22. Quality Change and Hedonics

22.1   Chapters 16 to 18, and 21 cover theoretical issues relating to the choice of index number formulas and are based on a simplifying assumption: that the aggregation was over the same matched $i = 1....n$ items in the two periods being compared. This meets the needs of the discussion of alternative index number formulas, since a measure of price change between two periods requires the quality of each item to remain the same. The practical compilation of XMPIs involves defining the *price basis* (quality specification and terms of sale) of a sample of items in an initial period and monitoring the prices of this matched sample over time, so that only pure price changes are measured, not price changes tainted by changes in quality. In practice this matching becomes imperfect. The quality of what is produced *does* change and, furthermore, new goods (and services) appear on the market that the matched sampling ignores. The relative price changes of these new goods may differ from those of the existing ones, leading to bias in the index if they are excluded. In this chapter, a theoretical framework is outlined that extends the definition of items to include their quality characteristics. The focus of the chapter is on the *economic* theory of the market for quality characteristics and its practical manifestation in hedonic regression outlined in Chapter 8, Section E.4. This chapter provides a *background* for the more practical issues relating to quality adjustments in Chapter 8 and item substitution in Chapter 9.

New and Disappearing Items and Quality Change: Introduction

22.2   The assumption in the previous chapters was that the same set of items was being compared in each period. Such a set can be considered as a sample from all the matched items available in periods 0 and t—the *intersection universe*, which includes only matched items. [1] Yet for many commodity markets old items disappear and new items appear. Constraining the sample to be drawn from this intersection universe is unrealistic. Establishments may produce an item in period 0, but it may not be sold in subsequent periods t.[2] New items may be introduced after period 0 that cannot be compared with a corresponding item in period 0. These items may be variants of the old existing one, or provide totally new services that cannot be directly compared with anything that previously existed. This universe of all items in periods 0 and t is the dynamic *double universe*.

22.3   There is a third universe from which prices might be sampled: a *replacement universe*. The prices reported by establishments are those for an agreed *price basis*—a detailed description of the item being sold and the terms of the transaction. The price basis for items in period 0 are first determined, and then their prices are monitored in subsequent periods. If the item is discontinued and there are no longer prices to record for a particular price basis, prices of a comparable replacement item may be used to continue the series of prices. This universe is a *replacement universe* that starts with the base-period universe, but it

---

[1]The terminology is credited to Dalén (1998a), see also Appendix 8.1.
[2]Its absence may be temporary, being a seasonal item, and specific issues and methods for such temporarily unavailable items are considered in Chapter 9. The concern here is with items that disappear permanently.

also includes one-to-one replacements when an item from the sample in the base period is missing in the current period.

22.4   When a comparable replacement is unavailable, a noncomparable one may be selected. In this case, an explicit adjustment has to be made to the price of either the old or the replacement item for the quality difference. Since the replacement is of a different quality to the old item, it is likely to have a different price basis. Alternatively, assumptions may be made so that the price change of the old item (had it continued to exist) follows those of other items, keeping to the matched universe. In this second case, an implicit adjustment is being made for quality changes, so that the difference in price changes for the group and the old item (had it continued to exist) is equivalent to their quality differences.[3] What is stressed here is that the problem of missing items is the problem of adjusting prices for quality differences.

22.5   Three practical problems emerge. First, is the problem of explicit quality adjustment between a replacement and old item. The item is no longer produced, a replacement is found that is not strictly comparable in quality, the differences in quality are identified, and a price has to be put on these differences if the series of prices for the new replacement item are to be used to continue those of the old series.

22.6   Second, in markets where the turnover of items is high, the sample space selected from the matched universe is going to become increasingly unrepresentative of the dynamic universe, as argued in detail in Chapters 8 and 9. Even the replacement universe may be inappropriate, as it will be made of series carrying with them quality adjustments in each period whose overall accuracy, given the rapidly changing technology, may be tenuous. In such cases, it may be that prices are no longer collected from a matched sample but from a sample of the main items available in each period even though they are of a different quality. A comparison between the average prices of such items would be biased if, say, the quality of the items was improving. The need for, and details of, mechanisms to remove the effects of such changes from the average price comparisons were discussed in some detail in Chapter 8, Section G.

22.7   Finally, there is the problem of new and disappearing goods and services—when the new item is not a variant of the old but provides a completely new service. It is not possible to use it as a replacement for an old item by adjusting a price for the quality differential because what it provides is, by definition, something new.

22.8   There are a number of approaches to quality adjustment, and these are considered in Chapter 8. One of the approaches is to make explicit adjustments to prices for the quality difference between the old and replacement item using the coefficients from hedonic regression equations. *Hedonic regressions* are regressions of the prices of individual models of a product on their characteristics—for example, the prices of television sets on screen size, stereo sound, and text retrieval. The coefficients on such variables provide estimates of the monetary values of different quantifiable characteristics of the product. They can be used to

---

[3]Such methods and their assumptions are outlined in detail in Chapter 8.

adjust the price of a noncomparable replacement item for quality differences compared with the old item—for example, the replacement television set may have text-retrieval facilities that the previous version did not. Yet, it is important that a clear understanding exists of the meaning of such estimated coefficients if they are to be used for quality adjustment, especially given that their use is being promoted.[4] To understand what these estimated parameters mean, it is first necessary to conceive of products as aggregates of their characteristics because, unlike items, characteristics have no separate prices attached to them. The price of the item is the price of a "tied" bundle of characteristics. One must also consider what determines the prices of these characteristics. Economic theory points toward examining demand and supply factors (Sections B.2 and B.3) and the interaction of the two to determine an equilibrium price (Section B.4). Having developed the analytical framework for such prices, it is then necessary to see what interpretation the economic theoretic framework allows us to put on these calculated coefficients (Section B.5). It will be seen that unless there is uniformity of buyers' tastes or 'technologies, an identification problem prevents an unambiguous supply or demand interpretation. Borrowing a framework by Diewert (2002d), a demand-side interpretation that assumes firms are competitive price takers is provided, which, under this user-value approach, shows the assumptions required to generate such meaningful coefficients (Section B6). All of the aforementioned analysis assumes competitive behavior, an assumption which will be relaxed in Section B7.

22.9   Chapter 8, Section G recommends two main approaches for handling industries with rapid turnover of items. If the sample in period 0 is soon outdated, the matched universe, with even one-on-one replacements, can become increasingly unrepresentative of the double universe, and repeated sampling from the double universe is required. In this case either chained indices are advised, as in Chapter 8, Section G.3, or one of a number of *hedonic indices*, described in Chapter 8, Section G.2. Such indices differ from the use of hedonic regression equations for adjusting prices for quality differences for a missing item. These indices use hedonic regressions, say, by including a dummy variable for time on the right hand side of the equation, to estimate the quality-adjusted price change, as outlined below in Section C and in Chapter 8. An understanding of hedonic regression equations requires that the economic theory of output price indices, outlined in Chapter 18, be developed to include goods that can be defined in terms of tied bundles of their characteristics. *Theoretical output price indices* are defined that include changes in the prices of characteristics. Yet, as with the output price indices for goods considered in Chapter 18, there are many formulations that hedonic indices can take, and analogous issues and formulas arise here when discussing alternative approaches in Sections C.3–C.6.

22.10   The estimation of hedonic regressions and the testing of their statistical properties are facilitated by the availability of user-friendly, yet powerful, statistical and econometric software. There are many standard issues in the estimation of regression equations, which can be examined by the diagnostics tests available in such software, as discussed in Kennedy (2003) and Maddala (1988). However, there are issues on functional form, the use of weighted least squares estimators, and specifications that are quite specific to the estimation

---

[4]See Boskin (1996; 1998) and Schultze and Mackie (2002) on this point.

of hedonic equations. While many of these are taken up in Chapter 8, where an illustration is provided, Appendix 21.1 considers some of the theoretical issues.

22.11   Finally, in Section D, economic theory will be used to advise on the problem of new and disappearing goods and services. This problem arises where differences between existing goods and services and the new goods and services are substantive and cannot be meaningfully compared with an old item, even with a quality adjustment. The economic theory of reservation prices will be considered and some issues about its practical implementation will be discussed.

Hedonic Prices and Implicit Markets

B.1   Items as tied bundles of characteristics

22.12   A *hedonic regression* is a regression equation that relates the prices of items, p to the quantities of characteristics, given by the vector $z = (z_1, z_2, ...,z_n)$, that is,

$$(22.1) \; p(z) = p(z_1, z_2, ...,z_n),$$

where the items are defined in terms of varying amounts of their characteristics. In practice, what will be observed for each item or variant of the commodity is its price, a set of its characteristics, and possibly the quantity and thus, the value sold. Empirical work in this area has been concerned with two issues: estimating how the price of an item changes as a result of unit changes in each characteristic—that is, the estimated coefficients of equation (22.1)— and estimating the demand and supply functions for each characteristic. The depiction of an item as a basket of characteristics, each characteristic having it own implicit (shadow) price, requires in turn the specification of a market for such characteristics, since prices result from the workings of markets. Houthakker (1952), Becker (1965), Lancaster (1966), and Muth (1966) have identified the demand for items in terms of their characteristics. The sale of an item is the sale of a tied bundle of characteristics to consumers, whose economic behavior in choosing between items is depicted as one of choosing between bundles of characteristics.[5] However, Rosen (1974) further developed the analysis by providing a structural market framework in terms of both producers and consumers. There are two sides: demand and supply. How much of each characteristic is supplied and consumed is determined by the interaction of the demand for characteristics by consumers and the supply of characteristics by producers. These are considered in turn.

B.2   Consumer or demand side

22.13   Figure 8.1 in Triplett (1987, p. 634) presents a simplified version of the characteristic space between two characteristics. This figure is reproduced as Figure 21.1 below. The hedonic surfaces denoted by p_1 and p_2 in that figure trace out all the combinations of the two characteristics z_1 and z_2 that can be purchased at prices p_1 and p_2. An indifference curve q_j*

---

[5] Consumers are typically assumed to have preferences over alternative combinations of characteristics that give rise to continuously differentiable price functions. However, for some models, the price functions are piecewise linear and hence continuous but not differentiable; e.g., see Lancaster (1971) or Gorman (1980).

maps the combinations of z_1 and z_2 that the consumer is indifferent against purchasing; that is, the consumer will derive the same utility from any point on the curve. The tangency of q_j^* with p_1 at A is the solution to the utility-maximization problem for a given budget (price p_1) and tastes (reflected in q_j^*).

22.14 The slope of the hedonic surface is the marginal cost of acquiring the combination of characteristics, and the slope of the utility function is the marginal utility gained from their purchase. The tangency at A is the utility-maximizing combination of characteristics to be purchased at that price. If consumers purchased any other combination of characteristics in the space of Figure 21.1, it would either cost them more to do so or lead to a lower level of utility. Position A', for example, has more of both z_1 and z_2, and the consumer receives a higher level of utility being on q_j, but the consumer also has to have a higher budget and pays p_2 for being there. Note that the hedonic surface depicted here is nonlinear, so that relative characteristic prices are not fixed. The consumer with tastes q_k^* chooses characteristic set B at p_1. Thus, the data observed in the market depends on the set of tastes. Triplett (2002) has argued that if tastes were all the same, then only one model of a personal computer would be purchased. But in the real world more than one model does exist, reflecting heterogeneous tastes and income levels. Rosen (1974) shows that of all the characteristic combinations and prices at which they may be offered, the hedonic surface traces out an envelope[6] of tangencies including those on q_j^* and q_k^* on p_1 in Figure 21.1. This envelope is simply a description of the locus of the points chosen. Since rational consumers who optimize are assumed, these are the points that will be observed in the market and are thus, used to estimate the hedonic regression. Alternative z points on the same indifference curve will allow the relative price of z_1 to z_2 to be determined. However, observed data are likely to result from a the locus of points on an expansion paths such as $A\,A'$. There may be expansion paths for consumers with different income levels and tastes, such as B, and this may give rise to conflicting valuations, so that the overall parameter estimates determined by the regression from transactions observed in the market are an amalgam of such data. And of course this would just be a reflection of the reality of economic life. What arises from this exposition is the fact that the form of the hedonic function is determined in part by the distribution of buyers and their tastes in the market.

22.15 The exposition is now formalized to include parameters for tastes and a numeraire commodity[7] against which combinations of other aggregates are selected following Rosen (1974). The hedonic function $p(z)$ describes variation in the market price of the items in terms of their characteristics. The consumer purchase decision is assumed to be based on utility-maximization behavior, the utility function being given by $U(z, x;\alpha)$, where x is a numeraire commodity, the maximization of utility being subject to a budget constraint given by income y measured as $y = x + p(z)$ (the amount spent on the numeraire commodity and the

---

[6]An envelope is more formally defined by letting $f(x,y,k) = 0$ be an implicit function of x and y. The form of the function is assumed to depend on k, the tastes in this case. A different curve corresponds to each value of k in the xy plane. The envelope of this family of curves is itself a curve with the property that it is tangent to each member of the family. The equation of the envelope is obtained by taking the partial derivative of f(x,y,k) with respect to k and eliminating k from the two equations $f(x,y,k) = 0$ and $fk(x,y,k) = 0$. (See Osgood, 1925.)

[7]The numeraire commodity represents all other goods and services consumed—it represents the normal nonhedonic commodities. The price of x is set equal to unity; $p(z)$ and income are measured in these units.

hedonic commodities), and α is a vector of the features of the individual consumer that describe their tastes. Consumers maximize their utility by selecting a combination of quantities of x and characteristics z subject to a budget constraint. The market is assumed to be competitive and consumers are described as price takers, they purchase only the one item, so their purchase decision does not influence the market price. The price they pay for a combination of characteristics, vector z, is given by $p(z)$. Since they are optimizing consumers the combination chosen is such that

22.16   (22.2) $[\partial U(z, y - p(z);\alpha) / \partial z_i] / [\partial U(z, y - p(z);\alpha) / \partial x] = \partial p(z) / \partial z_i \equiv p_i(z)$,

where $\partial p(z) / \partial z_i$ is the first derivative of the hedonic function in equation (22.1) with respect to each z characteristic. The coefficients of the hedonic function are equal to their shadow price p_i, which measure the utility derived from that characteristic relative to the numeraire good for given budgets and tastes.

