

Evaluating Welfare with Nonlinear Prices

Peter C. Reiss[†] and Matthew W. White[‡]

June 2006

Abstract

This paper examines how to evaluate consumer welfare when consumers face nonlinear prices. This problem arises in many settings, such as devising optimal pricing strategies for firms, assessing how price discrimination affects consumers, and evaluating the efficiency costs of many transfer programs in the public sector. We extend prior methods to accommodate a broad range of modern pricing practices, including menus of pricing plans. This analysis yields a simpler and more general technique for evaluating exact consumer surplus changes in settings where consumers face nonlinear prices. We illustrate our method using recent changes in mobile phone service plans.

[†]Graduate School of Business, Stanford University, Stanford, CA 94305-5015, and NBER.

[‡]The Wharton School, University of Pennsylvania, Philadelphia PA 19104-6372, and NBER.

Acknowledgments. Many thanks to Katja Seim for the internet access and mobile phone plan data.

1 Introduction

When sellers face heterogeneous buyers, it is often profitable to employ nonlinear pricing strategies that separate consumers according to willingness to pay. Familiar examples include mobile phone calling plans, deductible choices for insurance, point and rate options for home mortgages, and volume discount pricing of consumer goods. While an extensive theoretical literature examines how these practices benefit sellers, it also stresses their theoretically ambiguous consequences for consumers.¹ This ambiguity places consumer welfare questions squarely in the empirical domain.

In this paper, we examine how to evaluate exact consumer surplus when consumers face nonlinear prices. Researchers have long recognized that evaluating welfare when consumers face nonlinear prices poses an intricate problem if consumers' income elasticities are non-negligible. Partly as a result, this aspect of demand behavior is commonly ignored in analyses of price discrimination.² Yet empirical studies typically indicate that consumption varies with income for a wide variety of goods and services. Thus it seems useful to have empirical techniques for evaluating consumer welfare that do not ignore either income or nonlinear prices, but instead address these features in an integrated way.

Our proximate goal is to provide applied researchers with a practical method for evaluating consumer welfare with nonlinear prices, including menus of pricing plans. Elements of this problem are found in several literatures in economics, including demand analysis for rationed goods (Neary and Roberts 1980), the efficiency costs of income taxation (Hausman 1983), the benefits of in-kind subsidy programs (Schwab 1985), and durable goods' demand (Dubin and McFadden 1984). Analytically, questions involving consumer welfare in these settings are special cases of the general problem we consider here. Yet even in these environments, where prices and budget constraints are simple piecewise linear functions, existing methods for evaluating welfare tend to be computationally involved. Few (if any) studies attempt to measure exact consumer surplus facing the complex menus of nonlinear prices encountered in many product markets.³

The technique we present allows researchers to evaluate exact consumer welfare changes in complex pricing environments, and is simpler to implement than prior methods. Our approach is based on recasting the problem in 'supply and demand' space, rather than

¹Stole (2003), Wilson (1993), and Varian (1989) provide excellent surveys. A classical treatment is Roberts (1979).

²An exception is Wilson (1993), who observes that income effects can have significant consequences for the design of nonlinear pricing schemes and how they impact consumers.

³There is an extensive econometric literature on modeling choice with nonlinear budget constraints, principally in public finance (see, *e.g.*, Hausman 1985 or Moffit 1986, 1990). A much smaller literature examines how to evaluate consumer welfare in such settings. See Creedy and Kalb (2005), DeBorger (1989), Fullerton and Gan (2004), or Maddock (1989), which extend Hausman's (1983) technique. We relate our approach to these antecedents in detail further below.

proceeding from the budget constraint formulation typical of applied welfare analyses (see Slesnick 1998). This approach leads to comparatively simple equations for evaluating exact (variational) consumer welfare changes with budget sets of quite general forms. In the important special case of discrete nonlinear price schedules, which are common in applications, these equations are computationally straightforward to evaluate if an analytic representation for indirect utility is available. If not, existing and well-known numerical methods for obtaining expenditure functions from ordinary demand equations, such as Vartia (1983)'s method, can be readily applied in our setting as well. As a consequence, our methods do not require a researcher to know the consumer's directly utility specification to evaluate exact consumer welfare; all that is required is the ordinary demand function that is typically estimated in applied work.⁴

Since the economic motivation for nonlinear prices arises when sellers have imperfect information about buyers' tastes, we also consider the implications of unobserved heterogeneity for evaluating welfare. The literature has amply recognized that ignoring unobserved heterogeneity in preferences can yield misleading estimates of consumer welfare in many settings (McFadden 1999, Creedy and Duncan 2002, Fullerton and Gan 2004). In the presence of nonlinear prices, however, accounting for unobserved heterogeneity becomes complex because observationally-equivalent consumers with different preferences tend to self-select different marginal prices. As we show, this sorting behavior must be addressed properly in order to obtain accurate measures of welfare. We illustrate the potential error magnitudes that can arise using an example based on the changes in wireless calling plans that occurred in Washington D.C. in 2002 and 2003.

2 The Setting

A few preliminaries are useful concerning the definition of exact consumer surplus when consumers face nonlinear prices.

The analysis of welfare in applied work typically begins with empirical estimates of the consumer's ordinary demand functions, here written as $q(p, y)$. These functions solve

$$\max_q u(q) \quad \text{s.t.} \quad p \cdot q \leq y \quad (1)$$

where u is utility, p a vector of prices, and y is income. This formulation assumes the consumer can purchase any quantity of good i at a constant (marginal and average) price of p_i per unit. If the consumer faces nonlinear prices, however, the consumer's budget

⁴This aspect parallels the measurement of exact consumer surplus in the classical case where prices do not vary with quantity (Hausman 1981).

constraint differs from that in (1) and the standard constructs needed to evaluate welfare—indirect utility, the expenditure function, and so on—must be redefined appropriately.

Suppose now that one or more goods are subject to nonlinear pricing. Let $P_i(q_i)$ be the total price—that is, the consumer’s bill—for q_i units of good i . For example, if good i is offered with a two-part price consisting of an initial fee F_i and a constant marginal price of p_i per unit, then $P_i(q_i) = F_i + p_i q_i$. If the consumer faces a menu of pricing options, then we shall interpret $P_i(q_i)$ to be the *minimum* expense necessary to purchase q_i units. In this way, we may represent the consumer’s problem facing a menu of pricing options using a single, nonlinear price schedule. This representation is standard in theoretical work and quite general; it rests upon the result that a single nonlinear tariff can be implemented through a menu of simple (*e.g.*, two-part) tariffs, and vice versa.⁵

With nonlinear prices, the consumer’s problem becomes

$$\max_q u(q) \quad \text{s.t.} \quad \sum_{i=1}^n P_i(q_i) \leq y \quad (2)$$

where $q = \{q_1, q_2, \dots, q_n\}$ as before. Let the solution be $\tilde{q}(\mathcal{P}, y)$, where $\mathcal{P} = \{P_1, P_2, \dots, P_n\}$ comprises the collection of total price schedules for all goods. Note that a good available at a constant price p_i has $P_i(q_i) = p_i q_i$, so (2) accommodates uniformly-priced goods as well. As usual, the consumer’s indirect utility function is the value of problem (2): $\tilde{v}(\mathcal{P}, y) = u(\tilde{q}(\mathcal{P}, y))$. We use \tilde{v} to distinguish it from the conventional indirect utility function with uniform prices, $v(p, y) = u(q(p, y))$. This distinction is central to evaluating welfare with nonlinear prices.

The standard metrics of exact consumer surplus are Hicks’ compensating and equivalent variation. A precise definition in this setting uses $\tilde{v}(\mathcal{P}, y)$ or, equivalently, the expenditure function that applies with nonlinear prices. Specifically, let $\tilde{e}(\mathcal{P}, u)$ be the inverse of $\tilde{v}(\mathcal{P}, y)$ with respect to y . If prices change from an initial set \mathcal{P}^0 to a new set \mathcal{P}^1 , then exact consumer surplus is

$$\tilde{s}(\mathcal{P}^0, \mathcal{P}^1, y) = \tilde{e}(\bar{\mathcal{P}}, u^1) - \tilde{e}(\bar{\mathcal{P}}, u^0), \quad (3)$$

where $u^0 = \tilde{v}(\mathcal{P}^0, y)$ and $u^1 = \tilde{v}(\mathcal{P}^1, y)$. Choosing reference prices $\bar{\mathcal{P}} = \mathcal{P}^1$ gives the compensating variation and $\bar{\mathcal{P}} = \mathcal{P}^0$ gives the equivalent variation. In either case, $\tilde{s}(\mathcal{P}^0, \mathcal{P}^1, y)$ is the lump-sum change in income that would leave the consumer indifferent to the price changes.⁶ Thus, equation (3) provides a precise generalization of exact consumer surplus to situations where the consumer faces nonlinear pricing.

It is important to note that different nonlinear pricing schemes result in different func-

⁵See Wilson (1993). This representation assumes no intertemporal changes in preferences, however, which might be unduly restrictive in some applications.

⁶The compensating and equivalent variation differ because they assume different prices apply when income is spent.

tional expressions for the consumer’s expenditure function. For instance, if a good’s pricing changes from a two-part tariff (say) to a three-part tariff, that change will entail a new expression for the expenditure function $\tilde{e}(\mathcal{P}, u)$ on which welfare calculations are based. The dependence of the expenditure function on the precise form of nonlinear pricing is an important problem in applied work, as explained next.

3 Evaluating Welfare: Shadow Price Techniques

It is useful to clarify why computing exact consumer surplus with nonlinear prices tends to be a difficult problem. This occurs even before the complications of unobserved heterogeneity in preferences are considered. Accordingly, here we set aside the issue of preference heterogeneity altogether and consider the problem from the standpoint of consumer theory.

