years, the price increase was negative or did not cover storage costs. So, if a firm in Toledo stored corn routinely every year without hedging, the average increase would appear to have more than covered costs, but this does not reflect the outcomes in individual years. In sum, firms have choices about the decisions they make and the models that support these decisions.

## Price Determination Equation

In Chapters 5 and 12, we outlined models of price determination. The discussion emphasized the potential complexity of realistic models, and Chapter 12 discussed the simultaneous determination of cash and futures prices. Notwithstanding these complexities, analysts sometimes use graphs, tables, and simple regression models of price determination to summarize information. The research objective may be to provide a forecasting tool, but more generally, the analysis helps the researcher depict current economic conditions relative to the historical evidence. We illustrate such a model for feed grains. The reader will recall that the relationship among supplies and the uses of grains is defined by an identity (e.g., see Table 14-1). The identity demand equals supply, $D_t = S_t$, can be written

$$I_t + Q_t + X_t = I_{t-1} + A_t + M_t,$$

where $I_t$ is the ending inventory, $I_{t-1}$ is the beginning inventory, $Q_t$ is the domestic consumption, $X_t$ are the exports, $A_t$ is the domestic production, and $M_t$ are the imports. The variables on the right-hand side may be treated as predetermined. This is clearly true for beginning inventories and domestic production, which depend on past decisions. It is perhaps less true for imports, but for the grains in the United States, imports are small.

In this context, we specify a price determination, or inverse (derived) demand, function as

$$P_t = D^{-1}(D_t) = D^{-1}(I_t + Q_t + X_t).$$

In words, the right-hand-side quantities involve the sum of the various uses of the given supply of grain, and in equilibrium, these quantities must equal the total supply. Clearly a more complex model would specify separate equations for the demand for inventories, the demand for domestic uses, and the demand for exports. It also might involve modeling futures prices, i.e., expected prices.

For empirical estimates of a simple one-equation model, a common practice is to specify the farm price as a function of the ratio of ending inventories to the sum of the other uses during the *t*th time period. Letting $U_t$ be the total use in time $t$ ($Q_t + X_t$ in our example) and $I_t$ be the inventory at the end of time $t$,

$$P_t = f(I_t/U_t).$$

A variety of non-linear functional forms are used. This accommodates the idea that, as current stocks approach zero (but cannot be less than zero), prices must necessarily rise sharply to ration these stocks among competing demands. (In the more complete model in Chapter 12, this non-linearity is modeled via the supply of storage equation.)

Westcott and Hoffman (1999) provided examples for wheat and corn, using a double-logarithm functional form. Here, we illustrate a relationship for feed grains in the United States using data for the marketing years 1990–91 through 2012–13 (Table 16-2). The sample starts in 1990–91 to avoid complications related to changes in government programs. The marketing year runs from September 1 through August 31. Several functional forms were considered, and we report a semi-logarithmic function.

$$\underset{(1.78)}{P_t = -2.221} - \underset{(4.26)}{2.631}\ln(I_t/U_t), \quad \text{adjusted-}r^2 = 0.437,$$

where $I_t$ is the ending (August 31) inventory of all feed grains, in millions of metric tons; $U_t$ is the sum of the uses of feed grains for the marketing year, in millions of metric tons; $P_t$ is the weighted average farm price of corn for the marketing year in the United States, in dollars per bushel, and the *t*-ratios are shown in parentheses.

The marginal effect on farm price of a small change in the ratio $I/U$ can be estimated for different points on the function using the first derivative (see Table 14-2). The ratio in the sample period varied from 0.058 in 1995–96 to 0.203 in 2004–05. Thus, letting $I/U = Z$, we compute the marginal effects for small (0.06) and large (0.20) ratios as follows:

$$dP/dZ = -2.631/0.06 = -43.85 \quad \text{and} \quad dP/dZ = -2.631/0.2 = -13.16.$$

These results demonstrate that a small change in the ratio has a large (absolute value) effect on price when the ratio itself is small but has a small effect on price when the ratio is large. In economic terms, a given demand shock influences prices much more when the stocks-to-use ratio is small.