22.17   A *value function* θ can be defined as the value of expenditure a consumer with tastes α is willing to pay for alternative values of z at a given utility u and income y represented by $\theta(z;u,y,\alpha)$. It defines a family of indifference curves relating the z_i to foregone x, 'money'. For individual characteristics z_i, θ are the marginal rate of substitution between z_i and money, or the implicit marginal valuation the consumer with tastes α puts on z_i at a given utility level and income. It is an indication of the reservation demand price[8] for additional units of z_i.[9] The price in the market is $p(z)$, and utility is maximized when $\theta(z;u,y,\alpha) = p(z)$, that is the purchase takes place where the surface of the indifference curve θ is tangent to the hedonic price surface. If different buyers have different value functions (tastes), some will buy more of a characteristic than others for a given price function, as illustrated in Figure 21.1.

22.18   The joint distribution function of tastes and income sets out a family of value functions, each of which, when tangential to the price function, depicts a purchase and simultaneously defines the price function whose envelope is the market hedonic price function. The points of purchase traced out by the hedonic function thus, depend on the budget of the individual and the tastes of the individual consumer purchasing an individual set of characteristics. If demand functions are to be traced out, the joint probability distribution of consumers with particular budgets and tastes occurring in the market needs to be specified, that is, $F(y, \alpha)$. This function, along with equation (22.1) ,allows the demand equations to be represented for each characteristic.

---

[8]This is the hypothetical price that makes the demand for the characteristic equal to zero that is, it is the price that, when inserted into the demand function, sets demand to zero.
[9]The utility function is assumed strictly concave so that θ is concave in z, and the value function is increasing in z_i at a decreasing rate.

B.3   Producer or supply side.

22.19   Again referring to Triplett's (1987) Figure 8.1, it also shows the production side. In Chapter 18, Section B.1, a revenue-maximizing producer was considered whose revenue-maximization problem was given by equation (17.1)[10]:

$$(22.3)\ R(p,v) \equiv \max{}_q \left[\sum_{n=1}^{N} p_n q_n : q \text{ belongs to } S(v)\right],$$

where $R(p,v)$ is the maximum value of output, $\sum_{n=1}^{N} p_n q_n$, that the establishment can produce, given that it faces the vector of output prices p and given that the vector of inputs v is available for use, using the period t technology. Figure 17.1 illustrated in goods-space how the producer would choose between different combinations of outputs, q_1 and q_2. In Figure 21.1, the characteristics-space problem is analogous to the goods-space one with producers choosing here between combinations of z_1 and z_2 to produce for a particular level of technology and inputs $S(v)$. For a particular producer with level of inputs and technology $S*_G$ facing a price surface p_1, the optimal production combination is at A. However, a different producer with technology and inputs $S*_H$ facing a price surface p_1 would produce at B. At these points, the marginal cost of z_1 with respect to z_2 is equal to its marginal price from the hedonic surface as depicted by the tangency of the point. Production under these circumstances at any other combination would not be optimal. The envelope of tangencies such as $S*_G$ and $S*_H$ trace out the production decisions that would be observed in the market from optimizing, price-taking producers and be used as data for estimating the hedonic regressions. The hedonic function can be seen to be determined, in part, by the distribution of technologies of producers, including their output scale.

22.20   Rosen (1974) formalizes the producer side, whereby price-taking producers are assumed to have cost functions described by $C(M, z; \tau)$[11] where $Q = Q(z)$ is the output scale, that is, the number of units produced by an establishment offering specifications of an item with characteristics z. They have to decide which items to produce, that is, which package of z to produce. The solution for each producer is to choose the output that minimizes costs given its own technology: the output combinations each producer can produce with given input costs using its factors of production and factor prices the technology. The cost function includes τ, equivalent to $S(v)$ above, a vector of the technology and inputs of each producer. It is the variation in τ across producers that distinguishes producer A's decision about which combination of z to produce from that of producer B in Figure 21.1. Producers are optimizers who seek to maximize profits given by

$$(22.4)\ Q\,p(z) - C(Q, z; \tau)$$

---

[10]The time superscripts are not relevant in this context.
[11]The cost function is assumed to be convex with no indivisibilities. The marginal cost of producing one more item of a given combination of characteristics is assumed to be positive and increasing, and, similarly, the marginal cost of increasing production of each component characteristic is positive and nondecreasing.

by selecting Q and z optimally. The supplying market is assumed to be competitive, and producers are price takers so the producers cannot influence price by their production decision. Their decision about how much to produce of each z is determined by the price of z, assuming that the producer can vary Q and z in the short run.[12] Dividing equation (22.4) by Q and setting the resulting expression equal to zero, the first-order profit-maximizing conditions are given by

$$(22.5) \quad \frac{\partial p}{\partial z_i} = p_i = \frac{C_{zi}(Q, z; \tau)}{Q}$$

where $p = p(z_1, z_2, \ldots, z_n)$ as in equation (22.1)

22.21 The *marginal unit revenue* from producing characteristic z_i is given by its shadow price in the price function and its marginal cost of production. In the producer case, a knowledge of the probability distribution of the technologies of firms, $G(\tau)$, is necessary if the overall quantity supplied of items with given characteristic sets are to be revealed. Since it is a profit-maximization problem to select the optimal combination of characteristics to produce, marginal revenue from the additional attributes must equal their marginal cost of production per unit sold. Quantities are produced up to the point where unit revenues $p(z)$ equal marginal production costs, evaluated at the optimum bundle of characteristics supplied.

22.22 While for consumers a *value function* was considered, producers require an *offer function* $\phi(z; \pi, \tau)$. The offer price is the price the seller is willing to accept for various designs at constant profit level π, when quantities produced are optimally chosen, while $p(z)$ is the maximum price obtainable from those models in the market. Producer equilibrium is characterized by a tangency between a profit characteristics indifference surface and the market characteristics price surface, where $p_i(z_i) = \phi_{zi}(z; \pi, \tau)$ and $p(z) = \phi_z(z; \pi, \tau)$. Since there is a distribution of technologies $G(\tau)$, the producer equilibrium is characterized by a family of offer functions that envelop the market hedonic price function. The varying τ will depend on different factor prices for items produced in different countries, multiproduct firms with economies of scale, and differences in the technology, whether the quality of capital, labor, or intermediate inputs and their organization. Different values of τ will define a family of production surfaces.

B.4   Equilibrium

22.23 The theoretical framework first defined each item as a point on a plane of several dimensions made up by the $z_1, z_2, \ldots, z_n$ quality characteristics; each item was a combination of values $z_1, z_2, \ldots, z_n$. If only two characteristics defined the item, then each point in the positive space of Figure 21.1 would define an item. The characteristics were not bought

---

[12]Rosen (1974) considered two other supply characterizations: the short run in which only Q is variable, and a long run in which plants can be added and retired. The determination of equilibrium supply and demand is not straightforward. A function $p(z)$ is required such that market demand for all z will equate to market supply and clear the market. But demand and supply depend on the whole $p(z)$, since any adjustment to prices to equate demand and supply for one combination of items will induce substitutions and changes for others. Rosen (1974, pp. 44–48) discusses this in some detail.

individually but as bundles of characteristics tied together to make up an item. It was assumed that the markets were differentiated so that there was a wide range of choices to be made.[13] The market was also assumed to be perfectly competitive with consumers and producers as price takers undertaking optimizing behavior to decide which items (tied sets of characteristics) to buy and sell. Competitive markets in characteristics and optimizing behavior are assumed so that the quantity demanded of characteristics z must equal the quantity supplied. It has been shown that consumers' and producers' choices or "locations" on the plane will be dictated by consumer tastes and producer technology. Tauchen and Witte (2001, p. 4) show that the hedonic price function will differ across markets in accordance with the means and variances (and in some cases also higher moments) of the distributions of household and firm characteristics.

22.24   Rosen (1974, p. 44) notes that a buyer and seller are perfectly matched when their respective value and offer functions are tangential. The common gradient at that point is given by the gradient of the market-clearing implicit price function $p(z)$. The consumption and production decisions were seen in the value and offer functions to be jointly determined, for given $p(z)$, by $F(y, \alpha)$ and $G(\tau)$. In competitive markets there is a simultaneity in the determination of the hedonic equation, since the distribution of $F(y, \alpha)$ and $G(\tau)$ help determined the quantities demanded and supplied and also the slope of the function. Although the decisions made by consumers and producers are as price takers, the prices taken are those from the hedonic function. There is a sense in which the hedonic function and its shadow prices emerge from the operations of the market. The product markets implicitly reveal the hedonic function. Since consumers and producers are optimizers in competitive markets, the hedonic function, in principle, gives the minimum price of any bundle of characteristics. Given all of this, Rosen (1974, p. 44) asked: what do hedonic prices mean?

B.5   What do hedonic prices mean?

22.25   It would be convenient if, for MPI construction, the estimated coefficients from hedonic regressions were estimates of the marginal production cost or producer value of a characteristic or, for XPI construction, they were estimates of the marginal nonresident user value from a characteristic. But theory tells us that this is not the case and that the interpretation is not clear.

22.26   There was an erroneous perception in the 1960s that the coefficients from hedonic methods represented user-values as opposed to resource-costs. Rosen (1974), as has been shown, found that hedonic coefficients generally reflect both user-values and resource-costs; both supply and demand situations. The ratios of these coefficients may reflect consumers' marginal rates of substitution or producers' marginal rates of substitution (transformation) for characteristics. There is what is referred to in econometrics as an "identification"' problem in which the observed prices and quantities are jointly determined by supply and demand considerations, and their underlying effects cannot be separated. The data collected on prices

---

[13]In order to ensure that choices among combinations of z are continuous, assume further that $p(z)$ possesses continuous first order derivatives.

jointly arise from variations in demand by different consumers with different tastes and preferences, and from variations in supply by producers with different technologies.

22.27   First, it is necessary to come to terms with this simultaneity problem. Hedonic regressions are an increasingly important analytical tool, one implicitly promoted by the attention given to it in this *Manual* but also promoted in separate manuals by organizations such as the OECD (see Triplett, 2002), and Eurostat (2001), and widely used by the U.S. Bureau of Labor Statistics (Kokoski, Waehrer, and Rozaklis, 2001, and Moulton, 2001b). So how do economists writing on the subject shrug their intellectual shoulders in the light of these findings?

22.28   Rosen (1974, p. 43) refers to the hedonic function as "..a joint envelope of a family of value functions and another family of offer functions. An envelope function by itself reveals nothing about the underlying members that generate it; and they in turn constitute the generating structure of the observations."

22.29   Griliches (1988, p. 120) notes the following:

> My own view is that what the hedonic approach tries to do is to estimate aspects of the budget constraint facing consumers, allowing thereby the estimation of "missing" prices when quality changes. It is not in the business of estimating utility functions *per se*, though it can also be useful for these purposes….what is being estimated is the actual locus of intersection of the demand curves of different consumers with varying tastes and the supply curves of different producers with possible varying technologies of production. One is unlikely, therefore to be able to recover the underlying utility and cost functions from such data alone, except in very special circumstances.

22.30   Triplett (1987) states, "It is well-established—but still not widely understood—that the form of $h(\cdot)$ [the hedonic function] cannot be derived from the form of $Q(\cdot)$ and $t(\cdot)$ [utility and production functions], nor does $h(\cdot)$ represent a "reduced form" of supply and demand functions derived from $Q(\cdot)$ and $t(\cdot)$."

22.31   Diewert (2003, p. 320) with his focus on the consumer side, says;

> Thus, I am following Muellbauer's (1974, p. 977) example where he says that his "approach is unashamedly one-sided; only the demand side is treated…Its subject matter is therefore rather different from that of the recent paper by Sherwin Rosen. The supply side and simultaneity problems which may arise are ignored."

Diewert (2002e) has also considered the theoretical PPI indices with a focus on the producer side. He bases the optimizing problem the establishments face when deciding on which combinations of characteristics to produce, however, on the consumer's valuations, giving them precedence. There are many industries in which firms are effective price takers, and the prices taken are dictated by the consumer side rather than by cost and technological considerations. In Section B.6 this framework is outlined, which allows a more straightforward development of the theory of hedonic index numbers for XMPIs.

22.32   Second, the Rosen's theoretical framework allows the conditions to be considered under which the hedonic coefficients are determined by only demand side or supply side factors—the circumstances under which clear explanations would be valid. The problem is that because the coefficients of a hedonic function are the outcome of the interaction of consumer and producer optimizing conditions, it is not possible to interpret the function only in terms of, say, producer marginal costs or consumer marginal values. However, suppose the *production technology* τ was the same for each producing establishment. Buyers differ but sellers are identical. Then, instead of a confusing family of offer functions, there is a unique offer function with the hedonic function describing the prices of characteristics the firm will supply with the given ruling technology to the current mixture of tastes. The offer function becomes $p(z)$, since there is no distribution of τ to confuse it. There are different tastes on the consumer side, and so what appears in the market is the result of firms trying to satisfy consumer preferences all for a constant technology and profit level; the structure of supply is revealed by the hedonic price function. In Figure 21.1 only the expansion path traced out by, say S_H* akin to $A\,A'$, would be revealed. Now, suppose sellers differ, but *buyers' tastes* α are identical. Here the family of *value functions* collapses to be revealed as the hedonic function $p(z)$ which identifies the structure of demand, such as $A\,A'$ in Figure 21.1.[14] Diewert's (2003) approach follows a representative consumer, rather than consumers with different tastes, so that the demand side alone can be identified. Triplett (1987, p. 632) notes that of these possibilities, uniformity of technologies is the most likely, especially when access to technology is unrestricted in the long run, while uniformity of tastes is unlikely. There may, of course, be segmented markets where tastes are more uniform to which specific sets of items are tailored and for which hedonic equations can be estimated for individual segments.[15] In some industries there may be a prior expectation of uniformity of tastes against uniformity of technologies and interpretation of coefficients will accordingly follow. In many cases, however, the interpretation may be more problematic. The pure producer approach requires assumptions of uniformity of technology and input prices which cannot of course be generally assumed. But the key assumption that will not generally be satisfied in the producer context is that each *producer is able to produce the entire array of hedonic models* whereas, in the consumer context, it is quite plausible that each consumer has the possibility of purchasing and consuming each model.

22.33   Third, issues relating to the estimation of the underlying supply and demand functions for characteristics have implications for the estimation of hedonic functions. In Appendix 21.2, identification and estimation issues will be considered in this light. Finally,

---

[14]Correspondingly, if the supply curves were perfectly inelastic, so that a change in price would not affect the supply of any of the differentiated products, then the variation in prices underlying the data and feeding the hedonic estimates would be determined by demand factors. The coefficients would provide estimates of user values. Similarly, if the supplying market were perfectly competitive, the estimates would be of resource costs. None of the price differences between differentiated items would be due to, say, novel configurations of characteristics, and no temporary monopoly profit would be achieved as a reward for this, or as a results of the exercise of market power. See Berndt (1983).