A direct extension to this setting of ‘textbook’ methods for evaluating compensating and equivalent variation is the following. Starting from a specific functional form for the (direct) utility function $u(q)$, computing welfare with nonlinear prices proceeds by solving (2) to obtain $\tilde{q}(\mathcal{P}^0, y)$, $\tilde{q}(\mathcal{P}^1, y)$, and the indirect utility functions $\tilde{v}(\mathcal{P}^0, y)$, $\tilde{v}(\mathcal{P}^1, y)$. Inverting and differencing appropriately then gives (3). While conceptually straightforward, a significant impediment arises in implementing this method. Solving the optimization program in (2) can be cumbersome and difficult in practice, requiring general nonlinear programming techniques. Not surprisingly, few papers evaluate welfare using this method.⁷

Because of this difficulty, it is desirable to consider methods that avoid solving for $\tilde{v}(\mathcal{P}, y)$. An alternative, and quite general, method for evaluating welfare with nonlinear prices becomes clear by recasting the consumer’s problem in ‘demand and supply’ space. The economic logic of this technique is most transparent graphically, as shown next.

3.1 A Diagrammatic Analysis

Figure 1 interprets exact consumer welfare with nonlinear prices in terms of supply and demand at the individual level. For simplicity we assume here that prices change for a single (nonlinearly-priced) good, and all other goods’ prices remain constant; hence we drop the good index i temporarily.

In this figure, the price schedules $p^0(q)$ and $p^1(q)$ indicate the *marginal* price for the q th unit of a good purchased under the initial and new price schedules, respectively. Thus the

⁷Two early efforts are Blomquist (1983) and Hausman (1983). A more recent discussion can be found in Fullerton and Gan (2004). These approaches are based on mathematical programming techniques for optimizing convex objective functions with (multiple) linear constraints. Such constraints arise if the price schedules $P_i(q_i)$ are piecewise linear in q_i , which is common in applications.

total cost of purchasing q units at initial prices is the area under $p^0(q)$ from 0 to q . Note we have here assumed price schedules that are smooth functions of quantities; this simplification will make the economic intuition of how to evaluate exact consumer surplus most transparent. The mathematical treatment and examples in succeeding sections address discontinuous and discrete (multi-part) nonlinear prices, which require minor adjustments.

The shaded area in Figure 1 indicates the compensating variation if prices fall from $p^0(q)$ to $p^1(q)$. Here q^0 is the initial quantity chosen, p_*^0 the marginal price, and $u^0 = u(q^0)$ initial utility. Since this consumer faces higher prices than p_*^0 inframarginally, its compensating variation is not the conventional area under the compensated demand function $h(p, u^0)$ between p_*^0 and p_*^1 . Rather, the expenditure change that maintains constant utility is the region under h that lies between the two *schedules*. This region adjusts for the price changes at inframarginal quantities. Exact consumer surplus thus has its customary interpretation as an area under the compensated demand curve, but it is a complex area when prices vary with quantity.

This observation is nonetheless useful for applied work. Using integration by parts, the shaded region can be written as

$$\int_{p_*^1}^{p_*^0} h(p, u^0) dp + [d^0(p_*^0) - d^1(p_*^1)]. \quad (4)$$

The term $d^0(p_*^0)$ is the difference between the expenditure necessary to purchase q^0 under the price schedule $p^0(q)$ and the expenditure necessary to purchase the same quantity at a constant marginal price p_*^0 . In Figure 1, d^0 is the roughly triangular region bounded by $p^0(q)$ and p_*^0 . The interpretation of d^1 is analogous for the alternative price schedule.

Our insight is that equation (4) provides a means of calculating exact consumer surplus without solving for the expenditure function under nonlinear prices, $\tilde{e}(\mathcal{P}, u)$. Here it is important to observe that $h(p, u^0)$ is *not* the compensated demand curve that solves the consumer's expenditure minimization problem given nonlinear prices. Rather, $h(p, u)$ is the income-compensated demand curve that would apply if, *counter to fact*, the consumer faced (uniform) prices that do not vary with quantity.

The economic logic for using this compensated demand function to determine consumer welfare under nonlinear prices lies in a subtle aspect of its information content. Unlike the ordinary demand function, this constant-utility demand function $h(p, u)$ contains sufficient information about preferences to determine a key relationship: how total expenditure (even under nonlinear prices) varies with the consumer's marginal rate of substitution along a given indifference curve. This relationship also pins down which indifference curve is achieved, since total expenditure must equal total income. Together, these provide enough information to evaluate exact consumer surplus. In the compensat-

ing variation case, for example, we hold the initial utility level u^0 fixed and move along the demand curve $h(p, u^0)$ until it crosses the new price schedule. The difference between initial income and total expenditure (under the new price schedule) at the new bundle is the compensating variation, and by construction leaves the consumer indifferent to the price schedule change.

A more formal analysis of this approach is provided next. Its practical consequence is that an applied researcher never needs to solve nonlinear programming problems like (2) directly. Indeed, even integrating the function h to evaluate (4) is unnecessary, because its value can be determined using the conventional expenditure function assuming a linear budget constraint. This much is a familiar task in applied welfare analysis (see Hausman 1981 or Slesnick 1998).

There is one aspect of Figure 1 that merits further discussion. The values of $d^0(p_*^0)$ and $d^1(p_*^1)$ are readily calculated from observed price schedules in applied work, once the appropriate marginal valuations p_*^0 and p_*^1 are known. Determining these marginal prices (and the reference utility level) is typically straightforward using the same ‘supply equals demand’ conditions indicated in Figure 1. We elaborate on this step in section 3.3.

3.2 A Precise Characterization

This section provides a more formal analysis of how to evaluate consumer welfare under nonlinear prices, without directly solving the consumer’s problem (2) for the indirect utility $\tilde{v}(\mathcal{P}, y)$ or expenditure function $\tilde{e}(\mathcal{P}, y)$. It also provides a precise justification for the depiction of exact consumer surplus shown in Figure 1, and connects this procedure to related techniques in the literature.

The main insight lies in connecting the consumer’s expenditure minimization problem under nonlinear pricing to the classical expenditure problem with uniform prices. As above, let $h(p, u)$ be the consumer’s income-compensated demand function when facing uniform prices p ; the expenditure function with uniform prices is then

$$e(p, u) = p \cdot h(p, u). \quad (5)$$

Similarly, let $\tilde{h}(\mathcal{P}, u)$ be the solution to the consumer’s expenditure minimization problem when (one or more) goods have nonlinear prices:

$$\min_q \sum_{i=1}^n P_i(q_i) \quad \text{s.t.} \quad u(q) \geq \bar{u}. \quad (6)$$

The expenditure function with nonlinear prices is then

$$\tilde{e}(\mathcal{P}, u) = \sum_{i=1}^n P_i \tilde{h}(\mathcal{P}, u). \quad (7)$$

The expenditure functions in (5) and (7) have a useful relationship for evaluating welfare. This relationship arises from the equality of their respective compensated demand curves when evaluated at the consumer's marginal willingness to pay. For any expenditure minimizing bundle $\tilde{h}(\mathcal{P}, \bar{u})$, one can always construct a hyperplane tangent to the indifference curve that gives \bar{u} at quantity $\tilde{h}(\mathcal{P}, \bar{u})$. If the vector p_* is orthogonal to this tangent hyperplane, then

$$\tilde{h}(\mathcal{P}, \bar{u}) = h(p_*, \bar{u}). \quad (8)$$

The left-hand side is the least-cost bundle that achieves utility \bar{u} facing nonlinear prices \mathcal{P} , and the right-hand side is the least-cost bundle that achieves the same utility facing a linear budget constraint. Equation (8) simply states that the two are equal if the linear budget constraint is tangent to \bar{u} at the point $\tilde{h}(\mathcal{P}, \bar{u})$. Note that the ratio of any two prices in the vector p_* is the consumer's optimal marginal rate of substitution at $\tilde{h}(\mathcal{P}, \bar{u})$. For this reason we call p_* a *shadow price*.⁸

Equation (8) provides the link to relate the two forms of the expenditure function above. Differencing (5) from (7) and rearranging yields

$$\tilde{e}(\mathcal{P}, u) = e(p, u) + \sum_i P_i (\tilde{h}(\mathcal{P}, u)) - p \cdot h(p, u)$$

Since this holds for any p , it holds for a p_* that satisfies (8). Substituting $h(p_*, u)$ for $\tilde{h}(\mathcal{P}, u)$ gives

$$\begin{aligned} \tilde{e}(\mathcal{P}, u) &= e(p_*, u) + \underbrace{\sum_i P_i (h(p_*, u)) - p_* \cdot h(p_*, u)}_{d(p_*, u)} \\ &= e(p_*, u) + d(p_*, u) \end{aligned} \quad (9)$$

The expenditure function \tilde{e} facing the nonlinear schedules \mathcal{P} is equal to the conventional expenditure function evaluated at a constant shadow price plus an adjustment term that reflects the nonlinearity of prices. As in section 3.1, $d(p_*, u)$ is the difference between the minimum expenditure necessary to purchase the quantity $h(p_*, u)$ under nonlinear pricing and the expenditure necessary to purchase that bundle at a constant price p_* . We call $d(p_*, u)$ the *difference in inframarginal expenditure*.

With equation (9), we can obtain a useful simplification for exact consumer surplus

⁸The shadow price term originates in the analysis of demand under rationing, which is a special case of the analysis here. We clarify this connection in section 4.2.

with nonlinear prices. Recall that variational surplus under nonlinear prices is

$$\tilde{s}(\mathcal{P}^0, \mathcal{P}^1, y) = \tilde{e}(\bar{\mathcal{P}}, u^1) - \tilde{e}(\bar{\mathcal{P}}, u^0).$$

Evaluating the two right-hand terms using (9) yields

$$\tilde{s}(\mathcal{P}^0, \mathcal{P}^1, y) = e(\bar{p}_*, u^1) - e(\bar{p}_*, u^0) + [d(\bar{p}_*, u^1) - d^0(\bar{p}_*, u^0)],$$

where the reference price vector \bar{p}_* is proportional to the consumer's marginal rates of substitution at the expenditure-minimizing bundle facing \mathcal{P}_0 or \mathcal{P}_1 (depending whether the equivalent or compensating variation is desired, respectively). This yields two comparatively simple expressions for the compensating and the equivalent variation under nonlinear prices:

$$CV(\mathcal{P}^0, \mathcal{P}^1, y) = y - e(p_*^1, u^0) - d(p_*^1, u^0) \quad (10a)$$

and

$$EV(\mathcal{P}^0, \mathcal{P}^1, y) = e(p_*^0, u^1) + d(p_*^0, u^1) - y. \quad (10b)$$

The variational surplus measures for general nonlinear budget set changes can therefore be expressed in terms of the consumer's income, its conventional expenditure function under uniform prices, and the difference in inframarginal expenditures. There is no need to solve for the expenditure function $\tilde{e}(\bar{\mathcal{P}}, u)$ that applies with nonlinear prices.