**Table 16-2.** Quantities of feed grains and farm price of corn, United States, 1990–91 to 2012–13

| | Feed grains (mil. metric tons) | | |
|---|---|---|---|
| Marketing year[a] | Disappearance | Ending stocks | Price of corn ($/bu.)[b] |
| 1990–91 | 230 | 48 | 2.28 |
| 1991–92 | 234 | 34 | 2.37 |
| 1992–93 | 249 | 63 | 2.07 |
| 1993–94 | 226 | 27 | 2.50 |
| 1994–95 | 269 | 45 | 2.26 |
| 1995–96 | 243 | 14 | 3.24 |
| 1996–97 | 256 | 27 | 2.71 |
| 1997–98 | 252 | 38 | 2.43 |
| 1998–99 | 261 | 51 | 1.94 |
| 1999–00 | 268 | 49 | 1.82 |
| 2000–01 | 272 | 53 | 1.85 |
| 2001–02 | 272 | 45 | 1.97 |
| 2002–03 | 260 | 31 | 2.32 |
| 2003–04 | 279 | 29 | 2.42 |
| 2004–05 | 291 | 59 | 2.06 |
| 2005–06 | 304 | 55 | 2.00 |
| 2006–07 | 301 | 36 | 3.04 |
| 2007–08 | 344 | 45 | 4.20 |
| 2008–09 | 327 | 47 | 4.06 |
| 2009–10 | 350 | 48 | 3.55 |
| 2010–11 | 348 | 32 | 5.18 |
| 2011–12 | 331 | 28 | 6.22 |
| 2012–13[c] | 297 | 22 | 6.65–7.15 |

*Source*: USDA (2013a).
[a] September 1–August 31 year.
[b] Season average price, weighted by marketings.
[c] Preliminary estimates as of March 2013.



If one examines the plot of price against the ratio (Figure 16-2), it is clear that the observations starting in 2006–07 lie well above the remaining data points. An expedient way to accommodate this shift is to add a zero-one intercept variable to the model. In the following equation, $D_t$ equals 1 for 2006–07 to 2012–13 and 0 otherwise.

$$P_t = -0.679 + 1.956D_t - 1.558 \ln(I_t/U_t), \quad \text{adjusted-}R^2 = 0.792.$$
$$(0.848) \quad (6.07) \qquad (3.75)$$

The coefficient of the zero-one variable implies that, net of the effect of the ratio variable, the farm price of corn was about $1.96 per bushel higher starting in the 2006–07 marketing year. This same $D_t$ variable can be multiplied by the ratio variable, called an interaction term, to test whether the marginal effect changed between the two time periods. The following regression includes the slope as well as intercept $D_t$ variable:

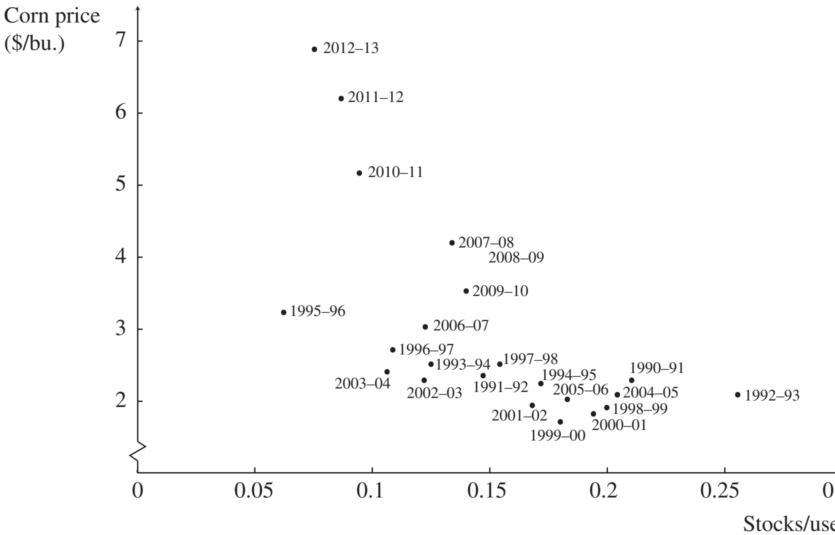



**Figure 16-2.** Relationship of price of corn to stocks-to-use ratio of feed grains, United States, 1990–01 to 2012–13

$$P_t = 0.614 - 6.631D_t - 0.874\ln(I_t/U_t) - 3.969D_t \ln(I_t/U_t),$$
$$(1.16) \quad (4.56) \quad (3.16) \qquad (5.96)$$
$$\text{adjusted-}R^2 = 0.924.$$

This equation can be interpreted as two separate equations, one for each time period:

$$P_t = 0.614 - 0.874\ \ln(I_t/U_t), \quad \text{for 1995–96 to 2005–06, and}$$
$$P_t = -6.017 - 4.843\ \ln(I_t/U_t) \quad \text{for 2006–07 to 2012–13.}$$

These results indicate that the marginal effect increased considerably over the seven-year period.

From a conceptual viewpoint, price and the right-hand-side variables (especially ending inventory) are simultaneously determined. The size of ending inventory will be influenced by the consumption demands and prices during the prior 12 months. Thus, a simultaneous equations estimator (rather than OLS) is a preferred way to estimate the slope parameter if the emphasis is on obtaining the marginal effects reported here. Although the alternative estimator would produce different numbers than those just estimated, the qualitative result would be similar, namely that the marginal effect of a change in the ratio depends on whether the ratio is small or large.

The least squares estimates of the price model could, in principle, be used to forecast price, conditional on the expected stocks-to-use ratio (Waugh 1961). The problem is, however, that, since the ratio and price are simultaneously determined, it is not logical to specify the level of *I*/*U* without knowing *P*, or vice versa. Both *U*/*I* and *P* depend on other exogenous variables. It should also be noted that forecasts at the data extremes can be influenced by the choice of the functional form. Nonetheless, analysts sometimes use the simple regression equation as a guide to expected economic conditions.

## Structural Models of the U.S. Dairy Sector

Econometric models of the U.S. dairy sector have been developed to study the impact of dairy policies, new technologies, advertising, and other issues on dairy markets (e.g., Kaiser 1997, 2010; Kaiser et al. 1994; Liu et al. 1991; Tomek and Kaiser 1999). As pointed out in Chapter 14, models should be built and evaluated relative to specific research objectives. In this section, we present an example of a model developed by Kaiser (2010) to examine the impacts of generic fluid milk promotion on U.S. milk markets.

Both dairy farmers and fluid milk processors pay an assessment on milk sales that finances a marketing program aimed at increasing milk consumption. (Dairy farmers also have promotion programs for cheese and other dairy products.) Since this program is sanctioned by Congress, there is a requirement that it be independently evaluated on an annual basis to determine whether it increases the consumption of milk and dairy products. To net out the effects of other demand determinants from generic milk advertising, the analysis used an econometric demand model, which included the following variables influencing per capita fluid milk demand: the CPI for fluid milk; the CPI for nonalcoholic beverages, which was used as a proxy for fluid milk substitutes; the percentage of the U.S. population younger than 6 years old; per capita disposable income; variables to capture seasonality in fluid milk demand; expenditures on food consumed away from home as a percentage of total food expenditures; expenditures on competing beverage advertising (bottled-water and soy beverage advertising combined), expenditures on generic fluid milk advertising, and expenditures on generic fluid milk non-advertising marketing activities. Since the goals of the farmer and processor marketing programs are the same for fluid milk, all generic fluid milk advertising by both programs were aggregated into a single advertising variable, and all generic fluid milk non-advertising marketing by both programs were aggregated into a single non-advertising marketing variable.