[15]Berry, Levinsohn, and Pakes (1995) provide a detailed and interesting example for automobiles in which makes are used as market segments, while Tauchen and Witte (2001) provide a systematic theoretical study of estimation issues for supply, demand, and hedonic functions where consumers and producers and their transactions are indexed across communities.

the subsequent concern with new products in Section D of this chapter refers to demand functions. However, attention is now turned to hedonic *indices*. In the next section, these are noted to have a quite different application than that for the quality adjustment of noncomparable replacement items.

B.6   An alternative hedonic theoretical formulation

22.34   This section is based on a formulation by Diewert (2002d). It assumes competitive price-taking behavior on the part of firms. In this approach, the user's valuations of the various models that could be produced flow to producers via the hedonic function in the same way that output prices are taken, as given in the usual theory of the output price index. It is necessary to set up the establishment's revenue-maximization problem assuming that it produces a single output, but in each period, the establishment has a choice of which type of model it could produce. Let the model be identified by a K dimensional vector of characteristics, $z \equiv [z_1,...,z_K]$. Before tackling the establishment's revenue maximization problem, it is necessary to characterize the set of output prices that the establishment faces in period t as a function of the characteristics of the model that the establishment might produce. It is assumed that in period t, the demanders of the output of the establishment have a cardinal utility function, $f^t(z)$, that enables each demander to determine that the value of a model with the vector of characteristics $z^1 \equiv [z_1^1,...,z_K^1]$ compared with a model with characteristics vector $z^2 \equiv [z_1^2,...,z_K^2]$ is $f^t(z^1)/f^t(z^2)$. Thus,, in period t, demanders are *willing to pay* the amount of money $P^t(z)$ for a model with the vector of characteristics z where:

22.35   (22.6) $\Pi^t(z) \equiv \rho^t f^t(z), \quad t = 0,1$.

The scalar ρ^t is inserted into the willingness-to-pay function because, under certain restrictions, ρ^t can be interpreted as a period t price for the entire family of hedonic models that might be produced in period t. These restrictions are

(22.7) $f^0 = f^1$,

that is, the *model relative utility functions* f^t are identical for the two periods under consideration. We will make use of the specific assumption in equation (22.7) later.

22.36   In what follows, it is assumed that econometric estimates for the period 0 and 1 *hedonic model price functions*, Π^0 and Π^1, are available, although we will also consider the case where only an estimate for Π^0 is available.[16]

---

[16]We will need some identifying restrictions in order to identify the parameters of f^0 and f^1 along with ρ^0 and ρ^1. One common model sets $\rho^0 = 1$ and $f^0 = f^1$. A more general model sets $\rho^0 = 1$ and $f^0(z^*) = f^1(z^*)$ for a reference characteristics vector, $z^* \equiv [z_1^*,...,z_K^*]$.

Now, consider an establishment that produces a single model in each period in the marketplace that is characterized by the hedonic model price functions, $\Pi^t(z)$, for periods $t = 0,1$. Suppose that in period t, the establishment has the *production function F^t*, where

(22.8) $q = F^t(z,v)$

is the number of models, each with vector of characteristics z, that can be produced if the vector of inputs v is available for use by the establishment in period t. As is usual in the economic approach to index numbers, we assume a competitive model, where each establishment takes output prices as fixed parameters beyond its control. In this case, there is an entire schedule of model prices that the establishment takes as given instead of just a single price in each period. Thus,, it is assumed that if the establishment decides to produce a model with the vector of characteristics z, then it can sell any number of units of this model in period t at the price $\Pi^t(z) = \rho^t f^t(z)$. Note that the establishment is allowed to choose which model type to produce in each period.

22.37   Now, define the establishment's *revenue function*, R, assuming the establishment is facing the period s hedonic price function $\Pi^s = \rho^s f^s$ and is using the vector of inputs v and has access to the period t production function F^t:

(22.9) $R(\rho^s f^s, F^t, Z^t, v) \equiv \max_{q,z} \{\rho^s f^s(z)q : q = F^t(z,v) ; z \text{ belongs to } Z^t\}$
$= \max_z \{\rho^s f^s(z)F^t(z,v) : z \text{ belongs to } Z^t\}$,

where Z^t is a *technologically feasible set of model characteristics* that can be produced in period t. The second line follows from the line above by substituting the production-function constraint into the objective function.

22.38   The actual period t revenue-maximization problem that the establishment faces is defined by the revenue function equation (22.9), except that we replace the period s hedonic price function $\rho^s f^s$ by the period t hedonic price function $\rho^t f^t$, and the generic input quantity vector v is replaced by the observed period t input quantity vector used by the establishment, v^t. Further, assume that the establishment produces q^t units of a single model with characteristics vector z^t and that $[q^t,z^t]$ solves the period t revenue-maximization problem—that is, $[q^t,z^t]$ is a solution to[17]

(22.10) $R(\rho^t f^t, F^t, Z^t, v^t) \equiv \max_{q,z} \{\rho^t f^t(z)q : q = F^t(z,v^t) ; z \text{ belongs to } Z^t\}; \qquad t = 0,1$
$= \rho^t f^t(z^t)q^t$

---

[17]If the establishment is competitively optimizing with respect to its choice of inputs as well, then the period t input vector v^t, along with q^t and z^t, are a solution to the following period t profit-maximization problem for the establishment: $\max_{q,z,v} \{\rho^t f^t(z)q - w^t \bullet v : q = F^t(z,v) ; z \text{ belongs to } Z^t\}$, where w^t is a vector of input prices that the establishment faces in period t and $w^t \bullet v$ denotes the inner product of the vectors w^t and v. It is possible to rework our analysis presented below, conditioning on an input price vector rather than on an input quantity vector.

where the period t establishment output q^t is equal to

$$(22.11) \; q^t = F^t(z^t, v^t) \; ; \quad t = 0,1.$$

Now, a family of *Konüs-type hedonic output price indices P* between periods 0 and 1 can be defined as follows:

$$(22.12) \; P(\rho^0 f^0, \rho^1 f^1, F^t, Z^t, v) \equiv R(\rho^1 f^1, F^t, Z^t, v)/R(\rho^0 f^0, F^t, Z^t, v).$$

22.39 Thus, a particular member of the above family of indices is equal to the establishment's revenue ratio, where the revenue in the numerator of equation (22.12) uses the hedonic model price function for period 1, and the revenue in the denominator of equation (22.12) uses the hedonic model price function for period 0. For both revenues, however, the technology of period t is used (that is, F^t and Z^t are used in both revenue-maximization problems), and the same input quantity vector v is used. This is the usual definition for an economic output price index, except that instead of a single price facing the producer in each period, we have a whole family of model prices facing the establishment in each period. Note that the only variables that are different in the numerator and denominator of equation (22.12) are the two hedonic model price functions facing the establishment in periods 0 and 1.

22.40 The right-hand side of equation (22.12) looks a bit complex. However, if the assumption in equation (22.7) holds (that is, the period 0 and 1 hedonic model price functions are identical except for the multiplicative scalars ρ^0 and ρ^1), then equation (22.12) reduces to the very simple ratio, ρ^1/ρ^0. To see this, use definitions equations (22.12) and (22.10) as follows:

$$(22.13) \; P(\rho^0 f^0, \rho^1 f^1, F^t, Z^t, v) \equiv R(\rho^1 f^1, F^t, Z^t, v) / R(\rho^0 f^0, F^t, Z^t, v)$$
$$= \max{}_z \; \{\rho^1 f^1(z) F^t(z,v^t) \; ; \; z \text{ belongs to } Z^t\} / \max{}_z \; \{\rho^0 f^0(z) F^t(z,v^t) \; ; \; z \text{ belongs to } Z^t\}$$
$$= \max{}_z \; \{\rho^1 f^0(z) F^t(z,v^t) \; ; \; z \text{ belongs to } Z^t\} / \max{}_z \; \{\rho^0 f^0(z) F^t(z,v^t) \; ; \; z \text{ belongs to } Z^t\}$$

using equation (22.7)
$$= [\rho^1/\rho^0]\max{}_z \; \{\rho^0 f^0(z) F^t(z,v^t) \; ; \; z \text{ belongs to } Z^t\} / \max{}_z \; \{\rho^0 f^0(z) F^t(z,v^t) \; ; \; z \text{ belongs to } Z^t\}$$
assuming ρ^0 and ρ^1 are positive and canceling terms
$$= \rho^1/\rho^0.$$

This is a very useful result since many hedonic regression models have been successfully estimated using equation (22.7). Under this assumption, *all* the theoretical hedonic establishment output price indices reduce to the observable ratio, ρ^1/ρ^0.

22.41 We return to the general case where the assumption in equation (22.7) is not made. As usual, it is always of interest to specialize equation (22.12) to the special cases where the conditioning variables that are held constant in the numerator and denominator of equation (22.12), F^t, Z^t, and v, are equal to the period 0 and 1 values for these variables, namely, F^0, Z^0, and v^0, and F^1, Z^1, and v^1. Thus, define the *Laspeyres-type hedonic output price index* between periods 0 and 1 for our establishment as follows:

(22.14) $P(\rho^0 f^0, \rho^1 f^1, F^0, Z^0, v^0) \equiv R(\rho^1 f^1, F^0, Z^0, v^0) / R(\rho^0 f^0, F^0, Z^0, v^0)$
$= R(\rho^1 f^1, F^0, Z^0, v^0) / \rho^0 f^0(z^0) q^0$, using equation (22.10) for $t = 0$
$\quad = \max_z \{\rho^1 f^1(z) F^0(z, v^0) ; z \text{ belongs to } Z^0\} / \rho^0 f^0(z^0) q^0$ using equation (22.9)
$\geq \rho^1 f^1(z^0) F^0(z^0, v^0) / \rho^0 f^0(z^0) q^0$ since z^0 is feasible for the maximization problem
$= \rho^1 f^1(z^0) q^0 / \rho^0 f^0(z^0) q^0$ \qquad\qquad using equation (22.11) for $t = 0$
$= \rho^1 f^1(z^0) / \rho^0 f^0(z^0)$
$\equiv P_{HL}$.

where the *observable hedonic Laspeyres output price index* P_{HL} is defined as

(22.15) $P_{HL} \equiv \rho^1 f^1(z^0) / \rho^0 f^0(z^0)$.

Thus, the inequality in equation (22.14) says that the unobservable theoretical Laspeyres-type hedonic output price index $P(\rho^0 f^0, \rho^1 f^1, F^0, Z^0, v^0)$ is bounded from below by the observable (assuming that we have estimates for ρ^0, ρ^1, f^0, and f^1) hedonic Laspeyres output price index P_{HL}. The inequality in equation (22.14) is the hedonic counterpart to a standard Laspeyres-type inequality for a theoretical output price index.

22.42   It is of modest interest to rewrite P_{HL} in terms of the observable model prices for the establishment in periods 0 and 1. Denote these prices by P^0 and P^1, respectively. Using equation (22.6):

(22.16) $P^0 = \rho^0 f^0(z^0)$ and $P^1 = \rho^1 f^1(z^1)$.

Now, rewriting equation (22.15) as follows:

(22.17) $P_{HL} \equiv \rho^1 f^1(z^0) / \rho^0 f^0(z^0)$
$= \rho^1 f^1(z^1)[f^1(z^0) / f^1(z^1)] / \rho^0 f^0(z^0)$
$= P^1[f^1(z^0)/f^1(z^1)] / P^0$ using equation (22.16)
$= [P^1/f^1(z^1)] / [P^0/f^1(z^0)]$.

The prices $P^1 / f^1(z^1)$ and $P^0 / f^1(z^0)$ can be interpreted as *quality adjusted model prices* for the establishment in periods 1 and 0, respectively, using the hedonic regression pertaining to period 1 to do the quality adjustment.

22.43   In the theoretical hedonic output price index $P(\rho^0 f^0, \rho^1 f^1, F^0, Z^0, v^0)$ defined by equation (22.14) above, we conditioned on F^0 (the base-period production function), Z^0 (the base-period set of models that were technologically feasible in period 0), and v^0 (the establishment's base-period input vector). We now define a companion period 1 theoretical hedonic output price that conditions on the period 1 variables, F^1, Z^1, v^1. Thus, define the

Paasche-type hedonic output price index between periods 0 and 1 for an establishment as follows:[18]

(22.18) $P(\rho^0 f^0, \rho^1 f^1, F^1, Z^1, v^1) \equiv R(\rho^1 f^1, F^1, Z^1, v^1) / R(\rho^0 f^0, F^1, Z^1, v^1)$
$= \rho^1 f^1(z^1)q^1 / R(\rho^0 f^0, F^1, Z^1, v^1)$ using equation (22.10) for $t = 1$
$= \rho^1 f^1(z^1)q^1 / \max{}_z \{\rho^0 f^0(z)F^1(z,v^1) ; z \text{ belongs to } Z^1\}$ using equation (22.9)
$\leq \rho^1 f^1(z^1)q^1 / \rho^0 f^0(z^1)F^1(z^1,v^1)$ since z^1 is feasible for the maximization problem
$= \rho^1 f^1(z^1)q^1 / \rho^0 f^0(z^1)q^1$ using equation (22.11) for $t = 1$
$= \rho^1 f^1(z^1) / \rho^0 f^0(z^1)$
$\equiv P_{HP}$,

where the *observable hedonic Paasche output price index* P_{HP} is defined as

(22.19) $P_{HP} \equiv \rho^1 f^1(z^1) / \rho^0 f^0(z^1)$.

Thus, the inequality in equation (22.18) says that the unobservable theoretical Paasche-type hedonic output price index $P(\rho^0 f^0, \rho^1 f^1, F^1, Z^1, v^1)$ is bounded from above by the observable (assuming that we have estimates for ρ^0, ρ^1, f^0, and f^1) hedonic Paasche output price index P_{HP}. The inequality in equation (22.18) is the hedonic counterpart to a standard Paasche-type inequality for a theoretical output price index.

22.44   Again, it is of interest to rewrite P_{HP} in terms of the observable model prices for the establishment in periods 0 and 1. Rewrite equation (22.19) as follows:

(22.20) $P_{HP} \equiv \rho^1 f^1(z^1) / \rho^0 f^0(z^1)$
$= \rho^1 f^1(z^1) / \{\rho^0 f^0(z^0)[f^0(z^1) / f^0(z^0)]\}$
$= P^1/\{P^0[f^0(z^1)/f^0(z^0)]\}$ using equation (22.16)
$= [P^1/f^0(z^1)]/[P^0/f^0(z^0)]$.