Last, a transformation of (10a) provides the basis for the interpretation of Figure 1 given in section 3.1. Since

$$y = \tilde{e}(\mathcal{P}^0, u^0) = e(p_*^0, u^0) + d(p_*^0, u^0),$$

the compensating variation can alternatively be expressed as

$$\begin{aligned} CV(\mathcal{P}^0, \mathcal{P}^1, y) &= \tilde{e}(\mathcal{P}^0, u^0) - \tilde{e}(\mathcal{P}^1, u^0) \\ &= e(p_*^0, u^0) - e(p_*^1, u^0) + [d(p_*^0, u^0) - d(p_*^1, u^0)] \\ &= \int_{p_*^1}^{p_*^0} h(p, u^0) dp + [d^0(p_*^0) - d^1(p_*^1)]. \end{aligned} \quad (11)$$

The last line equals the shaded area in Figure 1. This provides the precise justification for equation (4) and the interpretation of compensating variation as the region between the marginal price schedules under $h(p, u^0)$. The analogous expression for the equivalent variation is similarly derived.

Antecedents. This analysis is related to the “virtual price and income” technique introduced by Burtless and Hausman (1978) to model choice facing piecewise linear budget

constraints (see also Hausman, 1985). The prototypical application is labor supply analysis with multiple income tax brackets; see Blundell and McCurdy (1999) or Creedy and Duncan (2002) for recent appraisals. In that literature, the expenditure decomposition in equation (9), or $e(p_*, u) + d(p_*, u)$, is termed a consumer's 'virtual income' and the marginal price p_* a 'virtual price.' These constructs are used to model the choice $\tilde{q}(\mathcal{P}, y)$ that solves the consumer's problem (2) in terms of the ordinary demand (or labor supply) function $q(p, y)$ that applies facing a linear budget constraint. Terminology aside, this amounts to a standard mathematical programming technique for solving optimization problems with multiple linear constraints.

Evaluating welfare is more involved, however. In principle, one could iteratively 'guess' at trial income changes Δy , re-solve the consumer's problem in (2) by constructing the appropriate new set of virtual incomes and prices to determine $\tilde{q}(\mathcal{P}^1, y + \Delta y)$, and then assess whether this quantity yields the same direct utility as income y under original prices \mathcal{P}^0 . This becomes a computationally involved iterative procedure because the virtual price and income technique provides an algorithm, not a functional expression, for computing the value of $\tilde{q}(\mathcal{P}^1, y + \Delta y)$; the algorithm must be re-applied at each trial value of Δy until the value that leaves the consumer indifferent to the change in the price schedules (tax schedules, etc.) is obtained. The handful of papers that evaluate exact consumer welfare in public finance applications with nonlinear budget constraints appear to employ this procedure, including Blundell et al. (1988), Fullerton and Gan (2004), and Creedy and Kalb (2005).

The shadow price method summarized by Figure 1 solves this problem differently. The key insight lies in defining shadow prices using the compensated (Hicksian) demands in (8); in contrast, the 'virtual prices' employed in the literature are (implicitly) defined in terms of the ordinary (Marshallian) demands. The latter is useful for modeling choice, while the former is a more direct route for evaluating welfare. The reason is that defining shadow prices using the compensated demand function immediately reveals how much we must change income to hold utility constant *if* these marginal prices were constant. It is a simple matter to then determine the complete change in exact consumer surplus, by adding in the change in inframarginal expenditures. This is the logic of equation (11).

The benefit of this shadow-price method is that it allows one to avoid the cumbersome process of repeatedly re-solving the consumer's optimization problem in (2) in order to evaluate welfare with nonlinear prices, as occurs in previous work. Instead, our method involves finding a single shadow price, p_* , which is often a much simpler computational procedure. As shown next, it essentially amounts to finding where the consumer's compensated demand function intersects the supply (price) schedule.

In addition, there is an effective dimension reduction using our method. When the

researcher is evaluating welfare facing price changes for K goods, the iterative scheme used in prior studies involves (repeatedly) solving a nonlinear optimization problem in K dimensions—even if there is only one nonlinearly-priced good in the budget set. By contrast, the shadow price method above requires solving only a one-dimensional problem (for a single shadow price) when there is only one nonlinearly priced good among all K . This dimension-reduction occurs because the shadow prices for all other goods equal their observed (uniform) prices, which are given. If the problem is solved as in prior work, however, then one must solve the K -dimensional consumers' optimization problem because all goods' quantities change with prices. From a practical standpoint, this dimension reduction can greatly simplify the computational burden of evaluating exact consumer welfare with nonlinear prices when the researcher is modeling consumption choices for more than two goods.

Last, because our shadow price method makes it unnecessary to solve (2) directly, it is also unnecessary to have an explicit functional representation for the consumer's direct utility function. As Creedy and Duncan (2002) note, demand specifications that fit data well often have direct utility functions that are difficult to obtain analytically. To be clear about how one can determine the correct shadow prices without a direct utility function, we address this step next.

3.3 Determining Shadow Prices without Direct Utility

Two pieces of information are needed to evaluate exact consumer surplus with nonlinear prices using equation (10a) or (10b). The first is the expenditure function under uniform prices, $e(p, u)$, along with its derivative $h(p, u) = \partial e(p, u) / \partial p$. In applied work, it is common to begin with an econometric estimate of a consumer's ordinary demand function $q(p, y)$. Given a demand model, direct methods for recovering $e(p, u)$ are standard in the literature. Hausman (1981) describes analytic techniques based on the usual form of Roy's Identity for two-good models common in applied work, and Vartia (1983) and others provide simple numerical methods for more complex demand systems.

Second, it remains to determine p_*^0, p_*^1 , and the reference utility u^0 (or u^1 if equivalent variation is desired). There are several ways to do this, but the simplest follows from equating supply and demand. Consider again Figure 1, which describes the compensating variation case. The *supply schedule* for good i , $s_i(p_i)$, is the inverse of the marginal price schedule $p_i(q_i)$. Thus $s_i(p_i)$ indicates the set of quantities available at a marginal price of

p_i per unit. Since the initial shadow price sets demand equal to supply, we have⁹

$$h_i(p_*^0, u^0) = s_i^0(p_*^0). \quad (12)$$

With a single nonlinearly-priced good, this reduces to one equation in two unknowns, p_*^0 and u^0 . However, h also conveys additional information: $p \cdot h(p, u)$ indicates how total expenditure (under uniform prices) varies with marginal willingness to pay, p , along *any* indifference curve. The correct initial indifference curve, u^0 , equates total expenditure with nonlinear prices to total income:

$$y = \sum_i P_i(h_i(p_*^0, u^0))$$

or, equivalently,

$$y = e(p_*^0, u^0) + d^0(p_*^0). \quad (13)$$

The equivalence of these two expressions follows from (9). Note that the initial difference in inframarginal expenditure, $d^0(p_*^0)$, does not depend directly upon utility (it depends upon u only through p_*^0 ; see Figure 1). Solving (12) and (13) gives p_*^0 and u^0 .

Although determining these marginal prices and reference utility might first appear to be an involved procedure, solving these conditions in practice is often a simple matter. The reason lies in the economics of demand and sellers' pricing decisions: Theory requires the demand curve h to be monotonically decreasing in price, and price schedules chosen by sellers are typically monotone as well.¹⁰ We illustrate this in several examples below.

Given u^0 , we can then determine p_*^1 by moving along the demand curve $h(p, u^0)$ until it intersects the other price (supply) schedule:

$$h_i(p_*^1, u^0) = s_i^1(p_*^1), \quad (14)$$

The solution for p_*^1 given by (14) is indicated in Figure 1. The compensating variation is evaluated by inserting p_*^1 and u^0 into (10a).¹¹

If the equivalent variation is of interest, the procedure is similar. We first obtain the reference utility level u^1 appearing in (10b) by solving the marginal and total valuation

⁹If the supply schedule exhibits discontinuities, we can replace the right-hand side of (12) by its closed graph. This ensures that there will always exist a solution, given monotonicity of the compensated demand functions in prices. An example is provided in the next section.

¹⁰Non-monotonic marginal price schedules do occur for some goods, the most common product-market example being mobile phone contracts. We address this situation below.

¹¹There is a small gap in our discussion: Technically, (12)-(14) are necessary conditions for the shadow prices and reference utility level. The distinction between necessary and sufficient conditions becomes important if the supply schedule is decreasing but less price-elastic (that is, $p(q)$ decreases faster) than the compensated demand function. In that situation there may be multiple distinct solutions for (p_*^0, u^0) . Fortunately, this is easily resolved by selecting the pair achieving the highest utility. We illustrate this in section 4.1.

conditions:

$$h_i(p_*^1, u^1) = s_i^1(p_*^1)$$

and

$$y = e(p_*^1, u^1) + d(p_*^1, u^1).$$

Given u^1 , we determine p_*^0 from the intersection of (compensated) demand and the new price schedule:

$$h_i(p_*^0, u^1) = s_i^0(p_*^0).$$

The equivalent variation is then obtained by inserting p_*^0 and u^1 into (10b).

Depending on the problem being considered, there are other ways to determine the shadow prices and implement (10a) or (10b). For instance, if the initial consumption bundle q^0 is observed (in data) and the form of the direct utility function $u(q)$ is specified, then one can immediately determine the initial utility level from $u^0 = u(q^0)$. Steps (12)-(13) become unnecessary and only the new shadow price needs to be found, using (14), to evaluate (e.g.) the compensating variation.