The model was estimated with national quarterly data from 1995 through 2009. To account for the effects of inflation, prices and income were deflated by the CPI for all items. Generic fluid milk advertising and competing advertising expenditures were deflated by a media cost index computed from annual changes in advertising costs by media type. Generic fluid milk non-advertising marketing expenditures were deflated by the CPI for all items. Because advertising has a carry-over effect on demand, all advertising-expenditure variables were modeled using a distributed-lag structure.

The amount of food that is consumed away from home, measured in this model as per capita expenditures on food eaten away from home as a percentage of per capita expenditures on all food, has an elasticity of –0.685. This means that a 1 percent increase in the food consumed away from home would result in a 0.685 percent decrease in fluid milk demand when holding all other demand factors constant. This negative relationship may be due to the limited availability of fluid milk products and high availability of fluid milk substitutes at many eating establishments, which frequently offer only one or two types of fluid milk beverages. The percentage of the population under 6 years of age is also one of the most important factors affecting fluid milk consumption. This factor has an estimated elasticity of 0.561. This result is consistent with previous studies, which show that one of the largest fluid milk-consuming segments of the population is young children.

Per capita disposable income has a positive and statistically significant impact on per capita fluid milk consumption. A 1 percent increase in real per capita income would result in a 0.13 percent increase in per capita fluid milk demand, holding all other demand factors constant. Similar to the price elasticity in magnitude, the income elasticity is consistent with the notion of fluid milk products as a staple commodity in the United States. Not surprisingly, the retail price of fluid milk has a negative and statistically significant impact on per capita demand. The results indicate that a 1 percent increase in the real retail price of fluid milk would result in a 0.126 percent decrease in per capita fluid milk quantity demanded. The magnitude of this elasticity is relatively small, which indicates that U.S. consumers' fluid milk purchasing behavior is relatively insensitive to changes in the retail price. This result, which is consistent with other studies, is probably due to the fact that fluid milk is generally regarded as a staple commodity in the United States.

Combined soy beverage and bottled water advertising also has a negative impact on fluid milk demand during the study period. The estimated fluid milk demand elasticity with respect to soy beverage and bottled-water advertising is –0.013 and is statistically significant. The generic fluid milk marketing activities conducted by fluid milk processors and dairy farmers have a positive and statistically significant impact on per capita fluid milk demand. The average advertising elasticity is computed to be 0.037 and is statistically significantly different from zero. Thus, a 1 percent increase in generic fluid milk advertising would increase per capita fluid milk consumption by 0.037 percent, holding all other demand factors constant. The generic non-advertising marketing elasticity is computed to be 0.028 and is statistically significant. Thus, the advertising elasticity is estimated to be 1.3 times higher than the non-advertising elasticity, and the two are statistically significantly different.

To examine the impact of dairy farmer and fluid milk processor marketing expenditures on the total consumption of fluid milk, the estimated demand equation was simulated for two scenarios for the period from 1999 through 2009: (1) a baseline scenario in which the combined fluid milk marketing (advertising and non-advertising) expenditures were equal to the actual marketing expenditures under the two programs, and (2) a no-national-dairy-program, no-fluid-milk-processor-program scenario in which there was no fluid milk processor–sponsored marketing and the dairy farmer–sponsored fluid milk marketing was reduced to 42 percent of actual levels to reflect the difference in assessment before the national program was enacted. A comparison of these two scenarios provided a measure of the impact of the national dairy and fluid milk programs. These marketing activities were responsible for creating an additional 6.23 billion pounds of milk consumption each year on average. Put differently, had there not been generic fluid milk marketing conducted by the two national programs, fluid milk consumption would have been 11.3 percent less than it actually was during this time period. Overall, the research suggests that dairy farmers and milk processors received a high return on their investment in advertising. But, as Kaiser (2010) pointed out, the model probably has shortcomings and additional research is required to refine his estimates.

## Equilibrium Displacement Models

Equilibrium displacement models (EDMs) have a long history of use in agricultural economics (e.g., see Davis and Espinoza 1998). These models permit an analysis of the consequences of a change in one or more exogenous variables on the endogenous variables in the economic sector under analysis. The analyses may be qualitative or empirical. In contrast to the simulation of an econometric model discussed in the prior section, EDM analysis typically looks at the consequence of a one-time change in a particular exogenous variable. An example is the classic paper by Gardner (1975) who analyzed, *inter alia*, the effect of an exogenous change in farm product supply on the retail-farm price ratio.