The prices $P^1 / f^0(z^1)$ and $P^0 / f^0(z^0)$ can be interpreted as *quality adjusted model prices* for the establishment in periods 1 and 0, respectively, using the hedonic regression pertaining to period 0 to do the quality adjustment.

22.45   It is possible to adapt a technique originally credited to Konüs (1924) and obtain a theoretical hedonic output price index that lies between the observable Laspeyres and Paasche bounding indices, P_{HL} and P_{HP}, defined above. Recall the definition of the revenue function, $R(\rho {}^s f^s, F^t, Z^t, v)$, defined by equation (22.9) above. Instead of using either F^0, Z^0, v^0 or F^1, Z^1, v^1 as reference production functions, feasible characteristics sets, and input vectors for the establishment in equation (22.12), use a *convex combination* or *weighted average* of these variables in our definition of a theoretical hedonic output price index. Thus, for each scalar λ between 0 and 1, define the theoretical hedonic output price index between periods 0 and 1, $P(\lambda)$, as follows:

---

[18]Assume that all ρ^t, $f^t(z)$, and $F^t(z,v^t)$ are positive for $t = 0,1$.

(22.21) $P(\lambda) \equiv R(\rho^1 f^1,(1-\lambda)F^0 + \lambda F^1,(1-\lambda)Z^0 + \lambda Z^1,(1-\lambda)v^0 + \lambda v^1) / R(\rho^0 f^0,(1-\lambda)F^0 + \lambda F^1,$
$\quad (1-\lambda)Z^0 + \lambda Z^1, (1-\lambda)v^0 + \lambda v^1)$
$\quad = \max_z\{\rho^1 f^1(z)[(1-\lambda)F^0(z,(1-\lambda)v^0 + \lambda v^1) + \lambda F^1(z,(1-\lambda)v^0 + \lambda v^1)] : z \text{ belongs to } (1-\lambda)Z^0$
$+ \lambda Z^1\} /$
$\quad\quad \max_z\{\rho^0 f^0(z)[(1-\lambda)F^0(z,(1-\lambda)v^0 + \lambda v^1) + \lambda F^1(z,(1-\lambda)v^0 + \lambda v^1)] : z \text{ belongs to } (1-\lambda)Z^0 +$
$\lambda Z^1\}.$

When $\lambda = 0$, $P(\lambda)$ simplifies to $P(\rho^0 f^0, \rho^1 f^1, F^0, Z^0, v^0)$, the Laspeyres-type hedonic output price index defined by equation (22.14) above. Thus, using the inequality in equation (22.14), we have:

(22.22) $P(0) \geq P_{HL}$,

where P_{HL} is equal to $\rho^1 f^1(z^0)/\rho^0 f^0(z^0)$, the observable Laspeyres hedonic output price index defined by equation (22.15) above. When $\lambda = 1$, $P(\lambda)$ simplifies to $P(\rho^0 f^0, \rho^1 f^1, F^1, Z^1, v^1)$, the Paasche-type hedonic output price index defined by equation (22.18) above. Thus, using the inequality in equation (22.18), we have

(22.23) $P(1) \leq P_{HP} P_{HL}$,

where P_{HP} is equal to $\rho^1 f^1(z^1) / \rho^0 f^0(z^1)$, the observable Paasche hedonic output price index defined by equation (22.20) above.

22.46   If $P(\lambda)$ is a continuous function of λ between 0 and 1, then we can adapt the proof of Diewert (1983a, pp. 1060-1061), which in turn is based on a technique of proof by Konüs (1924), and show that there exists a λ^* such that $0 \leq \lambda^* \leq 1$ and either

(22.24) $P_{HL} \leq P(\lambda^*) \leq P_{HP}$ or $P_{HP} \leq P(\lambda^*) \leq P_{HL}$,

that is, there exists a theoretical hedonic output price index between periods 0 and 1 using a technology that is intermediate to the technology of the establishment between periods 0 and 1, $P(\lambda^*)$, that lies *between* the observable[19] Laspeyres and Paasche hedonic output price indices, P_{HL} and P_{HP}. However, to obtain this result, we need conditions on the hedonic model price functions, $\rho^0 f^0(z)$ and $\rho^1 f^1(z)$, on the production functions, $F^0(z,v)$ and $F^1(z,v)$, and on the feasible characteristics sets, Z^0 and Z^1, that will ensure that the maximum functions in the numerator and denominator in the last equality of equation (22.21) are continuous in λ. Sufficient conditions to guarantee continuity are:[20]

- The production functions $F^0(z,v)$ and $F^1(z,v)$ are positive and jointly continuous in z,v,
- The hedonic model price functions $f^0(z)$ and $f^1(z)$ are positive and continuous in z,
- ρ^0 and ρ^1 are positive, and

---

[19]We need estimates of the hedonic model price functions for both periods to implement these "observable" indices.

[20]The result follows using Debreu's (1952, pp. 889–90; 1959, p. 19) Maximum Theorem.

- The sets of feasible characteristics Z^0 and Z^1 are convex, closed, and bounded.

22.47   A theoretical output price index has been defined that is bounded by two observable indices. It is natural to take a symmetric mean of the bounds to obtain a best single number that will approximate the theoretical index. Thus, let $m(a,b)$ be a symmetric homogeneous mean of the two positive numbers a and b. We want to find a best $m(P_{HL},P_{HP})$. If we want the resulting index, $m(P_{HL},P_{HP})$, to satisfy the time reversal test, then we can adapt the argument of Diewert (1997, p. 138) and show that the resulting $m(a,b)$ must be the geometric mean, $a^{1/2}b^{1/2}$. Thus, a good candidate to best approximate a theoretical hedonic output price index is the following observable *Fisher hedonic output price index*:

$$(22.25)\ P_{HF} \equiv [P_{HL}P_{HP}]^{1/2}$$
$$= [\rho^1 f^1(z^0) / \rho^0 f^0(z^0)]^{1/2}[\rho^1 f^1(z^1) / \rho^0 f^0(z^1)]^{1/2} \quad \text{using equations (22.15) and (22.121)}$$
$$= [\rho^1 / \rho^0][f^1(z^0) / f^0(z^0)]^{1/2}[f^1(z^1)/f^0(z^1)]^{1/2} .$$

Note that P_{HF} reduces to ρ^1 / ρ^0 if $f^0 = f^1$; that is, if the hedonic model price functions are identical for each of the two periods under consideration, except for the proportional factors, ρ^1 and ρ^0.

22.48   Instead of using equations (22.15) and (22.17) in the first line of equation (22.7), equations (22.17) and (22.20) can be used.   The resulting formula for the Fisher hedonic output price index is

$$(22.26)\ P_{HF} \equiv [P_{HL}P_{HP}]^{1/2}$$
$$= \{[P^1/f^1(z^1)] / [P^0/f^1(z^0)]\}^{1/2}\{[P^1/f^0(z^1)] / [P^0/f^0(z^0)]\}^{1/2}.$$

Equation (22.26) is preferred. It is the geometric mean of two sets of quality-adjusted model price ratios, using the hedonic regression in each of the two periods to do one of the quality adjustments.

22.49   The above theory, for the quality adjustment of establishment output prices, is not perfect. It has two weak parts:

- Using a convex combination of the two reference period technologies may not appeal to everyone, and

- Our technique for converting the bounds to a single number is only one method out of many.

22.50   The initial Laspeyres-type bounds and Paasche-type bounds formalizes the bounds outlined in Section C.5 below and referred to in Section C.2. The quality adjustments in equations (22.13) and (22.14) will be seen from this approach, to be made using the user's model valuation functions, $f^0(z)$ and $f^1(z)$. Producers' costs or production functions enter into the quality adjustment only to determine z^0 and z^1; that is, only to determine which models the establishment will produce. Hence, establishments that have different technologies, primary inputs, or face different input prices will in general choose to produce different

models in the same period. The choice problem has only been modeled here facing a single establishment, although the generalization should be straightforward.

B.7   Markups and imperfect competition

22.51   In Section B.5 it was shown there was some ambiguity in the interpretation of hedonic coefficients. A user-value or resource-cost interpretation was possible if there was uniformity in buyer's tastes or suppliers' technologies, respectively. In Section B.6 an assumption of price-taking behavior on the part of firms was introduced and a formal setting given to a user value interpretation, albeit involving some restrictive assumptions. Yet the approaches in Sections B.5 and B.6 both assume perfectly competitive behavior, and the discussion extends now to the effects of markups in imperfect competition. Feenstra (1995) notes that in imperfect competition, when pricing is above marginal cost, the hedonic function should include a term for the price-cost markup.

22.52   Pakes (2001) has developed the argument focusing on the study of new products as the result of prior investments in product development and marketing. A competitive marginal cost-pricing assumption would require that either (i) products with identical characteristics are developed from such investments, so that the law of one price for these identical products will eliminate any margin, or (ii) all products lose their investment (markup) in the new products. Neither of these is reasonable. Indeed, varying markups are a feature of differentiated products (see Feenstra and Levinsohn, 1995, for example). Pakes (2001) argued that markups should change over time. When new products are introduced, the improvements, and associated markups, are directed to characteristics where markups have previously been high. The markups on existing products with these characteristics will fall, and hedonic coefficients will thus, change over time. Pakes (2001) also argued that there may be an ambiguity as to the signs of the coefficients—that there is no economic reason to expect a positive relationship between price and a desirable characteristic. Such a conclusion would be at odds with a resource-cost or user-value approach. If the characteristics being compared are *vertical*—that is, they are characteristics, of which everyone would like more—then we can expect the sign to be positive. However, Pakes (2001) has argued that the sign on *horizontal* characteristics—that is, for which the ordering of the desirable amounts of characteristics is not the same for all consumers—can be negative. The entry of new products aimed at some segments of the market may drive down the markup on products with more desirable attributes. For example, some consumers may have a preference for television sets with smaller screen sizes and be willing to pay a premium price. Indeed, the required technology for the production of these sets may have required increased investment and thus, increased expected markups. It may be that the quality of the picture on these sets is such that it drives down the price of large-sized sets, resulting in an inverse relationship between price and screen size, where the latter is taken as one variable over the full range of screen sizes. Prior (to the modeling) information on the two markets would allow the regression equation to be appropriately specified, with dummy slope and intercepts for the ranges of screen sizes with new and old technologies.

22.53   Pakes (2001) takes the view that no meaning can be attributed to estimated coefficients and predicted values should be used for price comparisons of models of different quality attributes, rather than the individual coefficients. There are many good reasons for

this, as discussed in Chapter 8, Section E.4.3 and Section G.2.2, and the Appendix 21.1 to this chapter. Yet, it must be stressed that for vertical characteristics the coefficients may be quite meaningful, and even for horizontal characteristics or new characteristics, embodied with the latest research and development, some sense can be made by recourse to the above considerations. But again, theory does not support any easy answer to the interpretation of the coefficients from hedonic regressions. Their relevance is that they emanate from market data, from the often complex interaction of demand and supply and strategic pricing decisions. That theory warns us not to give simplistic interpretations to such coefficients, and allows an understanding of the factors underlying them, is a strength of theory. Yet they remain and are generally regarded (Shultze and Mackie, 2002) as the most promising objective basis for estimating the marginal value of quality dimensions of products, even though a purist interpretation is beyond their capability. [21]

Hedonic Indices

C.1 The need for such indices

22.54 In Section A it was noted that hedonic functions are required for two purposes with regard to a quality adjustment. The first is when an item is no longer produced and the replacement item, whose price is used to continue the series, is of a different quality from the original price basis. The differences in quality can be established in terms of different values of a subset of the z price-determining variables. The coefficients from the hedonic regressions, as estimates of the monetary value of additional units of each quality component z, can then be used to adjust the price of the old item so that it is comparable with the price of the new[22]—so that, again, like is compared with like. This process could be described as "patching," in that an adjustment is needed to the price of the old (or new replacement) series for the quality differences, to enable the new series to be patched onto the old. A second use of hedonic functions referred to in Section A is for estimating *hedonic indices*. These are suitable when the pace and scale of replacements of items is substantial and an extensive use of patching might (i) lead to extensive errors if there were some error or bias in the quality-adjustment process and (ii) lead to sampling from a biased replacement universe as outlined in Section A. Hedonic indices use data in each period from a sample of items that should include those with substantial share of sales revenue—sampling in each period from the double universe. There is no need to establish a price basis and for respondents to keep quoting prices from that basis. What is required are samples of items to be redrawn in each month along with information on their prices, characteristics z_i, and, possibly, quantities or values. The identification of multiple characteristics in the hedonic regressions controls for quality differences, as opposed to the matching of price quotes on the same price basis by the

---

[21]Diewert (2002f) goes further in suggesting positive sign restrictions should be imposed on the coefficients in the econometric estimation, particularly when the hedonic regression is being used to adjust the price of a replacement item in order to make it comparable with the price of an item that has disappeared.

[22]Various mechanisms for such adjustments are varied, as outlined in Chapter 8, Section E.4.3, and Triplett (2004). They include using the coefficients from the salient set of characteristics or using the predicted values from the regression as a whole and, in either case, making the adjustment to the old for comparison with the new, or to the new for comparison with the old, or some effective average of the two.

respondents. A number of procedures for estimating hedonic indices are briefly considered below.

C.2 Theoretical characteristics price indices

22.55 In Chapter 18 theoretical output price indices were defined and practical index number formulas considered as estimates of these indices. Theoretical output index numbers are defined here not just on the goods produced, but also on their characteristics. $R(p,S(v))$ was defined in Chapter 18 as the maximum value of output that the establishment can produce, given that it faces the vector of output prices p and given that the vector of inputs v (using technology S) is available for use. The establishment's *output price index P* between any two periods, say period 0 and period 1, was defined as

$$(22.27)\ P(p^0,p^1,v) = R(p^1, S(v)) / R(p^0, S(v)),$$

where p^0 and p^1 are the vectors of output prices that the establishment faces in periods 0 and 1, respectively, and $S(v)$ is a constant reference vector of technology using v intermediate and primary inputs.[23] For theoretical indices in characteristic space, the revenue functions are *also* defined over goods made up of tied bundles of characteristics, hedonic commodities, represented by the hedonic function[24]

$$(22.28)\ P(p^0,p^1,v, z^0, z^1) = \frac{R(p^1, p(z^1), S(v))}{R(p^0, p(z^0), S(v))}.$$

Note that the establishment faces prices p^1 for regular outputs and, from (22.6), the entire hedonic schedule of prices, $p^1(z)=\rho^1 f^1(z)$ for the hedonic commodity, and similarly for period 0. This schedule is a user valuation schedule and hence is exogenous to the establishment. The establishment then decides which model to produce in the light of this schedule. Extending the framework in B.6 to include regular and hedonic commodities (22.28) is an extension of definition (22.12), where the extension is that the period $t=0,1$ establishment production function is now $q^t = F^t(q,z,v)$, in place of the old (22.11): $q^t = F^t(z,v)$, where q^t is the hedonic commodity, q is a vector of "regular" commodities, z is the vector of characteristics for the hedonic commodity, v is a vector of inputs, and F^t the production function).