The method described here has an important advantage, however, in that it does not require the form of the direct utility function to be known. Thus, a researcher can begin with an empirically satisfactory model for the ordinary demand function $q(p, y)$ (which may be parametric or non-parametric), and then compute exact consumer surplus under nonlinear prices without having to recover the direct utility function.¹² The method described here is quite general, accommodating convex or nonconvex budget sets and price schedules of arbitrary form.

4 Discrete Price Schedules

It is useful to consider a few examples that involve discrete and discontinuous price schedules. These are common in practice, as applications of nonlinear pricing often exhibit only a few possible marginal prices that change at specific quantity thresholds. Simple volume discounts are an oft-encountered example. This section also clarifies a few technicalities that can arise in using the previous section's methods in applied work.

4.1 Decreasing Marginal Prices and Volume Discounts

Figure 2 depicts a simple volume discount pricing scheme for a single good. Here the consumer initially pays a higher marginal price p_h^0 for each unit purchased up to the quantity

¹²Doing so assumes the demand model also satisfies the standard integrability conditions to represent a coherent set of consumer preferences, of course.

k^0 , and then a lower marginal price p_l^0 for each unit consumed thereafter. Volume discounts of this form are common in a wide array of products and services; examples span retail consumer goods, building materials, software licenses, car and hotel rental rates, and so on.

We now consider how to evaluate exact consumer surplus when a volume discount price schedule of this form is changed, using the methods described in section 3. As is conventional in applied work, we assume a composite outside good (serving as a price numeraire). Suppose the ordinary demand function (facing a constant price p) has the log-linear form

$$\log q = \alpha + \beta \log p + \gamma \log y.$$

Solving Roy's Identity and inverting on y gives the expenditure function that applies under uniform prices,

$$e(p, u) = \left[(1 - \gamma) \left(u + e^\alpha p^{1+\beta} / (1 + \beta) \right) \right]^{\frac{1}{1-\gamma}} \quad (15)$$

provided $\gamma \neq 0$ (see, e.g., Hausman 1981). Differentiating with respect to p yields the compensated demand function

$$h(p, u) = e^\alpha p^\beta \left[(1 - \gamma) \left(u + e^\alpha p^{1+\beta} / (1 + \beta) \right) \right]^{\frac{\gamma}{1-\gamma}}.$$

Now consider a consumer's compensating variation when the prices for this good changes from the initial price schedule $s^0(p)$ to the new, lower schedule $s^1(p) = \{p_h^1, p_l^1, k^1\}$ as shown in Figure 2. We can determine the compensating variation by first using (15) to evaluate the area to the left of $h(p, u^0)$ between the initial and new shadow prices p_l^0 and p_h^1 . According to (11), to this we add the initial inframarginal expenditure difference $d(p_*^0, u^0) = (p_h^0 - p_l^0)k^0$. This gives the compensating variation for this price change for this consumer, since $d(p_*^1, u^0) = 0$ under the new schedule. Note that the form of the direct utility function was not needed to determine the compensating variation.

Alternatively and equivalently, we can immediately apply (10a) using (15) to find

$$CV = y - \left[(1 - \gamma) \left(u^0 + e^\alpha (p_h^1)^{1+\beta} / (1 + \beta) \right) \right]^{\frac{1}{1-\gamma}}$$

The value of u^0 is obtained from (12) and (13) using $e(p, u)$ and $h(p, u)$ from (15).

There are two potential subtleties about decreasing marginal prices that are illustrated in this example. The first is that if the price schedule has a region where the marginal price falls faster than the slope of the compensated demand function, then it is possible for h to intersect the supply schedule at more than one point. In such cases, care must be exercised to select the solution values that minimize total expenditure. We elaborate on

this possibility in Appendix A.

The second point to note is that the ordinary demand function passing through supply at the initial consumption level q^0 is not $q(p, y)$, but rather is $q(p, y + d^0)$ where $d^0 = d(p_*^0, u^0)$. The reason the additional term d^0 enters the ordinary demand function is that if the income elasticity of demand is nonzero, then there is an income effect due to the higher inframarginal price up to k^0 . This additional inframarginal expenditure is $d(p_*^0, u^0)$.

4.2 Increasing Marginal Prices: Limited Discount Pricing, Rationing, and Lifeline Pricing

Figure 3 illustrates the compensating variation for a change in a discrete, *increasing* marginal price schedule. In applied work, pricing of this form commonly arises through limited-quantity discounts (e.g., when a seller limits purchases at a discounted price to n units per customer). Other examples occur in utility pricing, where schedules of this form are known as ‘lifeline’ pricing; see Reiss and White (2005) for an analysis in that setting. In public finance, limited-benefit programs (sometimes called ‘in-kind transfer programs’) give rise to the same structure of marginal prices; see Schwab (1985) and DeBorger (1989) for examples.

Unlike the decreasing case, with (weakly) increasing marginal prices the consumer’s budget set is convex. Assuming decreasing marginal utility, there will be only one point where the compensated demand curve crosses the (closed graph of the) price schedule. However, a technicality arises in that it is possible for the appropriate compensated demand curve to cross the supply (price) schedule at a ‘gap’ between two marginal prices. This possibility is illustrated in Figure 3.

When this occurs, the initial shadow price p_*^0 is still the consumer’s marginal willingness to pay at the preferred initial consumption bundle q^0 . However, this shadow price is *not* equal to either of the two initial marginal prices, p_h^0 or p_l^0 . Instead, the shadow price is the value of p where demand crosses the closed graph of the supply schedule. This value of p_*^0 is indicated in Figure 3. Similarly, the second shadow price p_*^1 is the point where the compensated demand function $h(p, u^0)$ crosses the closed graph of the new price schedule. Note that equations (12)-(14) still suffice to determine the shadow prices and initial utility level in this setting; there is no need for the direct utility function, even though there is a range of marginal rates of substitution that all yield consumption where the marginal price changes.

Having found the two shadow prices, the calculation of CV proceeds as in the previous example. First, we compute the area under $h(p, u^0)$ between the two shadow prices. We then add to this the rectangular area $d(p_*^0, u^0) = (p_*^0 - p_l^0)k^0$, and subtract the rectangular area $d(p_*^1, u^0) = (p_*^1 - p_l^1)k^1$. This leaves us with the compensating variation as the irregular

shaded region in Figure 3.

The analysis in this case with a two-step increasing price schedule has an interesting connection to research on the welfare consequences of rationing. Neary and Roberts (1980) and Latham (1980) present modern treatments, although this line of work evidently begins with Robarath (1941). The connection to rationing arises from interpreting the supply schedule for a rationed good as perfectly elastic until some limit, at which point supply turns vertical. If a consumer's demand is actively rationed, then its demand curve will cross the vertical segment of supply in a manner analogous to that in Figure 3. The indicated shadow price, p_{*}^0 , then equals the LaGrange multiplier on the supply constraint in the consumer's problem. (This case motivated our adoption of the shadow price terminology). The analysis of the welfare consequences of rationing becomes a special case of the general methods discussed in section 3.

4.3 Menus of Pricing Plans

In many settings, consumers face not a single nonlinear price schedule but rather a *menu* of pricing plans. Common examples include mobile phone service plans, membership levels at museums and nonprofit organizations, and deductible choices for insurance. Menus of prices can present the consumer with an effective schedule of marginal prices that is not monotonic in quantity. Here we illustrate how the methods of section 3 can be applied when consumers face a menu of prices.

Consider the example in Figure 4. Here the consumer faces a choice of one of three pricing plans, where each solid line in Figure 4 represents a different plan. Note that in this figure the vertical axis measures each plan's total price, *not* the marginal price. Plan *A* has a minimum fee F_A , allows up to k_A units to be consumed at a marginal price per unit of zero, and then imposes a positive marginal price p_A for each unit beyond k_A . Plan *B* has a similar structure, but with higher minimum fee F_B , a larger allowance k_B , and a lower marginal price p_B beyond this allowance. Plan *C* has a simpler structure: A higher minimum fee F_C than the other plans, but a marginal price of zero for all units consumed. As Figure 4 suggests, Plan *C* appeals to relatively high demand consumers, Plan *B* to moderate demand consumers, and Plan *A* to consumers with the lowest demand for this service.

This pricing scheme is used by a major European internet service provider for internet access in the residential market (quantity is in megabits downloaded per month). Lambrecht, Seim, and Skiera (2005) present an empirical analysis of consumer demand facing the price menu in Figure 4. The interesting consumer welfare question arises from the striking contrast between this form of internet access pricing and the flat-fee system that prevails in the United States. Thus, here we consider how the methods of section 3 can be used to evaluate the consumer welfare consequences of price discrimination in this form.

Simplifying Lambrecht *et al.* slightly, assume that the consumer's ordinary demand for this service—if available at a constant marginal and average price p —takes the linear form

$$q = \alpha + \beta p + \gamma y.$$

Solving Roy's Identity and inverting on y gives the expenditure function that applies under uniform prices,

$$e(p, u) = u \exp(\gamma p) - \frac{1}{\gamma} (\alpha + \beta p + \beta/\gamma), \quad (16)$$

$\gamma \neq 0$, and differentiating on p yields

$$h(p, u) = \gamma u \exp(\gamma p) - \beta/\gamma. \quad (17)$$

Now consider how to evaluate the change in income that would leave the consumer indifferent if its initial price was replaced by the set of pricing plans in Figure 4. Suppose that the initial pricing is a simple fixed fee, F , with no additional charge per unit used. In this case the initial shadow price p_*^0 equals the marginal price of zero, and the consumer's initial utility level is readily obtained by inserting (16) into the 'initial income equals initial expenditure' condition (13) and solving for u^0 , with $d(p_*^0, u^0) = F$:

$$u^0 = (y - F) + \frac{1}{\gamma} (\alpha + \beta/\gamma). \quad (18)$$

To calculate the compensating variation when the initial fixed fee pricing is replaced with the menu, it remains to find the new shadow price p_*^1 where demand $h(p, u^0)$ intersects supply. For this purpose we assume that the consumer chooses the expenditure minimizing plan and consumption that achieves utility u^0 . Because the *marginal* prices (the slopes along the lower envelope) of the price schedules in Figure 4 vary non-monotonically and discontinuously, the calculations are simplest to convey in a table. The five rows in Table 1 partition all possible quantities into intervals according to plan and marginal price. These five intervals are indicated in the second column of Table 1 (and in Figure 4), in terms of the allowance quantities k_A, k_B , and the threshold quantities τ_A, τ_B that separate expenditure-minimizing plans.