### Definition

To understand EDMs, we start with a simple structural model of a competitive market with a supply and a demand equation (Chapter 14).

$$A_t = \alpha_0 + \alpha_1 P_t + \alpha_2 R_t \text{ (supply)}$$

$$P_t = \beta_0 + \beta_1 Q_t + \beta_2 Y_t, \quad \beta_1 < 0 \text{ (inverse demand)}$$

$$A_t = Q_t \text{ (equilibrium condition).}$$

The model can be written in terms of the true parameters or in terms of their estimates. For conceptual applications, one thinks in terms of the true parameters and their logical signs. For empirical work, the parameters must be estimated.

As noted in Chapter 14, the structural equations can be solved for their reduced forms. In the foregoing model, $P_t$ and $Q_t$ are the current endogenous variables, and $R_t$ and $Y_t$ are the exogenous variables. We solve for the endogenous variables in terms of the exogenous variables.

$$Q_t = [(\alpha_0 + \alpha_1\beta_0)/(1-\beta_1)] + [\alpha_2/(1-\beta_1)]R_t + [\alpha_1\beta_2/(1-\beta_1)]Y_t, \text{ and}$$

$$P_t = [(\beta_0 + \beta_1\alpha_0)/(1-\alpha_1)] + [\beta_1\alpha_2/(1-\alpha_1)]R_t + [\beta_2/(1-\alpha_1)]Y_t,$$

or

$$Q_t = \pi_{10} + \pi_{11}R_t + \pi_{12}Y_t, \text{ and}$$

$$P_t = \pi_{20} + \pi_{21}R_t + \pi_{22}Y_t.$$

For example, $\pi_{21} = \beta_1\alpha_2/(1 - \alpha_1)$; each of the reduced-form coefficients depend on a set of structural coefficients. Thus, the net effect of a unit change in $R_t$ on $P_t$ or on $Q_t$ depends on the combination of structural coefficients.

An EDM is analogous to the foregoing reduced-form equations, but it specifies all the variables in percentage-change terms. For example, $R_t$ is replaced with $R^* = dR/R$, where $dR$ is a differential, which can be thought of as a small change. Or one can think of the variables as specified as logarithmic differences, as in the demand equation discussed in the previous section (Kaiser 2010). The parameters in the structural equations are elasticities. Thus, the parameters of the reduced-form equations depend on these elasticities. In sum, the analysis of the effects of a change in an exogenous variable on an endogenous variable is conducted in terms of percentage changes.

Commonly, an analyst would ask a question like, if $R$ changes 1 percent (or 10 percent), what is the percentage change in $P$? In terms of our notation, $P^* = \pi_{21}R^*$, and setting $R^* = 1$, the percentage change in $P$ can be evaluated from $\pi_{21}$. If estimates of the structural elasticities are available, these can be used to compute the reduced-form parameters.

In a qualitative analysis, the analyst may be able to deduce the sign of the reduced-form parameters from the signs of the structural parameters. The sign of $\pi_{21}$ depends on the signs of $\beta_1$, $\alpha_1$, and $\alpha_2$. The coefficient $\beta_1$ is the own-price elasticity of demand and hence is negative; the coefficient $\alpha_1$ is the own-price elasticity of supply and hence is positive; if $R^*$ is a measure of input costs, then $\alpha_2$ is negative. It follows that the numerator of $\pi_{21}$ is positive. One can further argue that the price elasticity of supply is price inelastic, meaning that $\alpha_1$ is between 0 and 1. Based on these arguments, the effect of $R^*$ on price is logically positive.

Of course, sometimes the sign of the effect cannot be determined, but this too can be useful information (as illustrated next). Without a formal analysis, the researcher may have an erroneous preconceived idea about the expected sign of a reduced-form coefficient.