---

[23]This concept of the output price index (or a closely related variant) was defined by F. M.Fisher and Shell (1972, pp. 56–58), Samuelson and Swamy (1974, pp. 588–92), Archibald (1977, pp. 60–61), Diewert (1980, pp. 460–61; 1983, p. 1055), and Balk (1998b, pp. 83–89). Readers who are familiar with the theory of the true cost of living index will note that the output price index defined by equation (17.2) is analogous to the true cost of living index which is a ratio of cost functions, say $C(u,p1)/C(u,p0)$, where u is a reference utility level: R replaces C, and the reference utility level u is replaced by the vector of reference variables $S(v)$. For references to the theory of the true cost of living index, see Konüs (1924), Pollak (1983), or ILO and others (2004), which is the CPI counterpart to this *Manual*.

[24]Triplett (1987) and Diewert (2002d), following Pollak (1975), consider a two-stage budgeting process whereby that portion of utility concerned with items defined as characteristics has its theoretical index defined in terms of a cost-minimizing selection of characteristics, conditioned on an optimum output level for composite and hedonic commodities. These quantities are then fed back into the second-stage overall revenue maximization.

22.56 The output price index defined by equation (22.28) is a ratio of hypothetical revenues that the establishment could realize, with a given technology and vector of inputs v to work with. Equation (22.28) incorporates substitution effects: if the prices of some characteristics increase more than others, then the revenue-maximizing establishment can switch its output mix of characteristics in favor of such characteristics. The numerator in equation (22.28) is the maximum revenue that the establishment could attain if it faced the output prices and implicit hedonic shadow prices of period 1, p^1 and $p(z^1)$, while the denominator in equation (22.28) is the maximum revenue that the establishment could attain if it faced the output and characteristic's prices of period 0, p^0 and $p(z^0)$. Note that all the variables in the numerator and denominator functions are exactly the same, except that the output price and characteristics price vectors differ. This is a defining characteristic of an output price index: the technology and inputs are held constant. As with the economic indices in Chapter 16, there is an entire *family* of indices depending on which reference technology and reference input vector v that is chosen. In Section C.5 some explicit formulations will be considered including a base-period 0 reference technology and inputs and a current-period 1 reference technology and inputs analogous to the derivation of Laspeyres and Paasche in Chapter 18, Section B.1. Before considering such hedonic indices in Section C.5, two simpler formulations are first considered in Section s C.3 and C.4: hedonic regressions using dummy variables on time and period-on-period hedonic indices. They are simpler and widely used because they require no information on quantities or weights. Yet, their interpretation from economic theory is therefore more limited. However, as will be shown, weighted formulations are possible using a weighted least squares (WLS) estimator, although they are first considered in their unweighted form.

C.3 Hedonic regressions and dummy variables on time

22.57 Let there be K characteristics of a product, and let model or item i of the product in period t have the vector of characteristics $z_i^t \equiv [z_{i1}^t,...,z_{iK}^t]$ for $i = 1,....,K$ and $t = 1,...,T$. Denote the price of model i in period t by p_i^t. A hedonic regression of the price of model i in period t on its characteristics set z_i^t is given by

$$(22.29) \quad \ln p_i^t = \gamma_0 + \sum_{t=2}^{T}\gamma_t D_t + \sum_{k=1}^{K}\beta_k z_{ik}^t + \varepsilon_i^t ,$$

where D_t are dummy variables for the time periods, D_2 being 1 in period $t = 2$, zero otherwise; D_3 being 1 in period $t = 3$, zero otherwise, and so on. The coefficients γ_t are estimates of quality-adjusted price changes, having controlled for the effects of variation in quality (via $\sum_{k=1}^{K}\gamma_k z_{tki}$)—although see Goldberger (1968) and Teekens and Koerts (1972) for the adjustment for estimation bias.

22.58 The above approach uses the dummy variables on time to compare prices in period 1 with prices in each subsequent period. In doing so, the γ parameters are constrained to be constant over the period $t = 1,...,T$. Such an approach is fine retrospectively, but in real time the index may be estimated as a fixed-base or chained-base formulation. The *fixed-base* formulation would estimate the index for period 1 and 2, $I_{1,2}$, using equation (22.29) for $t = 1$, 2; the index for period 3, $I_{1,3}$, would use equation (22.29) for t = 1, 3; for period 4, $I_{1,4}$, using equation (22.29) for $t = 1$, 4; and so forth. In each case the index constrains the parameters to

be the same over the current and base period. A fixed-base, bilateral comparison using equation (22.29) makes use of the constrained parameter estimates over the two periods of the price comparison. A *chained* formulation would estimate $I_{1,4}$, for example, as the product of a series of links: $I_{1,4} = I_{1,2} \times I_{2,3} \times I_{3,4}$.[25] Each successive binary comparison or link is combined by successive multiplication. The index for each link is estimated using equation (22.24). Because the periods of time being compared are close, it is generally more likely that the constraining of parameters required by chained-time dummy hedonic indices is considered to be less severe than that required of their fixed base counterparts.

22.59 There is no explicit weighting in these formulations, and this is a serious disadvantage. In practice, cut-off sampling might be employed to include only the most important items. If sales data are available, a WLS (weighted by relative sales shares—see Appendix 21.1 and Diewert (2005) estimator should be used instead of an ordinary least squares (OLS) estimator.[26] A WLS estimator is equivalent to replicating the sample in proportion to the weights and applying an OLS estimator.

C.4 Period-on-period hedonic indices

22.60 An alternative approach to comparing prices in period 0 and 1 is to estimate a hedonic regression for period 1 and insert the values of the characteristics of each model existing in period 0 into the period 1 regression to predict, for each item, its price $\hat{p}_i^1(z_i^0)$. This would generate predictions of the price of items existing in period 0, at period 1 shadow prices, $\hat{p}_i^t(z_i^1)$, $i = 1,...,N$. These prices (or an average) can be compared with (the average of) the actual prices of models $i = 1,.....N$ models in period 0. The averages may be arithmetic, as in a Dutot index, or geometric, as in a Jevons index. The arithmetic formulation is defined as follows:

$$(22.30a) \quad \frac{\sum_{i=1}^{N} (1/N)\hat{p}_i^1(z_i^0)}{\sum_{i=1}^{N} (1/N)p_i^0(z_i^0)}$$

22.61 Alternatively, the characteristics of models existing in period 1 can be inserted into a regression for period 0. Predicted prices of period 1 items generated at period 0 shadow prices (or an average) can be compared with (the average of) the actual prices in period t:

$$(22.30b) \quad \frac{\sum_{i=1}^{N} (1/N)p_i^1(z_i^1)}{\sum_{i=1}^{N} (1/N)\hat{p}_i^0(z_i^1)} \, .$$

---

[25]Chapter 16, Section F contains a detailed account of chained indices.
[26]Ioannidis and Silver (1999) and Bode and van Dalen (2001) compared the results from these different estimators, finding notable differences, but not in all cases (see also Heravi and Silver, 2007).

22.62   For a fixed-base bilateral comparison using either equation (22.30a) or (22.30b), the hedonic equation need be estimated only for one period. The denominator in equation (22.30a) is the average observed price in period 0, which should be equal to the average price a hedonic regression based on period 0 data will predict using period 0 characteristics. The numerator, however, requires an estimated hedonic regression to predict period 0 characteristics at period 1 hedonic prices. Similarly, in equation (22.30b), a hedonic regression is required only for the denominator. For reasons analogous to those explained in Chapters 16, 17, and 18, a symmetric average of these indices should have some theoretical support.

22.63   Note that all the indices described in Section sC.1 and C.2 use all the data available in each period. If there is a new item, for example, in period 4, it is included in the data set and its quality differences controlled for by the regression. Similarly, if old items drop out, they are still included in the indices in the periods in which they exist. This is part of the natural estimation procedure, unlike using matched data and hedonic adjustments on noncomparable replacements when items are no longer produced.

22.64   As with the dummy variable approach, there is no need for matched data . Yet there is also no explicit weighting in these formulations and this is a serious disadvantage. Were data on quantities or values available, it is immediately apparent that such weights could be attached to the individual $i = 1,….N$ prices or their estimates. This is considered in the next section.

C.5   Superlative and exact hedonic indices

22.65   In Chapter 18 Laspeyres and Paasche bounds were defined on a theoretical basis, as were superlative indices, which treat both periods symmetrically. These superlative formulas, in particular the Fisher index, were also seen in Chapter 16 to have desirable axiomatic properties. Furthermore, the Fisher index was supported from economic theory as a symmetric average of the Laspeyres and Paasche bounds and was found to be the most suitable such average of the two on axiomatic grounds. The Törnqvist index seemed to be best from the stochastic viewpoint and also did not require strong assumptions for its derivation from the economic approach as a superlative index. The Laspeyres and Paasche indices were found to correspond to (be *exact* for) underlying (Leontief) aggregator functions with no substitution possibilities while superlative indices were exact for flexible functional forms including the quadratic and translog forms for the Fisher and Törnqvist indices, respectively. If data on prices, characteristics, *and quantities* are available, analogous approaches and findings arise for hedonic indices (see Fixler and Zieschang, 1992, and Feenstra, 1995). Exact bounds on such an index were defined by Feenstra (1995). Consider the theoretical index in equation (22.28), but now defined only over items in terms of their characteristics. The prices are still of items, but they are wholly defined through $p(z)$. An arithmetic aggregation for a linear hedonic equation finds a Laspeyres lower bound (as quantities supplied are *increased* with increasing relative prices) is given by

$$(22.31a) \quad \frac{R(p(z)_1, S(v)_0)}{R(p(z)_0, S(v)_0)} \geq \frac{\sum\limits_{i=1}^{N} x_{i0}\hat{p}_{i1}}{\sum\limits_{i=1}^{N} x_{i0}p_{i0}} = \sum_{i=1}^{N} s_{i,0}\left(\frac{\hat{p}_{i1}}{p_{i0}}\right),$$

where $R(.)$ denotes the revenue at a set of output prices, input quantities, v, and technology, S, following the fixed input-output price index model. The price comparison is evaluated at a fixed level of period 0technology and inputs. s_{i0} are the shares in total value of output of product i in period 0, where $s_{i0} = x_{i0}p_{i0}\Big/\sum\limits_{i=1}^{N} x_{i0}p_{i0}$, and

$$(22.31b) \quad \hat{p}_{i1} \equiv p_{i1} - \sum_{i=1}^{N} \beta_{k1}(z_{ik1} - z_{ik0})$$

are prices in periods 1 adjusted for the sum of the changes in each quality characteristic weighted by their coefficients derived from a linear hedonic regression. As noted in Appendix 21.1, β_{k1} may be estimated using a weighted least squares estimator where the weights are the sales quantities. The summation is over the same i in both periods, since replacements are included when items are missing and equation (22.31b) adjusts their prices for quality differences.

22.66 A Paasche upper bound is estimated as

$$(22.32a) \quad \frac{R(p_1, S(v)_1)}{R(p_0, S(v)_1)} \leq \frac{\sum\limits_{i=1}^{N} x_{i1}\hat{p}_{i1}}{\sum\limits_{i=1}^{N} x_{i1}p_{i0}} = \left[\sum_{i=1}^{N} s'_{i1}\left(\frac{\hat{p}_{i1}}{p_{i0}}\right)\right]^{-1}$$

where $s'_{i1} = x_{i1}\hat{p}_{i1}\Big/\sum\limits_{i=1}^{N} x_{i1}\hat{p}_{i1}$, and

$$(22.32b) \quad \hat{p}_{i0} \equiv p_{i0} + \sum_{i=1}^{N} \beta_{k0}(z_{ik1} - z_{ik0}),$$

which are prices in periods 0adjusted for the sum of the changes in each quality characteristic weighted by its respective coefficients derived from a linear hedonic regression.

22.67 Following from the inequalities Chapter 18 where the Laspeyres P_L and Paasche P_P form bounds on their true, economic theoretic indexes,

$$(22.33) \quad P_L \leq P(p^0, p^1, \alpha) \leq P_P \text{ or } P_L \leq P(p^0, p^1, \alpha) \geq P_P$$

22.68 The superlative and exact hedonic index (SEHI) approach thus, first applies the coefficients from hedonic regressions to changes in the characteristics to adjust observed prices for quality changes equations (22.31b and 21.32b). Second, it incorporates a weighting system using data on the value of output of each model and its characteristics, rather than treating each model as equally important equations (22.31a and 21.32a). Finally, it has a direct correspondence to formulation defined from economic theory.

22.69 Semilogarithmic hedonic regressions would supply a set of β coefficients suitable for use with these base-period and current-period geometric bounds:

$$(22.34a) \quad \prod_{i=1}^{N} \left(\frac{p_{i1}}{\hat{p}_{i0}} \right)^{S_{i1}} \geq \frac{R(p(z)_1, q, T)}{R(p(z)_0, q, T)} \geq \prod_{i=1}^{N} \left(\frac{\hat{p}_{i1}}{p_{i0}} \right)^{S_{i0}}$$

$$(22.34b) \quad \hat{p}_{i0} \equiv p_{i0} \exp[\sum_{i=1}^{N} \beta_{k0}(z_{ik1} - z_{ik0})]$$

$$\hat{p}_{i1} \equiv p_{i1} \exp[-\sum_{i=1}^{N} \beta_{k1}(z_{ik1} - z_{ik0})].$$

22.70 In equation (22.34a) the two bounds on their respective theoretical indices have been shown to be brought together. The calculation of such indices is no small task. For examples see Silver and Heravi (2001a; 2003) and Chapter 8, Section G.2, for comparisons over time, Kokoski, Moulton, and Zieschang (1999) for price comparisons across areas of a country, and Heravi, Heston and Silver (2005a) for comparisons across countries.

22.71 The methods outlined above are outlined to show how practical hedonic indices relate to theoretical counterparts. There are many more variants of such practical formulas some of which are outlined in Chapter 8. Their nature depends on the approach adopted, time dummy or period-on-period indices, whether the indices are fixed base or chained, whether geometric, arithmetic or harmonic aggregators are used, and whether base period, current period, or some average of the two period's weights are used. Heravi and Silver (2006a) in a meta analysis of the results of a whole variety of such measures explore such differences.