The fourth column of Table 1 shows the expenditure-minimizing consumption level that achieves u^0 , but assuming each interval's marginal price is constant for all units consumed. These are obtained from (17), with u^0 in (18). Here there are only three distinct values of $h(p, u^0)$, as there are only three distinct marginal prices under the plans being offered.

To determine the shadow price p_*^1 , we now check each interval (row) in Table 1 to see if demand exceeds supply or supply exceeds demand. This is done in the last column

TABLE 1
COMPENSATING VARIATION WITH A MENU OF PRICES

Plan	Quantity Interval	Marginal Price	Income-compensated Demand, $h(p, u^0)$	Excess Demand or Supply?
A	$0 < q \leq k_A$	0	$\gamma u^0 - \beta/\gamma$	<i>Excess demand: $h(0, u^0) > k_A$</i>
	$k_A < q \leq \tau_A$	p_A	$\gamma u^0 - \beta/\gamma - \exp(\gamma p_A)$	<i>Excess demand: $h(p_A, u^0) > \tau_A$</i>
B	$\tau_A < q \leq k_B$	0	$\gamma u^0 - \beta/\gamma$	<i>Excess demand: $h(0, u^0) > k_B$</i>
	$k_B < q \leq \tau_B$	p_B	$\gamma u^0 - \beta/\gamma - \exp(\gamma p_B)$	<i>Neither; $k_B < h(p_B, u^0) \leq \tau_B$</i>
C	$\tau_B < q$	0	$\gamma u^0 - \beta/\gamma$	<i>Excess supply: $h(0, u^0) < \tau_B$</i>

of the table; it evaluates whether the quantity demanded (in column 4) falls within the admissible interval in column 2. For the sake of illustration we have assumed that this occurs in fourth interval (row), which implies the shadow price is $p_*^1 = p_B$.

The compensating variation now follows immediately from applying (10a) using (16):

$$CV = y - u^0 \exp(\gamma p_B) - \frac{1}{\gamma} (\alpha + \beta p_B + \beta/\gamma) - [F_B - k_B p_B].$$

The final term in square brackets is the difference in inframarginal expenditure, $d(p_*^1, u^0)$, with Plan B relative to a constant (marginal and average) shadow price p_B :

$$d(p_*^1, u^0) = [F_B + (q - k_B)p_B] - q p_B = F_B - k_B p_B.$$

Although the calculations summarized in Table 1 might appear a formidable procedure at first, they are easy to implement and evaluate by computer. The same procedure can be easily extended to situations where the price structure has more marginal prices, or more plans are offered, and so on. Calculating the equivalent variation is slightly more involved in this (particular) example, because we have to do calculations analogous to Table 1 to find the reference utility level as well as the new shadow price simultaneously. This extension is straightforward, however, and employed in section 6.

There are two technical points that merit note. Both mirror situations illustrated earlier in sections 4.1 and 4.2. First, it is possible for compensated demand to match supply on more than one interval, rather than one as indicated in Table 1. This outcome is analogous to the situation depicted in Figure 2, where $h(p, u^0)$ crosses the marginal price schedule at two distinct points. As in that example, the values of $h(p, u^0)$ will differ at the each candidate shadow price, and this situation is resolved by selecting from among these the one with the minimal total expenditure (under the actual, nonlinear price schedule(s)). This is a simple step in practice.

The second possibility parallels the situation in Figure 3. In this event, is possible that

demand exceeds supply (strictly) *or* supply exceeds demand (strictly) for each interval (row) in Table 1. This occurs when $h(p, u^0)$ crosses supply at a ‘gap’ where the marginal price abruptly (*i.e.* discontinuously) increases. When this occurs, the shadow price will lie between two marginal prices and $h(p, u^0)$ will equal their common interval boundary. The shadow price in this situation is determined by the value of p that sets $h(p, u^0)$ equal to this common interval boundary, as indicated by p_*^1 in Figure 3.

In sum, this general procedure would appear to provide a useful and straightforward means to evaluate exact consumer surplus in applied work when a consumer faces a menu of prices. Nevertheless, the reader may note at this point that our analysis of this problem remains incomplete. So far, we have considered how to evaluate exact consumer surplus for a single consumer. In contrast, the price discrimination question that motivated this subsection asked how to determine whether consumers—as a whole—are better off under one pricing scheme versus another. Such consumers are necessarily heterogeneous in their preferences; it would not be worthwhile for a seller to employ a menu of prices if otherwise. This raises the question of how to integrate the preceding analyses into empirical models that accommodate unobserved heterogeneity in preferences. We address this next.

5 Welfare with Heterogeneous Preferences

We now turn from evaluating welfare for a single consumer with known preferences to settings more typical of applied work. In practice, heterogeneity in willingness-to-pay among consumers is an important feature of demand behavior, and a component of this variation is inevitably due to factors the researcher does not observe. In this situation, the best we can do is to evaluate expected welfare given the characteristics of consumers that are observed.

Applied welfare analyses are usually based on econometric models of (ordinary) demand. Such models represent an individual consumer’s demand by a function of the form $q(p, y, z, \nu, \theta)$ where p is price, y income, z a set of observable characteristics, ν unobservable characteristics, and θ a set of parameters. Note that this form of demand corresponds to behavior facing a uniform price, p , that does not vary with quantity. If the consumer faces nonlinear prices that change from \mathcal{P}_0 to \mathcal{P}_1 , then application of the methods in sections 3 and 4 will yield an expression for exact consumer surplus that depends on the same arguments: $CV(\mathcal{P}_0, \mathcal{P}_1, y, z, \nu, \theta)$. To evaluate expected welfare for a single consumer, given y and z , we then need to integrate this expression against the distribution of ν .

This logic extends naturally to evaluating the sum of all compensating (or equivalent) income changes necessary to make a heterogeneous population of consumers indifferent

to the price changes.¹³ In the compensating case, for example, with a heterogeneous population of N consumers this sum is determined by integrating $CV(\cdot)$ against the joint distribution $P_{\nu,y,z}$ of all observed and unobserved characteristics in the population:

$$N \int_{\{\nu,y,z\}} CV(\mathcal{P}^0, \mathcal{P}^1, y, z, \nu, \theta) P_{\nu,y,z}(d\nu dy dz) . \quad (19)$$

Although this expression appears formidable, the integrals involved are inexpensive to compute using simulation methods. Geweke (1996) and Judd (1999) discuss standard techniques.

The importance of integrating over consumers' characteristics when evaluating welfare changes in a heterogeneous population is well noted in the literature.¹⁴ Nonetheless, it remains common empirical practice to evaluate welfare, even when income effects are present, using 'representative consumer' techniques that ignore preference heterogeneity altogether (see Slesnick's 1998 survey). The most common such approach amounts to ignoring the integral signs in (19) and simply evaluating $CV(\cdot)$ at the mean values of consumers' observed and unobservable characteristics. Because these characteristics commonly enter $CV(\cdot)$ in a nonlinear fashion, however, representative consumer methods provide welfare measures that typically differ from results based on (19). The standard concern is that because they ignore heterogeneity in preferences, representative consumer approaches may yield biased estimates of consumer welfare changes.

Here we wish to emphasize another concern that exacerbates this bias, and arises specifically when consumers face nonlinear prices. When two consumers with different preferences face the same nonlinear price schedule, they may choose quantities with different marginal prices (or, equivalently, choose different pricing plans from a common menu). This sorting response to nonlinear pricing is not unusual; rather, it is precisely why sellers find nonlinear pricing a profitable endeavor. Yet it implies the two consumers' compensating (or equivalent) variation values will differ not only because of their underlying preferences, but also because they face different price changes at the margin. Any representative consumer approach will assume all individuals face the same marginal price—thus mis-specifying the price changes that heterogeneous consumers actually experience. The self-selection of different marginal prices (or pricing plans) by heterogeneous consumers introduces a form of error that will not 'average out' if welfare is evaluated using a representative consumer approach.

¹³There is an extensive literature concerned with the conceptual foundations of aggregate welfare comparisons with heterogeneous individuals (see Blundell, Preston, and Walker (1995) or Slesnick (1998) for surveys). We assume that the total payments necessary to make all individuals indifferent to a price or policy change (that is, the sum of each individuals' compensating or equivalent variation values) is a matter of direct interest.

¹⁴See McFadden (1999) for a useful discussion in the context of discrete choice models. Fullerton and Gan (2004) explore this issue in continuous models using nonlinear transfer program examples.

The methods for evaluating consumer welfare with nonlinear prices described earlier allow one to avoid this problem. That is, one first evaluates each consumer's compensating (or equivalent) variation using the techniques in sections 3 and 4, contingent on the consumer's characteristics. Population welfare changes can then be obtained by aggregating these values, according to (19).

In sum, if consumers face nonlinear prices it appears to be particularly important to account for heterogeneity in preferences when evaluating welfare. Of course, how badly one might go awry using a representative consumer approach (or other welfare approximation method) is an empirical matter. Accordingly, it is useful to consider an application that can convey some magnitudes in a practical context.

6 Application

The pricing of wireless phone service has changed substantially over the past two decades, with consumers facing ever richer sets of pricing menus. Consumer welfare questions related to these changes arise in a number of areas, including the efficiency costs of wireless excise taxes (Ingraham and Sidak 2004, Gao et al. 2005), the consumer welfare losses from regulatory actions (Hausman 1997), and the construction of exact price indices for this service. In this section, we illustrate how to evaluate the consumer welfare gain from a richer menu of pricing plans in applied work. As noted at the outset of this paper, finer levels of price discrimination tend to benefit sellers but it is theoretically ambiguous whether they make consumers—as a whole—better or worse off. The numerical results below indicate how the methods in sections 3 and 4 can be applied in a complex multi-part pricing plan setting. Our calculations contrast an exact answer that recognizes observed and unobserved consumer heterogeneity with cruder representative agent and Marshallian consumer surplus approximations.