## Example Equilibrium Displacement Models

Many of the topics and models discussed in this book can be illuminated by the use of EDMs. The total elasticity concept introduced in Chapter 3 can be viewed as derived from an EDM. The general point of the total elasticity concept is that the own-price elasticity in a structural equation can be misleading because of the *ceteris paribus* clause. The example given in Chapter 3 was based on a four-equation model of supply and demand for two commodities. The solution for the reduced form in a system of equations is, however, best done with matrix algebra, a topic beyond the scope of this book.

A second example from Chapter 3 involves the relationship of price elasticities to price flexibilities and can be illustrated using a two-equation model. Conventional demand equations make quantity a function of prices, and for simplicity, we assume two products. All the factors influencing demand, other than the prices, are collapsed into the intercept terms. To obtain an EDM for this case, the demand equations are written as

$$Q_1^* = \beta_{10} + \beta_{11}P_1^* + \beta_{12}P_2^*$$
$$Q_2^* = \beta_{20} + \beta_{21}P_1^* + \beta_{22}P_2^*.$$

In this model, $\beta_{11}$ and $\beta_{22}$ are the own-price elasticities of demand and $\beta_{12}$ and $\beta_{21}$ are the cross-price elasticities of demand. The own-price elasticities are negative, and the cross-price elasticities of substitutes are expected to be positive.

For agricultural products, however, it is the quantities that are treated as predetermined, and the inverse demand functions make the respective prices a function of the two quantity variables. Solving for the prices in terms of quantities gives

$$P_1^* = \pi_{10} + \pi_{11}Q_1^* + \pi_{12}Q_2^*,$$

and an analogous equation for $P_2^*$. The coefficient $\pi_{11}$ is the own-price flexibility coefficient, and it depends on the elasticity parameters in the quantity-dependent functions. We state (without deriving) that $\pi_{11} = \beta_{22}/(\beta_{11}\beta_{22} - \beta_{12}\beta_{21})$. This shows, first, that $1/\pi_{11} = \beta_{11}$ only if one or the other of the cross-price elasticities is zero. That is, if the last term in the denominator is zero, then $\pi_{11} = 1/\beta_{11}$, or $1/\pi_{11} = \beta_{11}$. Thus, in general, the reciprocal of the flexibility is not equal to the own-price elasticity.

It is logical that the cross-price elasticities are both positive and that their product is smaller than the product of the own-price elasticities. This implies that $|1/\pi_{11}| < |\beta_{11}|$. The own-price elasticity will equal or exceed, in absolute value, the reciprocal of the own-price flexibility. Although the cross-price elasticities are positive for substitutes, the cross-price flexibilities are negative. Note, $\pi_{12} = -\beta_{12}/(\beta_{11}\beta_{22} - \beta_{12}\beta_{21})$. The numerator changes signs, and as suggested above, the denominator is positive.

Expanding on the previous example, a product's price may be a function of advertising as well as quantities and other predetermined variables. If the analyst is considering the effects of advertising and if two competing goods (say, beef and pork) are both advertised, then the reduced form for (say) the price of beef (subscript 1) can be written

$$P_1^* = \pi_{10} + \pi_{11}A_1^* + \pi_{12}A_2^* + \pi_{13}V_1^* + \pi_{14}V_2^*,$$

where $P_1$ is the price of beef, $A_1$ is the production of beef, $A_2$ is the production of pork, $V_1$ is the advertising of beef, and $V_2$ is the advertising of pork. (Other variables that would logically appear in the reduced form are ignored for simplicity.)

Kinnucan (1996) and Kinnucan and Miao (2000) pointed out that, logically, one cannot determine the expected sign of $\pi_{13}$, which is the effect of the advertising of beef on the price of beef. A superficial view is that advertising should have a positive effect on the own-price, but the reduced-form coefficient is a composite of structural parameters, including the cross effects with pork. Advertising beef, if effective, will reduce the demand for pork (a cross effect), and the reduced demand for pork will reduce its price, which in turn will reduce the demand for beef. This is an example of a case in which the sign of a reduced-form parameter, $\pi_{13}$, is indeterminate.