C6. The difference between the period on period and time dummy approaches

22.72 The dummy variable method outlined in section C3 and the period-on-period hedonic indexes, outlined in sections C4 and C5—also referred to as "hedonic imputation indexes" by Heravi and Silver (2007) and as "characteristic price index numbers" by Triplett (2004) —not only correct price changes for changes in the quality of items purchased, but also allow the indexes to incorporate matched and unmatched models. They provide a means by which price changes can be measured in product markets where there is a rapid turnover of differentiated models. However, they can yield quite different results. Diewert, Heravi, and Silver (2007) provide a formal exposition of the factors underlying such differences and the implications for choice of method. This was undertaken for the Törnqvist index, but the

analysis can be readily extended to other formulas. They found that differences between the two approaches may arise from both parameter instability over the two periods compared and changes over the two periods compared in the characteristics of the models sold, and that such differences are compounded when both such changes occur. They further showed that similarities between the two approaches resulted if there was little difference in either component change.

22.73 The above in section C has illustrated how weighted index number formulas might be constructed using data on prices, quantities, and characteristics for an item when the data are not matched. But for analytical purposes it is useful to decompose price changes into that due to matched price changes, that due to unmatched new models introduced, and that due to unmatched old models that are retired. The analysis is useful for determining the bias in just using matched models.

C.7   Decomposing price changes into matched and unmatched components

Following Silver and Heravi (2005b) the hedonic formulation in (22.29) is used to derive the basic matched model result for hedonic time dummy indexes over two periods, due originally to Triplett and McDonald (1977). However, we reformulate (22.29) as

(22.35) **Error! Objects cannot be created from editing field codes.**
$$m \in S(t); t = 1,2,...,T$$

where $S(t)$ is the set of models available in period t, p_{tm} is the period t price of model m, D_t is a time dummy variable that is 1 if the left hand side observation is the log of a period t price and is 0 otherwise, z_{mtk} is the amount of characteristic k model m in period t possesses, and ε_{tm} is an error term. Let the number of models available in period t be $N(t)$; i.e. there are $N(t)$ models in the set $S(t)$ for each t. The coefficients α_t and β_k are typically estimated using least squares. It should be mentioned that there is no constant term in (22.351); rather, there is a time dummy for every period. It is straightforward to show that this specification is equivalent to the usual hedonic model with time dummies that has a constant term. α_t is an estimate of the (logarithm of the) average price of models in period t having controlled for the z_{mtk} characteristics (though see Teekens and Koerts (1972) for an adjustment). Consider a special case of the general equation (22.35) in which there are only two periods so $T=2$ and assume that the models are matched in each of the two periods so that $S(1) = S(2)$ and **Error! Objects cannot be created from editing field codes.**so that the same M models are available in each period. Hence the model characteristics are the same in each, i.e. we have:

(22.36) $z_{mtk} = z_{mk}$ say, for $t =1,2$, $m=1,…, M$ and $k=1,…K$.

With these restrictions the least squares estimates for the unknown parameters in equation (22.35) are denoted by **Error! Objects cannot be created from editing field codes.** and **Error! Objects cannot be created from editing field codes.** and **Error! Objects cannot be created from editing field codes.** for $k=1,…K$.

Define *price levels* for periods 1 and 2, P_1 and P_2 respectively, in terms of the least squares estimates for α_1 and α_2 as follows:

(22.37) **Error! Objects cannot be created from editing field codes.**; **Error! Objects cannot be created from editing field codes.**.

Hence the logarithm of the *price index going from period 1 to 2* is defined as

(22.38) **Error! Objects cannot be created from editing field codes.** .

A property of least squares regression estimates is that the column vector of least squares residuals is orthogonal to each column vector of exogenous variables (this follows a. technique of proof used by Diewert (2001). Using this property for the first two columns of exogenous variables corresponding to the time dummy variables leads to the following two equations (using (22.36) as well):

(22.39) **Error! Objects cannot be created from editing field codes.**

(22.40) **Error! Objects cannot be created from editing field codes.**

Divide both sides of (22.39) and (22.40) by M and solve the resulting equations for the least squares estimates, **Error! Objects cannot be created from editing field codes.** and **Error! Objects cannot be created from editing field codes.**. Substituting these expressions for **Error! Objects cannot be created from editing field codes.** and **Error! Objects cannot be created from editing field codes.** into (22.38) leads to the following formula for the log of the *hedonic price index*:

(22.41) **Error! Objects cannot be created from editing field codes.**.

Exponentiating both sides of (22.41) shows that the hedonic model price index going from period 1 to 2 under the above matched model conditions is equal to *the equally weighted geometric mean of the M model price relatives*, which would be a conventional matched model statistical agency estimate of the price index for this elementary group of commodities.

Now let us relax the matched model restriction, but still assume that $T=2$, i.e., that there are only two periods in the hedonic regression model defined by (22.35). Some additional notation is required in order to model this case. Define the following sets of models:

(22.42) $S(1 \cap 2) \equiv S(1) \cap S(2)$;

(22.43) $S(1 \neg 2) \equiv S(1) \neg S(2)$;

(22.44) $S(2 \neg 1) \equiv S(2) \neg S(1)$;

Thus $S(1 \cap 2)$ is the set of models that are present in both periods 1 and 2, $S(1 \neg 2)$ is the set of models that are present in period 1 but not period 2 and $S(2 \neg 1)$ is the set of models that are present in period 2 but not period 1. Let the number of models in the sets $S(1 \cap 2)$, $S(1 \neg 2)$ and $S(2 \neg 1)$ be denoted by $N(1 \cap 2)$, $N(1 \neg 2)$ and $N(2 \neg 1)$ respectively. Relating our new notation to the total number of models in periods 1 and 2, $N(1)$ and $N(2)$ respectively, it can be seen that:

(22.45) $N(1) = N(1 \cap 2) + N(1 \neg 2)$ and

(22.46) $N(2) = N(1 \cap 2) + N(2 \neg 1)$

The least squares estimates for the equation defined by (22.35) when *T*=2 can now be obtained. Again recalling that the column vector of least squares residuals is orthogonal to each column vector of exogenous variables, we obtain the following two equations, where this orthogonality property was used for the first two columns of the exogenous variables corresponding to the time dummy variables:

(22.47) **Error! Objects cannot be created from editing field codes.**

(22.48) **Error! Objects cannot be created from editing field codes.**

If equations (22.47) and (22.48) are divided by the number of common models in the two periods, $N(1 \cap 2)$, expressions for **Error! Objects cannot be created from editing field codes.** and **Error! Objects cannot be created from editing field codes.** can be obtained. Substituting these expressions into (22.38) and using $z_{m1k} = z_{m2k}$ for the common models $m \in S(1 \cap 2)$ leads to the following formula for the log of the *hedonic price index*:

(22.49) **Error! Objects cannot be created from editing field codes.**

Error! Objects cannot be created from editing field codes.

Error! Objects cannot be created from editing field codes.

22.74 The first set of terms on the right hand side of (22.49) is the *matched model contribution* to the overall index, **Error! Objects cannot be created from editing field codes.**. The next two set of terms are respectively the change in price due to unmatched models existing in period 2, but not in 1, and unmatched models existing in period 1 but not in 2. These expressions are not captured in a matched models index. If the second set of terms, **Error! Objects cannot be created from editing field codes.**, is positive, then the matched model price index is too low and must be adjusted upward. Consider a new model *m* introduced in period 2. If (the logarithm of) its price $(ln\, p_{m2})$ is above that predicted from a period 2 hedonic regression $\left(\sum_{k=1}^{K} z_{m2k} \beta_k^* - \alpha_2^* \right)$, then this will raise the overall price index and a matched model index would be too low if it ignored such new models (Triplett and McDonald (1977) have a similar interpretation). Similarly, consider the last set of terms in (22.49) and an unmatched old model, introduced in period 1 but no longer available in period 2. If it was priced in period 1 above its period 1 predicted price then the matched model price index would be too high (note the negative sign). The extent and nature of the bias depends on the pricing strategy of new and old models.

22.75 The extent of any difference depends, in this unweighted formulation, on the proportions of old and new items leaving and entering the sample and on the price changes of old and new items relative to those of matched items. If the market for commodities is one in which old quality-adjusted prices are unusually low while new quality-adjusted prices are unusually high, then the matched index will understate price changes (see Silver and Heravi, 2005b, and Berndt, Ling, and Kyle, 2003, for examples). Different market behavior will lead to different forms of bias. The above expression is for unweighted price changes, but the principles extend to similar findings for weighted price changes and, by association, weighted index numbers, as shown in Silver and Heravi (2005b). As noted in the Appendix

to this chapter, and argued in Diewert (2005), different weighting systems in a weighted least squares hedonic regression correspond to different index number formula.

New Goods and Services

22.76 This section briefly highlights issues relating to the incorporation of new goods into the index. Practical issues were outlined in Chapter 9.D.3. The term new goods will be used here to refer to those that provide a substantial and substantive change in what is provided, as opposed to more of a currently available set of service flows, such as a new model of an automobile that has a bigger engine. In this latter instance, there is a continuation of a service and production flow, and this may be linked to the service flow and production technology of the existing model. The practical concern with the definition of new goods as against quality changes is that the former cannot be easily linked to existing items as a continuation of an existing resource base and service flow, because of the very nature of their "newness." There are alternative definitions; Oi (1997) directs the problem of defining new goods to that of defining a monopoly. If there is no close substitute, the good is new. A monopoly supplier may be able to supply an item with new combinations of the hedonic z characteristics because of a new technology and have a monopoly power in doing so, but in practice the new good can be linked via the hedonic characteristics set to the existing ones. In this practical sense, such goods are not considered new for the purposes of the *Manual*.

22.77 Merkel (2000, p. 6) takes a similar practical line in devising a classification scheme that will meet the practical needs of XMPI compilation. He considers *evolutionary* and *revolutionary* goods. The former are defined as

> …extensions of existing goods. From a production inputs standpoint, evolutionary goods are similar to pre-existing goods. They are typically produced on the same production line and/or use largely the same production inputs and processes as pre-existing goods. Consequently, in theory at least, it should be possible to quality adjust for any differences between a pre-existing good and an evolutionary good.

In contrast, revolutionary goods are goods that are substantially different from pre-existing goods. They are generally produced on entirely new production lines or with substantially new production inputs and processes than those used to produce preexisting goods. These differences make it virtually impossible, both from a theoretical and practical standpoint, to quality adjust between a revolutionary good and any preexisting good. The main concern regarding the incorporation of new goods into the XMPIs is the decision on the need and timing for their inclusion. Waiting for a new good to be established or waiting for the rebasing of an index before incorporating new products may lead to errors in the measurement of price changes if the unusual price movements at critical stages in the product life cycles are ignored. There are practical approaches to the early adoption of both evolutionary and revolutionary goods. These are outlined in Chapter 9, Section D.3. For evolutionary goods, such strategies include the rebasing of the index, resampling of items and introduction of new goods as directed *sample substitutions* (Merkel, 2000). Also of use are hedonic quality adjustments and indices outlined in Chapter 8, Section E.4, and Section C above that facilitate the incorporation of such evolutionary goods, since they possess a similar characteristics set to existing ones but deliver different quantities of these

characteristics. The modified short-run or chained framework outlined in Chapter 8, Section H–G may also be more appropriate for product areas with high turnover of items. These approaches can incorporate the price change of new goods into the index as soon as prices are available for two successive periods, although issues relating to the proper weighting of such changes may remain.

22.78   However, for revolutionary goods, substitution may not be appropriate. First, they may not be able to be defined within the existing classification systems. Second, they may be primarily produced by a new establishment, which will require extending the sample to such establishments. Third, there will be no previous items to match against and make a quality adjustment to prices, since by definition, they are substantially different from preexisting goods. And, finally, there is no weight to attach to the new establishment or item(s). *Sample augmentation* is appropriate for revolutionary goods, as opposed to sample substitution for evolutionary goods. It is necessary to bring the new revolutionary goods into the sample in addition to what exists. This may involve extending the classification, the sample of establishments, and item list within new or existing establishments (Merkel, 2000).

Appendix 21.1: Some Econometric Issues

22.79   Hedonic regression estimates will have been seen in Chapter 8 to have potential use for the quality adjustment of prices. There are a number of issues arise from the specification and estimation of hedonic regressions, the use of diagnostic statistics, and courses of action when the standard OLS assumptions are seen to break down. Many of these issues are standard econometric ones and not the subject of this *Manual*. This is not to say they are unimportant. The use of hedonic regressions will require some econometric or statistical expertise, but suitable texts are generally available. See Berndt (1991)—particularly the chapter on hedonic regressions—and Maddala (1988) and Kennedy (2003), among many others. Modern statistical and econometric software have adequate diagnostic tests for testing when OLS assumptions break down. There remain, however, some specific issues that merit attention, although it must be stressed that these points are over and above, and should not be taken to diminish, the important standard econometric issues found in econometric texts.

Identification and appropriate estimators

22.80   Wooldridge (1996, pp. 400–01) has shown on standard econometric grounds that the estimation of supply and demand functions by OLS is biased *and this bias carries over to the estimation of the hedonic function.* It is first useful to consider estimation issues in the supply and demand functions. These functions are rarely estimated in practice. The more common approach is to estimate offer functions, with the marginal price offered by the firm dependent upon chosen attributes (product characteristics) and firm characteristics, and to estimate *bid* or value functions, with the marginal prices paid by a consumer dependent on chosen attributes and consumer characteristics.[27] As noted earlier, the observed prices and quantities are the result of the interaction of structural demand and supply equations and the

---

[27]These are equivalent to inverse demand (supply) functions, with the prices dependent upon the quantities demanded (supplied) and the individual consumer (producer) characteristics.

distributions of producer technologies and consumer tastes, they and cannot reveal the parameters of these offer and value functions. Rosen (1974, pp. 50–51) suggested a procedure for determining these parameters. Since these estimates are conditioned on tastes (α) and technologies (τ), the estimation procedure needs to include empirical measures or "proxy variables" of α and τ. For the tastes α of consumers, the empirical counterparts may be sociodemographic and economic variables, which may include age, income, prices and quantities of nonhedonic commodities demanded by households,[28] education, and geographical region. For technologies τ, variables may include technologies and factor prices. First, the hedonic equation is estimated without these variables in the normal manner using the best-fitting functional form. This is to represent the price function consumers and producers face when making their decisions. Then, an implicit marginal price function is computed for each characteristic as $\partial p(z)/\partial z_i = \hat{p}_i(z)$, where $\hat{p}(z)$ is the estimated hedonic equation. Bear in mind that in normal demand and supply studies for *products*, the prices are observed in the market. For *characteristics* they are unobserved, and this first stage must be to estimate the parameters from the hedonic regression. The actual values of each z_i bought and sold is then inserted into each implicit marginal price function to yield a numerical value for each characteristic. These marginal values are used in the second stage[29] of estimation as endogenous variables for the estimation of the demand side:

(A21.1) $\hat{p}_i(z) = G(z1,\ldots,zK, \alpha^*)$,

where α^* are the proxy variables for tastes.