Table 2 details the terms of two sets of national mobile phone calling plans. The terms of these plans closely approximate those offered by a major wireless carrier in the Washington, D.C. area in 2002 and 2003.¹⁵ In 2003 this firm added two new calling plans, and changed the fixed monthly fee, allowance minutes, and marginal price (for minutes above the allowance) on all four prior plans. Inspection of Table 2 reveals that, from a consumer's standpoint, the 2003 plans do not uniformly dominate, nor are dominated by, the smaller previous choice set. More work needs to be done to draw welfare conclusions, as the change in welfare depends on the individual's consumption, price sensitivity, and marginal utility of income.

¹⁵The plans in Table 2 differ from this firm's actual offerings slightly, in order to preserve anonymity and to help illustrate the methodological points below.

TABLE 2
A WIRELESS PHONE CALLING PLAN MENU, 2002 AND 2003

Plan Name	Fixed Fee (\$ per month)	Allowance (min. per month)	Marginal Price above Allowance (\$ per min.)
<i>2002 Plans:</i>			
National 500	69.99	500	.35
National 700	89.99	700	.35
National 1000	109.99	1000	.25
National 1300	139.99	1300	.25
<i>2003 Plans:</i>			
National 400	69.99	400	.40
National 600	79.99	600	.40
National 800	89.99	800	.40
National 1100	119.99	1100	.40
National 1400	149.99	1400	.40
National 1800	179.99	1800	.40

A Model. Evaluating overall consumer welfare changes requires a model of preferences and how they vary across consumers. Suppose that, facing a constant (average and marginal) price p , individual i 's demand for wireless service has the usual linear form

$$q_i = \alpha_i - \beta_i p + \gamma_i y_i,$$

where q_i is consumption (in minutes per month) and y_i income. This specification, or minor variations of it (log or semi-log form), is widely used in econometric studies of telecommunications demand including wireless service (Hausman 1997, 2000, Ingraham and Sidak 2004, Iyengar 2005, Lambrecht et al. 2005, Rappoport et al. 2004). In the two-good framework typically employed (explicitly or implicitly) in empirical work, this demand function corresponds (facing a constant price p) to a partial direct utility function of the form

$$u_i(q, x) = \frac{\gamma_i q - \beta_i}{\gamma_i^2} \exp\left(\gamma_i \frac{\alpha_i + \gamma_i x - q}{\gamma_i q - \beta_i}\right),$$

where x is expenditure on the composite outside good (serving as a price numeraire). To capture demand variation, we follow Iyengar (2005) and others by assuming the parameters α_i , β_i , and γ_i vary across the population in a parametric way. Specifically, we assume $(\alpha_i, \beta_i, \gamma_i)' \sim N(\mu, \Sigma)$. We select parameters of this distribution to match aggregate statistics on wireless service consumption, as described in Appendix B. This model implies mean price and income elasticities of $-.44$ and 0.51 , respectively (when evaluated at a constant price of \$0.25 per minute). These are broadly consistent with the elasticity estimates

reported in econometric studies of mobile phone use.¹⁶ In addition, average consumption at the prices in Table 2 (approximately 600 minutes per month) matches national wireless plan use.

Calculating Welfare Changes. Now consider how to evaluate consumer welfare when a seller changes the set of available pricing plans. The compensating change in income that would leave a consumer just as well off facing the 2003 plan options as she would be with the 2002 plan choice set is, from (10a),

$$CV(\mathcal{P}^0, \mathcal{P}^1, y_i, \nu_i) = y_i - e(p_*^1(y_i, \nu_i), u^0(y_i, \nu_i)) - d(p_*^1(y_i, \nu_i), u^0(y_i, \nu_i))$$

where $\nu_i = (\alpha_i, \beta_i, \gamma_i)$ denotes the unobserved parameters that vary across individuals. The shadow price p_*^1 and initial utility level u^0 depend upon these preference parameters as well as the consumer's income y . The appropriate values of p_*^1 and u^0 equate the consumer's 'supply and demand' conditions (12)-(14), as before. For the population overall, the compensating change in income (per capita) that makes consumers indifferent between the two sets of pricing plans is then

$$\int_{\{y, \nu\}} CV(\mathcal{P}^0, \mathcal{P}^1, y, \nu) P_{y, \nu}(dy d\nu) . \quad (20)$$

Positive values of this integral indicate that consumers are better off overall; that is, they would prefer (on average) the richer choice of plans in 2003 to the limited set offered in 2002.

We calculate this exact welfare measure using Monte Carlo integration. Specifically, we draw at random a consumer's preference parameters $\nu_i = (\alpha_i, \beta_i, \gamma_i)$ and income y_i . The income draws are from a log-normal approximation to the 2000 Census PUMS income data (see Appendix B). For these draws we calculate the consumer's initial utility level $u^0(y_i, \nu_i)$ and initial shadow price under the 2002 plans using (12) and (13); this step parallels the example in 4.3. The 2003 shadow price, $p_*^1(y_i, \nu_i)$, is determined similarly as the solution to (14). The difference in inframarginal expenditure, $d(p_*^1(y_i, \nu_i), u^0(y_i, \nu_i))$, is then obtained directly from the initial price schedules. The expenditure function under uniform pricing, $e(p_*^1(y_i, \nu_i), u^0(y_i, \nu_i))$, is given by (16). Each consumer's compensating variation, $CV(\mathcal{P}^0, \mathcal{P}^1, y_i, \nu_i)$ is then readily determined from (10a).

We repeat this process 20,000 times to evaluate (20). We use 20,000 draws because the simulation error at that level is effectively negligible. The entire process takes approximately five seconds on a 2 GHz computer.

¹⁶Gao et al. 2005, Hausman 1997, 2000, Ingraham and Sidak 2004, or Rappoport et al. 2004. All of these studies appear to evaluate elasticities at average prices, analogous to those reported above. For consumption information, see for example www.telephia.com.

Alternative Measures. To illustrate the consequences of heterogeneity discussed in section 5, we also evaluate three alternative welfare measures commonly encountered in the literature. These are:

1. *CV assuming a representative consumer, with nonlinear pricing:* $CV(\mathcal{P}^0, \mathcal{P}^1, \bar{y}, \bar{v})$. Here a representative consumer is assumed to have the ‘mean’ tastes and characteristics of the heterogeneous consumers in the population.
2. *CV assuming a representative consumer, but ignoring nonlinear pricing:* $CV_{\text{lin}}(\bar{p}^0, \bar{p}^1, \bar{y}, \bar{v})$. Here the representative consumer is assumed to have ‘mean’ tastes and characteristics of the heterogeneous consumers in the population, but nonlinear pricing is ignored: The consumer is assumed, incorrectly, to face a linear budget constraint at the average price (in the population) for this service.
3. *Change in Marshallian consumer surplus, using average prices:* $CS_{\text{lin}}(\bar{p}^0, \bar{p}^1, \bar{y}, \bar{v})$. This is the area under the ordinary (Marshallian) aggregate demand curve, evaluated between the average prices (in the population) before and after the price changes.

The first of these alternative welfare metrics avoids having to evaluate the integral in (20) that arises when welfare is evaluated with heterogeneous consumers. The second alternative welfare metric, which assumes a representative consumer and incorrectly assumes that consumer faces a constant (marginal and average) price, is of interest because it is easily implemented when a researcher has only aggregate price and quantity data available. For this reason, this welfare measure is commonly encountered in the literature (e.g., Hausman 1997, Ingraham and Sidak 2004). We evaluate the third measure, the change in Marshallian consumer surplus, because it remains a widely employed metric for assessing consumer welfare changes in applied economic research.

Discussion. Table 3 shows the numerical results we obtain for exact compensating variation and these three approximations. Each row assumes the true distribution of preferences in the population follows the model of heterogeneous demand behavior described earlier. The first row is the mean compensating variation for the population, as given by equation (20). It indicates that, on average, consumers are better off under the 2003 choice set than under the 2002 choice set by \$17.76 per user per year.

Because consumers face price discrimination based on individuals’ demand intensities, it is natural to expect the three representative consumer-based welfare approximations to diverge from the exact calculation. The second row of Table 3 represents what one might calculate in order to accommodate a menu of nonlinear prices in a representative consumer framework. This measure overestimates the mean consumer welfare gain with a substantial error of +532 percent. To appreciate why, note that the typical consumer

TABLE 3
EXACT CONSUMER SURPLUS AND THREE COMMON APPROXIMATIONS

Method	Formula	Welfare Change	Approximation Error
Exact compensating variation, per capita	$\int CV(\mathcal{P}^0, \mathcal{P}^1, y, \nu) dP_{y,\nu}$	\$ 17.76	—
Representative consumer approximation, with nonlinear prices	$CV(\mathcal{P}^0, \mathcal{P}^1, \bar{y}, \bar{\nu})$	\$ 112.20	+532%
Representative consumer approximation, with average prices	$CV_{\text{lin}}(\bar{p}^0, \bar{p}^1, \bar{y}, \bar{\nu})$	\$ -2.64	-114%
Marshallian consumer surplus with average prices, per capita	$CS_{\text{lin}}(\bar{p}^0, \bar{p}^1, \bar{y}, \bar{\nu})$	\$ -2.64	-114%

uses just over 600 minutes per month and would prefer the 700-minute plan under the initial 2002 choice set, at \$89.99 per month (see Table 2). Under the 2003 plans, the typical consumer would instead prefer the 600-minute plan for \$79.99 and incur small overage charges. Since the typical consumer's demand is fairly price inelastic the net gain is slightly less than the fixed fee decrease of \$10 per month, amounting to \$112.20 per annum. In sum, by ignoring heterogeneity in preferences and consumption, this method mis-specifies the marginal price that most consumers face and poorly estimates consumer welfare.