An important message of this example is that analysts must be careful about making superficial judgments about the expected signs of parameters of reduced-form equations. If the analyst fit an equation like the beef price equation directly by least squares and obtained a negative sign for $\pi_{13}$, the analyst might modify the model (pretest) in a search for a model with the "logical" positive sign. A more complete analysis, however, using the concepts of an EDM can provide more precise insights about the interrelationships implicit in the structure of the market. Note also that, if both beef and pork are being advertised, then both advertising variables should appear in the reduced-from equation. Omitting the pork variable from the beef equation is probably a specification error.

In this section, we have limited ourselves to two equation models, but for most realistic applications, larger models are required. Gardner's (1975) six-equation model of margins is an example. One of the implications of this model, highlighted by Kinnucan and Forker (1987), is that elasticities of farm-to-retail price transmission[2] depend on whether the change in the margin arises from exogenous changes in retail demand or in farm supply.

2. The farm-to-retail price transmission elasticity is the percentage change in the retail price given a 1 percent change in the farm price. Kinnucan and Forker's (1987) analysis suggests that this elasticity is not 1.

If the analyst makes the additional assumption that the supply of marketing inputs is perfectly elastic (review Chapter 6), however, the elasticity of transmission reduces to the farmers' share of the consumers' dollar whether or not the margin change originates at the farm or retail level. This result implies that, for highly processed retail products, the elasticity of price transmission is small. This result is not surprising because, as noted in Chapter 6, a large percentage change in a commodity's price may have only a small percentage effect on the retail price of the product in which the commodity's input value is small. One should not expect to find a one-to-one correspondence between farm and retail price changes. This result contrasts with the argument, made occasionally, that competitive markets require a transmission elasticity of 1. This example, again, demonstrates that a formal model of the structure, when converted to an EDM framework, can clarify one's thinking about relationships among variables.

### Limitations

Although EDMs are useful, they also have limitations. As emphasized in Chapters 14 and 15, models are not perfect. The analyst must be sensitive to whether the model specification is appropriate for addressing the research problem.

Even if the model appears appropriate, there is still a question about the quality of the parameter estimates. Empirical EDMs frequently rely on elasticity estimates from a variety of sources. These elasticities may be econometric estimates from related models or judgmental values. Clearly, if the empirical estimates of the structural elasticities are in error, then the estimated coefficients in the EDM will also be in error.

It is common to undertake sensitivity analyses. The values of the structural parameters are varied, and the reduced-form parameters are computed for the various combinations used. The researcher presumably can make a judgment about the plausible range of the structural parameters, and the hope is that the range of results over the various combinations will still be informative. A potential problem is, of course, that the range of results is so large that it is not helpful in decision making. Kinnucan and Forker's (1987) results for transmission elasticities built on Gardner's (1975) model and were reported for six combinations of underlying structural parameters. Davis and Espinoza (1998) presented a more formal approach to sensitivity analysis in a Bayesian framework, and they also used Gardner as a point of departure.

Another question is, should the elasticities used to compute the EDM coefficients be thought of as short- or long-run elasticities? Presumably, if the analyst is interested in the total effect of the initial change in an exogenous variable on the endogenous variable, then long-run elasticities are the relevant coefficients. They represent the total response of an endogenous variable to a given change in an exogenous variable. This would be the case, for example, in evaluating advertising effects. Clearly, whether one uses a short-run or long-run elasticity can make a large difference in a final result. A related notion is that all the elasticities used in the EDM should be internally consistent.

EDMs are specified in terms of percentage changes, and as noted previously, the functional form can make a difference in the results as one gets further from the middle of the data. The EDM assumes that the relationships can be represented by percentage changes, and typical applications should be interpreted in terms of small percentage changes occurring at the mean of the data.

## Concluding Remarks

The examples presented in this chapter make at least two important points. First, price analyses arise from specific problems. The model, data, and analysis must be consistent with the intended application. This is true both for structural and time-series modeling. The first question a researcher must ask is, why am I doing this?