The supply side estimating equations might look like:

(A21.2) $\hat{p}_i(z) = G^*(z1,\ldots,zK, \tau^*)$,

where τ^* are the proxy variables for technologies.

The variables τ^* drop out when there is no variation in technologies and $\hat{p}_i(z)$ is an estimate of the offer function. Similarly the variables α^* drop out when sellers differ and buyers are identical and cross-section estimates trace out compensated demand functions.

22.81 Epple (1987) has argued that Rosen's modeling strategy is likely to give rise to inappropriate estimation procedures of the demand and supply parameters. The hedonic approach to estimating the demand for characteristics has a difficulty arising from the fact that marginal prices are likely to be endogenous—they depend on the amount of each characteristic consumed and must be estimated from the hedonic function rather than observed directly. There are two resulting problems. First, there is an identification problem (see Epple, 1987) because both the marginal price of a characteristic and the inverse bid

---

[28] The consumer theory approach used by Diewert (2003) to deriving the hedonic function rested on rather strong separability assumptions on consumer preferences. Once these separability assumptions are relaxed, the demand for nonhedonic commodities will provide a means for identification of the hedonic preferences.
[29] This two-stage approach is common in the literature, though Wooldridge (1996) discusses the joint estimation of the hedonic and demand and supply side functions as a system.

depend on the levels of characteristics consumed. Second, if important characteristics are unmeasured and they are correlated with measured characteristics, the coefficients on measured characteristics will be biased. This applies to all econometric models, but it is particularly relevant to hedonic models; on this point see Wooldridge (1996, 400–01). The equilibrium conditions for characteristic prices imply functional relationships among the characteristics of demanders, suppliers, and products. This in turn reduces the likelihood that important excluded variables will be uncorrelated with the included variables of the model (see also Bartik, 1988, on this point). The bias arises because buyers are differentiated by characteristics (y,α) and sellers by technologies τ. The type of item buyers will purchase is related to (y,α) and the type sellers provide to τ. On the plane of combinations of z transacted, the equilibrium ones chosen may be systematically related; the characteristics of buyers are related to those of sellers. Epple (1987) uses the example of stereo equipment: the higher income of some buyers leads to purchases of high-quality equipment and the technical competence of sellers leads them to provide it. The consumer and producer characteristics may be correlated.

22.82 Wooldridge (1996, pp. 400–01) suggests that individual consumer and firm characteristics such as income, education, and input prices should be used as instruments in estimating hedonic functions. In addition, variables other than a good's characteristics should be included as instruments if they are price determining, such as geographical location—say proximity to ports, good road systems, climate, and so on. Communities of economic agents are assumed, within which consumers consume and producers produce for each other at prices that vary across communities for identical goods. Variables on the characteristics of the communities will not in themselves enter the demand and supply equation but are price determining for observed prices recorded across communities. Tauchen and Witte (2001) provide a systematic investigation of the conditions under which consumer and producer and community characteristics will affect the hedonic parameter estimates for a single-regression equation estimated across all communities. A key concern is whether the hedonic price function error term represents factors that are unobserved by both the economic agents and the researcher, or by the researcher only. In the latter case the error term may be correlated with the product attributes and instrumental variable estimation is required. If the error term is *not* correlated with the product characteristics—preferences are quasi-linear—then a properly specified hedonic regression, including community-specific characteristics or appropriate slope dummies, can be estimated using OLS. In other cases, depending on the correlation between consume and producer characteristics, assumptions about the error term and the method of incorporating community characteristics into the regression, instrumental variables, including consumer or producer or community dummy or characteristics, may need to be used.

Functional form

22.83   Triplett (1987; 2002) argues that neither classical utility theory nor production theory can specify the functional form of the hedonic function.[30] This point dates back to Rosen (1974, p. 54) who describes the observations as being "..a joint-envelope function and cannot by themselves identify the structure of consumer preferences and producer technologies that generate them." A priori judgments about what the form should look like may be based on ideas about how consumers and production technologies respond to price changes. These judgments are difficult to make when the observations are jointly determined by demand and supply factors but not impossible in rare instances. However, it is complicated when pricing is with a markup, the extent of which may vary over the life cycle of a product. Some tied combinations of characteristics will have higher markups than others. New item introductions are likely to be attracted to these areas of characteristic space, and this will have the effect of increasing supply and thus, lowering the markup and price (Cockburn and Anis, 1998; Feenstra, 1995, p. 647; and Triplett, 1987, p. 38). This again must be taken into account in any a priori reasoning—not an easy or straightforward matter.

22.84   It may be that in some cases the hedonic function's functional form will be very straightforward. For example, prices on the websites for options for products are often additive. The underlying cost and utility structure are unlikely to jointly generate such linear functions, but the producer or consumer are also paying for the convenience of selling in this way and are willing to bear losses or make gains if the cost or utility at higher values of z are priced lower/worth more than the price set. But, in general, the data should convey what the functional form should look like, and imposing artificial structures simply leads to specification bias. For examples of econometric testing of hedonic functional form, see Cassel and Mendelsohn (1985); Cropper, Deck, and McConnell (1988),\; Rasmussen and Zuehlke (1990); Bode and van Dalen (2001); and Curry, Morgan, and Silver (2001).

22.85   The three forms prevalent in the literature are linear, semilogarithmic, and double-logarithmic (log-log). A number of studies have used econometric tests, in the absence of a clear theoretical statement, to choose between them. There have been a large number of hedonic studies, and, as illustrated in Curry, Morgan, and Silver (2001), in many of these the quite simple forms do well, at least in terms of the $\bar{R}^2$ presented, and the parameters accord with a priori reasoning, usually on the consumer side. Of the three popular forms some are favored in testing. For example, Murray and Sarantis (1999) favored the semilogarithmic form, while in others—for example Hoffmann (1998)—the three functional forms were found to scarcely differ in terms of their explanatory power. That the parameters from these simple forms accord with a priori reasoning, usually from the consumer side, is promising, but researchers should be aware that such matters are not assured. Of the three forms, the semilogarithmic form has much to commend it. The interpretation of its coefficients is quite straightforward—the coefficients represent proportionate changes in prices arising from a

---

[30]Arguea, Hsiao, and Taylor (1994) propose a linear form on the basis of arbitrage for characteristics, held to be likely in competitive markets, although Triplett (2004) argues that this is unlikely to be a realistic scenario in most commodity markets.

unit change in the value of the characteristic.[31] This is a useful formulation since quality adjustments are usually undertaken by making multiplicative instead of additive adjustments (see Chapter 8, Section C.3). The semilogarithmic form, unlike the log-log model, can also incorporate dummy variables for characteristics that are either present, $z_i = 1$, or not, $z_i = 0$.[32]

22.86   More complicated forms are possible. Simple forms have the virtue of parsimony and allow more efficient estimates to be made for a given sample. However, parsimony is not something to be achieved at the cost of misspecification bias. First, if the hedonic function is estimated across multiple independent markets, then interaction terms are required (see Mendelsohn, 1984, for fishing sites). Excluding them is tantamount to omitting variables and inappropriately constraining the estimated coefficients of the regression. Tauchen and Witte (2001) have outlined the particular biases that can arise from such omitted variables in hedonic studies. Second, it may be argued that the functional form should correspond to the aggregator for the index—linear for a Laspeyres index, logarithmic for a geometric Laspeyres index, translog for a Törnqvist index, and quadratic for a Fisher index (Chapter 18). However, as Triplett (2002) notes, the purpose of estimating hedonic regressions is to adjust prices for quality differences, and imposing a functional form on the data that is inconsistent with the data might create an error in the quality adjustment procedure. Yet, as Diewert (2003) notes, flexible functional forms encompass these simple forms. The log-log form is a special case of the translog form as in equation 17.11, and the semi-log form being a special case of the semi-log quadratic form as in equation 17.16. If there are a priori reasons to expect interaction terms for specific characteristics, as illustrated in the example in Chapter 8, Section E.4, then these more general forms allow this, and the theory of hedonic functions neither dictates the form of the hedonic form nor restricts it.

Changing tastes and technologies

22.87   The estimates of the coefficients may change over time. Some of this will be attributed to sampling error, especially if multicollinearity is present, as discussed below. But, in other cases, it may be a genuine reflection of changes in tastes and technologies. If a subset of the estimated coefficients from a hedonic regression is to be used to quality adjust a noncomparable replacement price, then the use of estimated out-of-date coefficients from some previous period to adjust the prices of the new replacement model would be inappropriate. There would be a need to update the indices as regularly as the changes

---

[31]It is noted that the anti-log of the OLS-estimated coefficients are not unbiased—the estimation of semilogarithmic functions as transformed linear regressions requires an adjustment to provide minimum-variance unbiased estimates of parameters of the conditional mean. A standard adjustment is to add one-half of the coefficient's squared standard error to the estimated coefficient (Goldberger, 1968, and Teekens and Koerts, 1972).

[32]Diewert (2002f) argues against the linear form on the grounds that, while the hedonic model is linear, the estimation required is of a nonlinear *regression* model, and the semi-log and log-log models are linear *regression* models. He also notes that semi-log form has the disadvantage against the log-log of not being able to impose constraints of constant returns to scale. Diewert (2002d) I think that you call this Diewert (2003) elsewhere; i.e., my NBER paper that started out as a comment on your paper. also argues for the use of nonparametric functional forms and the estimation of linear generalized dummy variable hedonic regression models. This has been take up in Curry, Morgan, and Silver (2001), who use neural networks that are shown to work well, although the variable set required for their estimation has to be relatively small.

demanded.[33] For estimating hedonic indices, the matter is more complicated. The coefficients in a simple dummy time-period model as in Section C.3 now have different estimates of the parameters in each period. Silver (1999), using a simple example, shows how the estimate of quality adjusted price change from such a dummy-variable model requires a reference basket of characteristics. This is apparent for the hedonic imputation indices where separate indices using base-and current-period characteristics are estimated. A symmetric average of such indices is considered appropriate. A hedonic index based on a time dummy variable implicitly constrains the estimated coefficients from the base and current periods to be the same. Diewert (2003) formalizes the problem of choosing the reference characteristics when comparing prices over time when the parameters of the hedonic function may themselves be changing over time. He finds the results of hedonic indices to *not* be invariant to the choice of reference-period characteristic vector set z. The use of a sales (quantity) weighted average vector of characteristics proposed by Silver (1999) is considered, but Diewert notes that over long time periods this may become unrepresentative.[34] Of course, if the dummy-variable approach is used in a chained formulation as outlined in Section C.3, the weighted averages of characteristics remain reasonably up to date, though chaining has its own pros and cons (see Chapter 16). A fixed-base alternative noted by Diewert (2003) is to use a Laspeyres-type comparison with the base-period parameter set, and a Paasche-type current-period index with the current-period parameter set, and take the geometric mean of the two indices for reasons similar to those given in Chapter 18, Section B.3. The resulting Fisher-type index is similar to that given in equation (22.32) proposed by Feenstra (1995).[35] A feature of the time dummy approach in is that it implicitly takes a symmetric average of the coefficients by constraining them to be the same. But what if, as is more likely the case, only base-period hedonic regression coefficients are available? Since hedonic indices based on a symmetric average of the coefficients are desirable, the spread or difference between estimates based on either a current or a reference-period characteristics set is an indication of potential bias, and estimates of such spread may be undertaken retrospectively. If the spread is large, estimates based on the use of a single period's characteristics set, say the current period, should be treated with caution. More regular updating of the hedonic regressions is likely to reduce spread because the periods being compared will be closer and the characteristics of the items in the periods compared more similar.

Weighting

22.88 OLS estimators implicitly treat each item as being of equal importance, although some items will have quite substantial sales, while for others sales will be minimal. It is axiomatic that an item with sales of more than 5,000 in a month should not be given the same influence in the regression estimator as one with a few transactions. Commodities with very low sales may be at the end of their life cycles or be custom made. Either way, their (quality-

---

[33]In Chapter 16, Section C.3, the issue of adjusting the base versus the current period's price is discussed, since there are different data demands.

[34]Other averages may be proposed—for example, the needs of an index representative of the "typical" establishment would be better met by a trimmed mean or median.

[35]Diewert (2002c) also suggests matching items where possible and using hedonic regressions to impute the prices of the missing old and new ones. Different forms of weighting systems, including superlative ones, can be applied to this set of price data in each period for both matched and unmatched data.

adjusted) prices and price changes may be unusual.[36] Such observations with unusual prices should not be allowed to unduly influence the index.[37] The estimation of hedonic regression equations by a WLS estimator is preferable. This estimator minimizes the sum of *weighted* squared deviations between the actual prices and the predicted prices from the regression equation, as opposed to OLS estimation, which uses an equal weight for each observation. There is a question as to whether to use quantity (volume) or expenditure weights. The use of quantity weights can be supported by considering the nature of their equivalent "price." Such prices are the average (usually the same) price over a number of transactions. The underlying sampling unit is the individual transaction, so there is a sense that the data may be replicated as being composed of, say, 12 individual observations using an OLS estimator, as opposed to a single observation with a weight of 12 using a WLS estimator. Both would yield the same result. Inefficient estimates arise if the variance of the errors, $V(u_i)$, is not constant—that is, they are heteroskedastic. WLS is equivalent to assuming that the error variances are related to the weights in a multiplicative manner, say $V(u_i) = \sigma^2 w_i^2$.[38] A priori notions as to whether a hedonic regression model predicts better/worse at different levels of quantities or expenditures may help in identifying which weights are appropriate; however, statistical tests or plots of heteroskedasticity may be more useful.