The third line of Table 3 table reports a conventional representative consumer compensating variation calculation, one that ignores the nonlinearity of prices entirely. It is meant to mimic what a researcher might do if only the consumer's total monthly bill is observed (i.e., the number of minutes used and the dollar amount billed). Here we use the seller's average revenue per minute (over all consumers) as the consumer's price, and assume a linear budget constraint. This method makes an approximation error of -114%, which has the wrong sign. This misleading result occurs because the representative consumer's average price rises slightly from 2002 to 2003 (by approximately four hundredths of a cent per minute); at average use of roughly 600 minutes per month and a price-inelastic representative consumer, we obtain a loss of \$2.64 annually. This welfare loss conclusion is incorrect, and arises from two different sources of aggregation bias: it incorrectly assumes that consumers face the average price on the margin, and it ignores demand heterogeneity by adopting a representative consumer framework.

The final row in Table 3 reports the consumer surplus measure used in a classic 'Harberger' deadweight welfare triangle calculation. This measure of welfare has three sources of error: It ignores income effects, it aggregates heterogeneous individuals consumer welfare changes incorrectly, and it incorrectly employs average instead of marginal prices

when consumers face nonlinear pricing. In this particular case, there turns out to be little difference between this measure of consumer welfare and the previous average price measure: For this small change in average prices and with income-inelastic mean demand (of .5), there is little difference between the representative consumer's Hicksian and Marshallian demand curves. The central point to observe here is that even if income elasticities are small (on average), Marshallian consumer surplus changes can yield poor estimates of actual (exact) welfare changes when consumers face nonlinear prices.

Overall, these results indicate that common approximation methods for evaluating consumer welfare changes can yield substantial errors. These errors would appear to be a potential concern whenever the population of interest is subject to a nonlinear pricing scheme that exploits the heterogeneity in consumers' preferences. Because they ignore the (induced) price variation across consumers with different demand levels, representative consumer approximations should be expected to be somewhat inaccurate in these settings. While these errors could turn out to be small in other applications, it remains a challenge to identify the circumstances in which a researcher can safely conclude as much *a priori*.

7 Concluding Remarks

This paper proposes a simplified method for evaluating exact consumer surplus when consumers face nonlinear prices, including menus of pricing plans. It also accommodates the preference heterogeneity that leads consumers to self-select along a menu of price schedules for unobserved reasons. Our approach complements and extends the prior methods used for special cases in the literature, including welfare analysis with rationed goods (Neary and Roberts 1980) and the piecewise-linear budget sets that arise in income tax change simulations (see, *e.g.*, Hausman 1983 or Moffitt 1990). In addition to generalizing these earlier methods to pricing practices of most any form, our technique is convenient to implement.

The techniques in this paper can be used to analyze a variety of welfare issues. For example, nonlinear prices and the need to evaluate welfare figure prominently in regulated markets and tax policy discussions. Most practical applications of this paper's methods will build upon econometric models of demand behavior, since—as we have indicated—it is important to account for the heterogeneity in individuals' preferences when consumers face nonlinear prices. Given a model of consumer demand, with these techniques a practitioner can *prospectively* evaluate consumers' well-being under any pricing system one wishes to consider.

These methods may also prove useful for addressing classical policy questions in situations where a consumer's choices affect the prices it faces. For instance, consider the famil-

iar problem of measuring the deadweight loss from an excise or ad valorem tax. While the classical 'welfare triangle' logic is clear when a good has a constant price, the deadweight loss is no longer so transparent when consumption decisions are made facing a menu of nonlinear prices. This paper's methods provide a general means to handle such problems, using the conventional (Diamond and McFadden 1974) definition of deadweight loss as the difference between total tax revenue and the total lump-sum transfers needed to leave each consumer indifferent to the tax. Evaluating the latter when consumers face nonlinear prices is an immediate application of the techniques presented here.

Appendix A.

This appendix provides additional detail regarding the determination of the shadow prices in Figure 2, as noted in section 4.1.

Consider the analytic determination of the initial utility level u^0 and shadow price p_*^0 in Figure 2. Inserting the expenditure function (15) into the ‘income equals initial expenditure’ condition (13) and rearranging gives

$$u^0 = \frac{1}{1-\gamma} [y - (p_h^0 - p_*^0)k^0]^{1-\gamma} - \frac{1}{1+\beta} p^{1+\beta} e^\alpha \quad (21)$$

This expresses the initial utility level as a function of the shadow price, p_*^0 . To determine the shadow price, note that the consumer’s initial marginal willingness to pay in this example is either the high or the low marginal price; which value applies depends whether the expenditure-minimizing consumption level is below or above the quantity k^0 where the price schedule changes. Equation (12) implies

$$p_*^0 = \begin{cases} p_h^0 & \text{only if } h(p_h^0, u^0) \leq k^0 \\ p_l^0 & \text{only if } h(p_l^0, u^0) > k^0. \end{cases}$$

This is a feasibility condition; in economic terms, it ensures that the quantity demanded at a particular marginal price will be supplied at that marginal price. The equality case assignment is determined by the terms of the price schedule (here assumed to be left-continuous at k^0). Now if $h(p, u^0)$ crosses the supply schedule more than once, then both cases can hold (thus the ‘only if’). This is the situation depicted in Figure 2. When this occurs, the correct shadow price is the value that minimizes the total expenditure required to achieve u^0 :

$$p_*^0 = \begin{cases} p_h^0 & \text{if } e(p_h^0, u^0) \leq e(p_l^0, u^0) + (p_h^0 - p_l^0)k^0 \\ p_l^0 & \text{if otherwise.} \end{cases} \quad (22)$$

The right-hand side is readily calculated using (21) and (15). Figure 2 is drawn such that the second case applies and the initial shadow price is p_l^0 . Note the equality case assignment in (22) is arbitrary because the consumer’s initial utility and total expenditure is the same either way.

Appendix B.

This appendix describes the parameterization of the model used in Section 6.

Distribution of Consumer Income. As noted in the text, consumer incomes are assumed to be log-normally distributed in the population. We estimated the parameters of this distribution using percentiles of the US household income data obtained from the One Percent Public Use Microdata Files of the 2000 Decennial Census. Mean monthly total income is approximately \$ 4,700.

Distribution of Consumer Preferences. The distribution of consumers’ preference parameters $\nu_i = (\alpha_i, \beta_i, \gamma_i)$ is obtained in two steps. First we fix the parameter means, using information from several sources. Telephia Customer Value Metrics (www.telephia.com) reports 2005-Q1 cell phone use averaged 679 minutes nationally. Discounting back at their estimated 5% p.a. growth rate implies average use of 600 minutes during 2002-2003. In the absence of overages, our demand

specification has $600 = \bar{\alpha} + \bar{\gamma}\bar{y}$, where the bars denote averages. We set $\bar{\gamma} = 0.065$ so as to produce an average income elasticity of 0.5, which is comparable to income elasticities reported in other telecommunications studies (see text). This elasticity implies $\bar{\alpha}$ is approximately 300 minutes per month. We then set the mean price coefficient $\bar{\beta}$ so that the reservation price for usage is \$1.50/minute. At a (uniform) price of \$0.40/minute this implies a mean price elasticity of -0.36 .

Consistent with prior econometric work, the distribution of the parameters $(\alpha_i, \beta_i, \gamma_i)$ are assumed to be normal and independent across consumers and parameters (see, e.g., Iyengar 2005). We estimate the variances by matching the model's moments to six sample consumption moments. Four moments are average consumption (in minutes) for the four 2002 plans (which are 450, 630, 900, and 1170 minutes per month). Two additional moments are the standard deviation of overall usage (300 minutes per month) and an average frequency of 0.20. These six sample consumption moments are based on related cell plan use statistics reported in Iyengar (2005).

Last, in our simulations we also impose, if needed, a constraint that an individual's price coefficient must yield quasi-concave utility at 1,000 minutes. This issue arises because a linear demand specification does not necessarily yield quasi-concave utility for all possible values of income and demand elasticities and because the support of these parameters is unbounded (this is a well-known problem with linear demand models; see Hausman 1981). For the parameters we employ, however, this problem is rarely encountered in our simulations.

References

- Bolmquist, S. (1983) "The Effect of Income Taxation on the Labor Supply of Married Men in Sweden." *Journal of Public Economics* 22: 169-97.
- Blundell, R. and T. MaCurdy (1999) "Labor Supply: A Review of Alternative Approaches." Chapter 27 in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics* Vol. 3. Amsterdam: North-Holland.
- Blundell, R., C. Meghir, and E. Symons (1988) "Labor Supply Specification and the Evaluation of Tax Reforms." *Journal of Public Economics* 36: 23-52.
- Burtless, G., and J.A. Hausman (1978) "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment." *Journal of Political Economy* 86: 1103-30.
- Creedy, J. and A. Duncan (2002) "Behavioral Microsimulation with Labor Supply Responses." *Journal of Economic Surveys* 16(1): 1-39.
- Creedy, J. and G. Kalb (2005) "Measuring Welfare Changes in Labor Supply Models." *The Manchester School* 73(6): 664-85.
- DeBorger, B. (1989) "Estimating the Welfare Implications of In-Kind Government Programs: A General Numerical Approach." *Journal of Public Economics* 38: 215-26.
- Diamond, P., and D. McFadden (1974) "Some Uses of the Expenditure Function in Public Finance." *Journal of Public Economics* 3: 3-21.
- Dubin, J. and D. McFadden (1984) "An Econometric Analysis of Residential Appliance Holdings and Consumption." *Econometrica* 52: 345-62.
- Fullerton, D. and L. Gan (2004) "A Simulation-Based Welfare Loss Calculation for Labor Taxes with Piecewise-Linear Budgets." *Journal of Public Economics* 88: 2339-59.
- Gao, M., A. Hyytinen, and O. Toivanen (2005) "Demand for Mobility: Evidence from a Real-World Pricing Experiment." Discussion Paper No. 43, Helsinki Center for Economic Research, University of Helsinki.
- Geweke, J. (1996) "Monte Carlo Simulation and Numerical Integration." Chapter 15 in H.M. Amman, D.A. Kendrick, and J. Rust, eds., *Handbook of Computational Economics*, Vol. 1. Amsterdam: North-Holland.
- Hausman, J.A. (1981) "Exact Consumer Surplus and Deadweight Loss." *American Economic Review* 71: 662-676.
- Hausman, J.A. (1983) "Stochastic Problems in the Simulation of Labor Supply." In M. Feldstein, ed., *Behavioral Simulation Methods in Tax Policy Analysis*. Chicago: University of Chicago Press.
- Hausman, J.A. (1985) "The Econometrics of Nonlinear Budget Sets." *Econometrica* 53: 1255-82.
- Hausman, J.A. (2000) "Efficiency Effects on the US Economy from Wireless Taxation." *National Tax Journal* 53: 733-9.
- Ingraham, A., and G. Sidak (2004). "Do States Tax Wireless Services Inefficiently? Evidence on the Price Elasticity of Demand." *Virginia Tax Review* 24(2): 249-62.
- Iyengar, R. (2005) "A Structural Demand Analysis for Wireless Services under Nonlinear Pricing Schemes." Working Paper, The Wharton School, University of Pennsylvania.
- Judd, K.L. (1999) *Numerical Methods in Economics*. Cambridge, MA: MIT Press.