Second, economic structure matters. Even when we are estimating simple one-equation, two-variable models, they arise from an implicit structural context. An equation that makes farm price a function of the inventory-to-use ratio may be informative, but it is important to remember that the data-generating process for prices is more complex than this simple equation implies.

Likewise, it is possible to estimate reduced-form equations directly by least squares, and such an equation may be used for forecasting. But, if the analyst wants to have a deeper understanding of the effects of exogenous variables, it is helpful to understand how the reduced-form parameters relate to the structure.

## References

Davis, G. C., and M. C. Espinoza. 1998. "A Unified Approach to Sensitivity Analysis in Equilibrium Displacement Models," *Am. J. Ag. Econ.* 80:868–879.

Gardner, B. L. 1975. "The Farm-Retail Price Spread in a Competitive Food Industry," *Am. J. Ag. Econ.* 57:399–409.

Heifner, R. G. 1966. "The Gains from Basing Grain Storage Decisions on Cash-Futures Spreads," *J. Farm Econ.* 48:1490–1495.

Hranaiova, J., and W. G. Tomek. 2002. "Role of Delivery Options in Basis Convergence," *J. Fut. Mkts.* 22:783–809.

Irwin, S. H., P. Garcia, D. L. Good, and E. L. Kunda. 2011. "Spreads and Non-Convergence in Chicago Board of Trade Corn, Soybean, and Wheat Futures: Are Index Funds to Blame?" *Applied Econ. Persp. & Policy* 33:116–142.

Kaiser, H. M. 1997. "Impact of National Generic Dairy Advertising on Dairy Markets," *J. Ag. & Applied Econ.* 29:303–313.

———. 2010. *Measuring the Impacts of Generic Fluid Milk and Dairy Marketing*. Res. Bul. 2010-01, Charles H. Dyson School of Applied Economics and Management, Cornell Univ.

Kaiser, H. M., O. D. Forker, J. Lenz, and C. H. Sun. 1994. "Evaluating Generic Dairy Advertising Impacts on Retail, Wholesale, and Farm Milk Markets," *J. Ag. Econ. Res.* 44(4):3–17.

Kinnucan. H. W. 1996. "A Note on Measuring Returns to Generic Advertising in Interrelated Markets," *J. Ag. Econ.* 47:261–267.

Kinnucan. H. W., and O. D. Forker. 1987. "Asymmetry in Farm-Retail Price Transmission for Major Dairy Products," *Am. J. Ag. Econ.* 69:285–292.

Kinnucan. H. W., and Y. Miao. 2000. "Distributional Impacts of Advertising on Related Commodity Markets," *Am. J. Ag. Econ.* 82:672–678.

Liu, D. J., H. M. Kaiser, T. D. Mount, and O. D. Forker. 1991. "Modeling the U.S. Dairy Sector with Government Intervention," *West. J. Ag. Econ.* 16:360–373.

Peck, A. E., and J. C. Williams. 1991. "Deliveries on the Chicago Board of Trade Wheat, Corn, and Soybean Contracts, 1964/65–1988/89," *Food Res. Inst. Studies* 22:129–225.

Tomek, W. G., and H. M. Kaiser. 1999. "On Improving Econometric Analyses of Generic Advertising Impacts," *Agribusiness* 15:485–500.

USDA. 2013a. *World Agricultural Supply and Demand Estimates*. Office of the Chief Economist. WASDE-516, March 8.

———. 2013b. *Daily Grain Review*. AMS Livestock, Poultry and Grain Market News. Selected dates available through http://www.ams.usda.gov/mnreports/lsddgr.pdf.

Waugh, F. V. 1961. "The Place of Least Squares in Econometrics," *Econometrica* 29:386–396.

Westcott, P. C., and L. A. Hoffman. 1999. *Price Determination for Corn and Wheat: The Role of Market Factors and Government Programs*. USDA Tech. Bul. 1878.

Working, H. 1953. "Hedging Reconsidered," *J. Farm Econ.* 35:544–561.

# Index