22.89 The sole use of statistical criteria for deciding on which weighing system to use has rightfully come under some criticism. Diewert (2002c and 2005) and Silver (2002) have argued that what matters is whether the estimates are representative of the target index in mind. Conventional target index numbers such as Laspeyres, Paasche, Fisher, and Törnqvist weight price changes by expenditure shares, and the latter two formulas have received support from the axiomatic, stochastic, fixed-base, and economic theoretic approaches, as shown in Chapters 16–18. Thus, value weights are preferred to quantity weights: "The problem with quantity weighting is this: it will tend to give too little weight to cheap models that have low amounts of useful characteristics" (Diewert, 2002c, p. 8). He continues to argue that for a WLS estimator of hedonic time dummy variable indices, expenditure *share* weights should be used, as opposed to the *value* of expenditure, to avoid inflation increasing period 1 value weights, resulting in possible heteroskedastic residuals. Furthermore, for a semilogarithmic hedonic function when models are present in both periods, the average expenditure shares in periods 0 and 1 for *m* items, $\frac{1}{2}(s_{m0} + s_{m1})$, should be used as weights in the WLS estimator. If only matched models exist in the data, then such an estimator may be equivalent to the Törnqvist index. If an observation *m* is only available in one of the periods, its weight should be s_{m0} or s_{m1} accordingly, and the WLS estimator provides a *generalization* of the Törnqvist index.

---

[36]Such observations have higher variances of their error terms, leading to imprecise parameter estimates. This would argue for the use WLS estimators with quantity sold as the weight. This is one of the standard treatments for heteroskedastic errors (see Berndt, 1991).

[37]See Berndt, Ling, and Kyle (2003), Cockburn and Anis (1998), and Silver and Heravi (2005b) for examples. Silver and Heravi (2005b) show old items have above-average leverage effects and below-average residuals. Not only are they different, but they exert undue influence for their size (number of observations).

[38]Estimating an equation for which each variable is divided by the square root of the weight using OLS is an equivalent procedure.

22.90   Silver (2002) has shown that a WLS estimator using value weights will not necessarily give each observation a weight equal to its relative value. The estimator will give more weight to those observations with high leverage effects and residuals. Observations with values of characteristics with large deviations from their means—say, very old or new models—have relatively high leverage. New and old models are likely to be priced at quite different prices than those predicted from the hedonic regression, even after taking into account their different characteristics. Such prices result, for example, from a pricing strategy designed to skim segments of the market willing to pay a premium for a new model, or from a strategy to charge relatively low prices for an old model to dump it to make way for a new one. In such cases the influence these models have on deriving the estimated coefficients will be over and above that attributable to their value weights. Silver (2002) suggests that leverage effects should be calculated for each observation, and those with high leverage and low weights should be deleted, and the regression re-run. Thus, while quantity or value weights are preferable to no weights (that is, OLS), value weights are more appropriate than quantity ones and, even so, account should be taken of observations with undue influence.

22.91   Diewert (2002f) has also considered the issue of weighting with respect to the time dummy hedonic indices outlined in Section C.6. The use of WLS by value involves weights being applied to observations in both periods. However, if, for example, there is high inflation, then the sales values for a model in the current period will generally be larger than those of the corresponding model in the base period, and the assumption of homoskedastic residuals is unlikely to be met. Diewert (2002f and 2005) suggests the use of expenditure *shares* in each period, as opposed to values, as weights for WLS for time dummy hedonic indices. He also suggests that an average of expenditure shares in the periods being compared be used for matched models.

22.92   Data on sales are not always available for weights, but the major selling items can generally be identified. In such cases, it is important to restrict the number of observations of items with relatively low sales, the extent of the restriction depending on the number of observations and the skewness of the sales distribution. In some cases, items with few sales provide the variability necessary for efficient estimates of the regression equation. In other cases, their low sales may be due to factors that make them unrepresentative of the hedonic surface, their residuals being unusually high. An example is low-selling models about to be dumped to make way for new models. Unweighted regressions may thus, suffer from a sampling problem—even if the prices are perfectly quality adjusted, the index can be biased because it is unduly influenced by low-selling items with unrepresentative price-characteristic relationships. In the absence of weights, regression diagnostics have a role to play in helping to determine whether the undue variance in some observations belongs to such unusual low-selling items.[39]

---

[39]A less formal procedure is to take the standardized residuals from the regression and plot them against model characteristics that may denote low sales, such as certain brands (makes) or vintage (if not directly incorporated) or some technical feature that makes it unlikely that the item is being bought in quantity. Higher variances may be apparent from the scatter plot. If certain features are expected to have, on average, low sales, but seem to have high variances, leverages, and residuals (see Silver and Heravi, 2005b), a case exists for at least downplaying their influence. Bode and van Dalen (2001) use formal statistical criteria to decide between

22.93   There is a situation in which an unweighted OLS estimator is preferred. This is when markets are in perfect hedonic equilibrium. Observations with unusual characteristics, say old or new models, would take values which were particularly dispersed from their means and thus increase the variation of the sample for the same underlying model. Such increased variation leads to an increase in the efficiency of the estimates. However, theory and empirical observation (see Sillver and Heravi, 2005b) find that such outlier do not have the same structural relationships as other models. If the sales shares of these new and old models is low relative to the number of models they represent in the market, then an OLS regression would give them undue weight.

22.94   Multicollinearity.

22.95   There are a priori reasons to expect for some commodities that the variation in the values of one characteristic will not be independent of one or a linear combination of other z characteristics. As a result, parameter estimates will be unbiased, yet imprecise. To illustrate this, a plot of the confidence interval for one parameter estimate against another collinear one is often described as elliptical, since the combinations of possible values they may take can easily drift from, say, high values of β_1 and low β_2 to higher values of β_2 and low of β_1. Since the sample size for the estimates is effectively reduced, relatively small additions to and deletions from the sample may affect the parameter estimates more than would be expected. These are standard statistical issues, and the reader is referred to Maddala (1988) and Kennedy (2003). In a hedonic regression, multicollinearity might be expected as some characteristics may be technologically tied to others. Producers including one characteristic may need to include others for it all to work, while for the consumer side, purchasers buying, for example, an up-market brand may expect a certain bundle of features to come with it. Triplett (2002) argues strongly for the researcher to be aware of the features of the product and consumer market. There are standard, though not completely reliable, indicators of multicollinearity (such as variance inflation factors), but an exploration of its nature is greatly aided by an understanding of the market along with exploration of the effects of including and excluding individual variables on the signs and coefficients and on other diagnostic test statistics (see Maddala, 1988).[40]

22.96   If a subset of the estimated coefficients from a hedonic regression are to be used to quality adjust a noncomparable replacement price, and if there is multicollinearity *between* variables in this subset *and* other independent variables, then the estimates of the coefficients to be used for the adjustment will be imprecise. The multicollinearity effectively reduces the sample size, and some of the effects of the variables in the subset may be wrongly ascribed to the other independent variables. The extent of this error will be determined by the strength of the multiple-correlation coefficient between all such "independent" variables (the multicollinearity), the standard error or "fit" of the regression, the dispersion of the independent variable concerned, and the sample size. These all affect the precision of the estimates, since they are components in the standard error of the t-statistics. Even if

---

different weighting systems and compare the results of OLS and WLS, finding, as with Ioannidis and Silver (1999), that different results can arise.

[40]Triplett (2004) stresses the point that $\bar{R}^2$ alone is insufficient for this purpose.

multicollinearity is expected to be quite high, large sample sizes and a well-fitting model may reduce the standard errors on the *t*-statistics to acceptable levels. If multicollinearity is expected to be severe, the predicted value for an item's price may be computed using the whole regression and an adjustment made using the predicted value, as explained in Chapter 8, Section E.4, since there is a sense in which it would not matter whether the variation was wrongly attributed to either β_1 or β_2. If dummy variable hedonic *indices* are being calculated (Section B.3 above), the time trend will be collinear with an included variable if a new feature appears in a new month for the vast majority of the items, so that the data are not rich enough to allow the separate effects of the coefficient on the time dummy to be precisely identified. The extent of the imprecision of the coefficient on the time dummy will be determined by the aforementioned factors. A similar argument holds for omitted variable bias.

Omitted variable bias

22.97   The exclusion of tastes and technology and community characteristics has already been discussed. The concern here is with product characteristics. Consider again the use of a subset of the estimated coefficients from a hedonic regression to quality adjust a noncomparable replacement price. It is well established that multicollinearity of omitted variables with included variables leads to bias in the estimates of the coefficients of included ones. If omitted variables are *independent* of the included variables, then the estimates of the coefficients on the included variables are unbiased. This is acceptable in this instance; the only caveat is that it may be that the quality adjustment for the replacement item also requires an adjustment for these omitted variables, and this, as noted by Triplett (2002), has to be undertaken using a separate method and data. But what if the omitted variable is multicollinear with a subset of included ones, and these included ones are to be used to quality adjust a noncomparable item? In this case, the coefficient on the subset of the included variables may be wrongly picking up some of the omitted variables' effects. The coefficients will be used to quality adjust prices for items that differ only with regard to this subset of included variables, and the price comparison will be biased if the characteristics of both included and omitted variables have different price changes. For hedonic *indices* using a dummy time trend, the estimates of quality-adjusted price changes will suffer from a similar bias if excluded from the regression are omitted variables multicollinear with the time change. What are picked up as quality-adjusted price changes over time may, in part, be changes due to the prices of these excluded variables. This requires that the prices on the omitted characteristics follow a different trend. Such effects are most likely when there are gradual improvements in the quality of items, such as the reliability and safety of consumer durables,[41] which are difficult to measure, at least for the sample of items in real time. The quality–adjusted price changes will thus, overstate price changes in such instances.

---

[41]There are some commodity areas, such as airline comfort, that have been argued to have overall patterns of decreasing quality.

Figure 21.1. Consumption and Production Decisions for Combinations of Characteristics
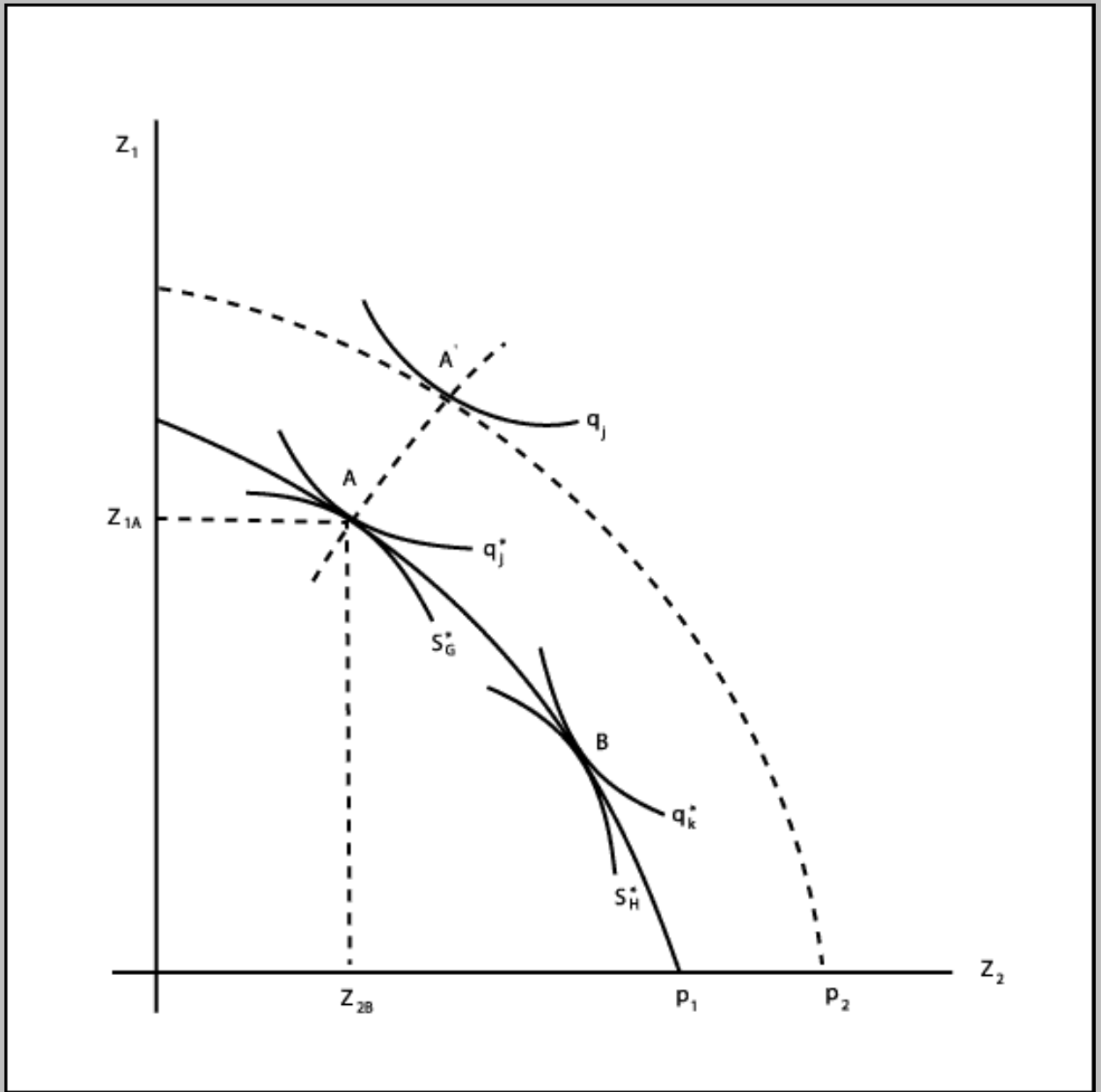
Heravi, Saeed and Mick Silver, 2007, "Hedonic Indexes: A Study of Alternative Methods," in *Hard-to-Measure Goods and Services: Essays in Honour of Zvi Griliches*, ed. by Berndt and Hulten, NBER/CRIW (Chicago: University of Chicago Press).

Silver, Mick and Saeed Heravi, 2005a, "Purchasing Power Parity Measurement and Bias from Inappropriate Item Selection and Quality Comparisons: Principles and an Empirical Study," *Journal of Official Statistics*, Vol. 21, No. 3, pp. 436-87.

Silver, Mick and Saeed Heravi, 2005b, "A Failure in the Measurement of Inflation: Results from a Hedonic and Matched Experiment Using Scanner Data," *Journal of Business and Economic Statistics*, Vol. 23, No. 3 (July), pp. 269-81.

Diewert, Erwin, 2005, "Weighted Country Product Dummy Variable Regressions and Index Number Formulas," *Review of Income and Wealth*, Vol. 51, No. 4, December, pp. 561–70.

Diewert, Erwin , Heravi, Saeed and Mick Silver, 2007 "Hedonic imputation indexes versus time dummy hedonic indexes", (with Erwin Diewert and Saeed Heravi), <u>IMF Working Paper Series</u> No. 07/234, October.