- Lambrecht, A., K. Seim and B. Skiera (2005) "Does Uncertainty Matter? Consumer Behavior under Three-Part Tariffs." Manuscript, Univ. of Pennsylvania, Department of Business and Public Policy.
- Latham, R. (1980) "Quantity Constrained Demand Functions." *Econometrica* 48(2): 308-14.
- McFadden, D. (1999) "Computing Willingness To Pay in Random Utility Models." Chapter 15 in J. Moore, R. Riezman, and J. Melvin, eds., *Trade, Theory, and Econometrics: Essays in Honor of John S. Chipman*. London: Routledge.
- Maddock, R. (1989) "Welfare Analysis with Rising Block Pricing." *Resources and Energy* 11(4): 329-36.
- Moffitt, R. (1986) "The Econometrics of Piecewise-Linear Budget Constraints: A Survey and Exposition of the Maximum Likelihood Method." *Journal of Business and Economic Statistics* 4(3): 317-28.
- Moffitt, R. (1990) "The Econometrics of Kinked Budget Constraints." *Journal of Economic Perspectives* 4: 119-39.
- Neary, J.P., and K. W.S. Roberts (1980) "The Theory of Household Behavior Under Rationing." *European Economic Review* 13:25-42.
- Rappoport, P., J. Alleman, and L.D. Taylor (2004) "Household Demand for Wireless Telephony: An Empirical Analysis." Working Paper, Columbia Institute for Tele-Information, Columbia University.
- Reiss, P. and M. White (2005) "Household Electricity Demand, Revisited." *Review of Economic Studies* 72: 853-83.
- Roberts, K.W.S. (1979) "Welfare Effects of Nonlinear Pricing." *Economic Journal* 89(353): 66-83.
- Schwab, R.M. (1985) "The Benefits of In-Kind Government Programs." *Journal of Public Economics* 27: 195-210.
- Slesnick, D.T. (1998) "Empirical Approaches to the Measurement of Welfare." *Journal of Economic Literature* 36(4): 2108-65.
- Stole, L. (2003) "Price Discrimination and Imperfect Competition." Forthcoming in M. Armstrong and R.H. Porter, eds., *Handbook of Industrial Organization* Vol. 3. Amsterdam: North-Holland.
- Vartia, Y. (1983) "Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Market Demand Functions." *Econometrica* 51: 79-98.
- Varian, H. (1989) "Price Discrimination." Chapter 10 in R. Schmalensee and R.D. Willig, eds., *Handbook of Industrial Organization* Vol. 1. Amsterdam: North-Holland.
- Wilson, R. (1993) *Nonlinear Pricing*. Oxford: Oxford University Press.

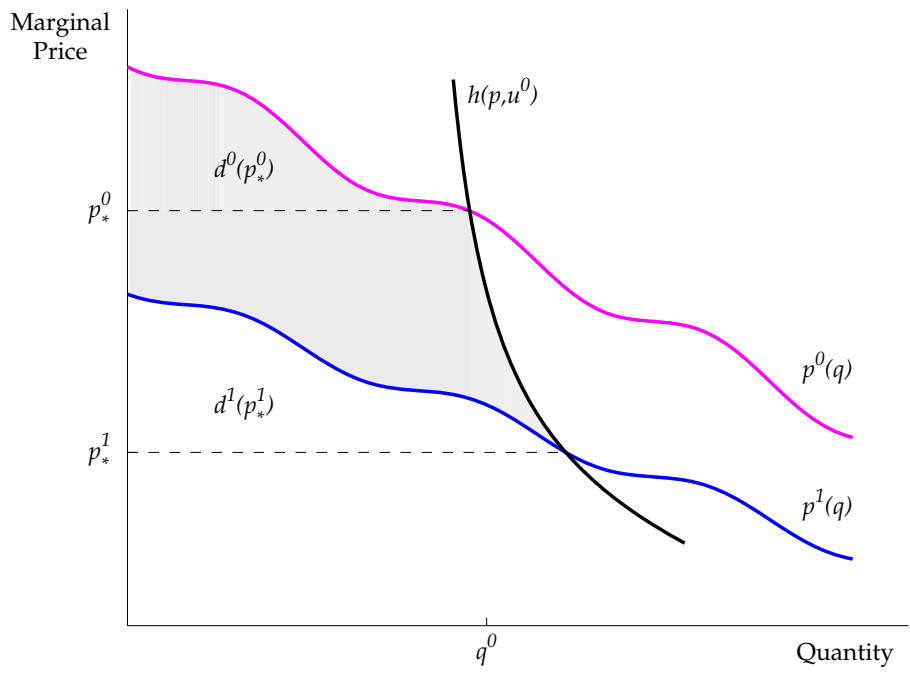


FIGURE 1

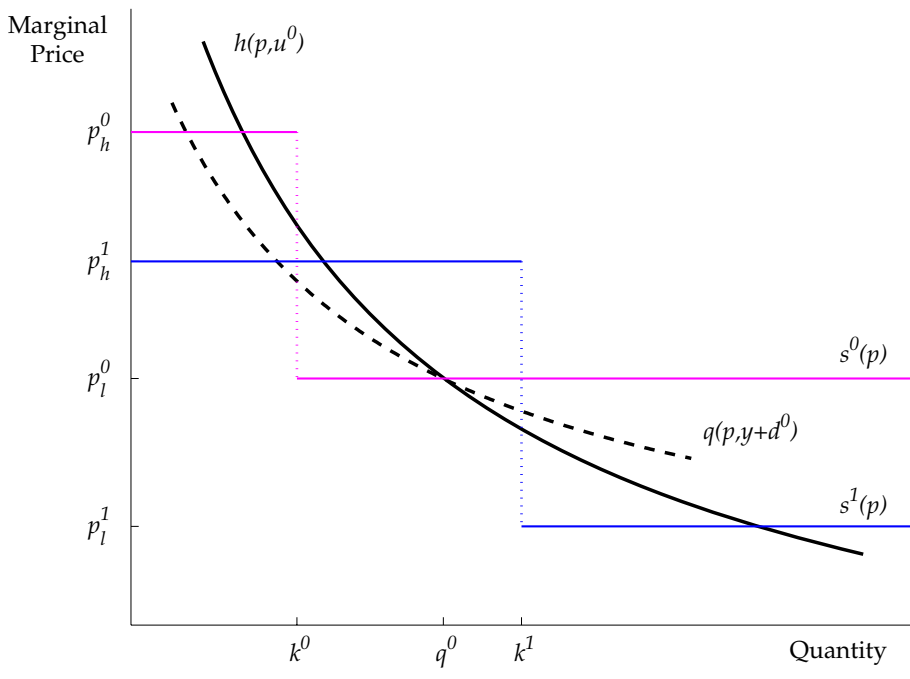


FIGURE 2

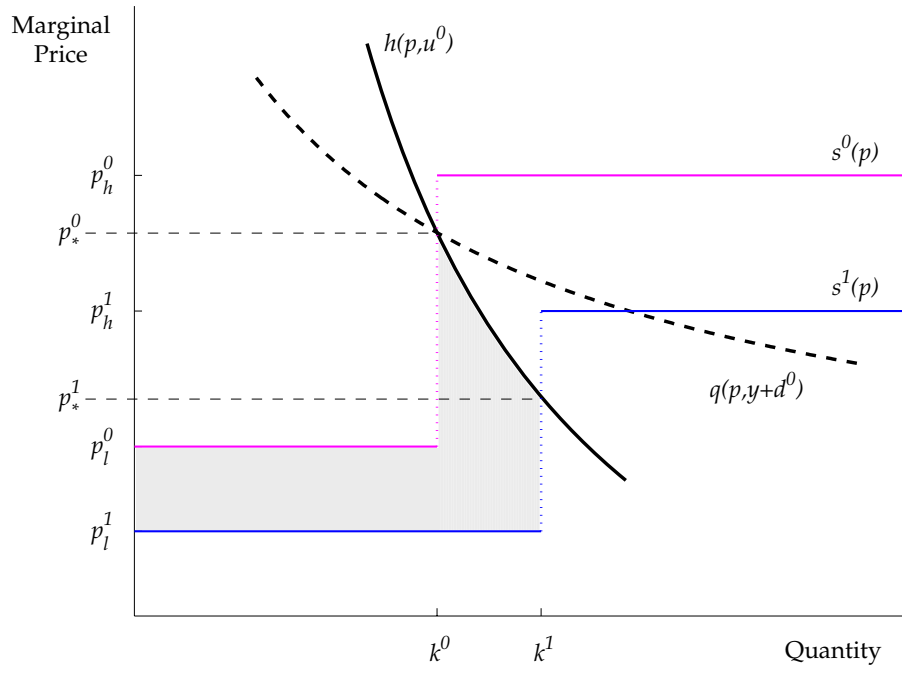


FIGURE 3

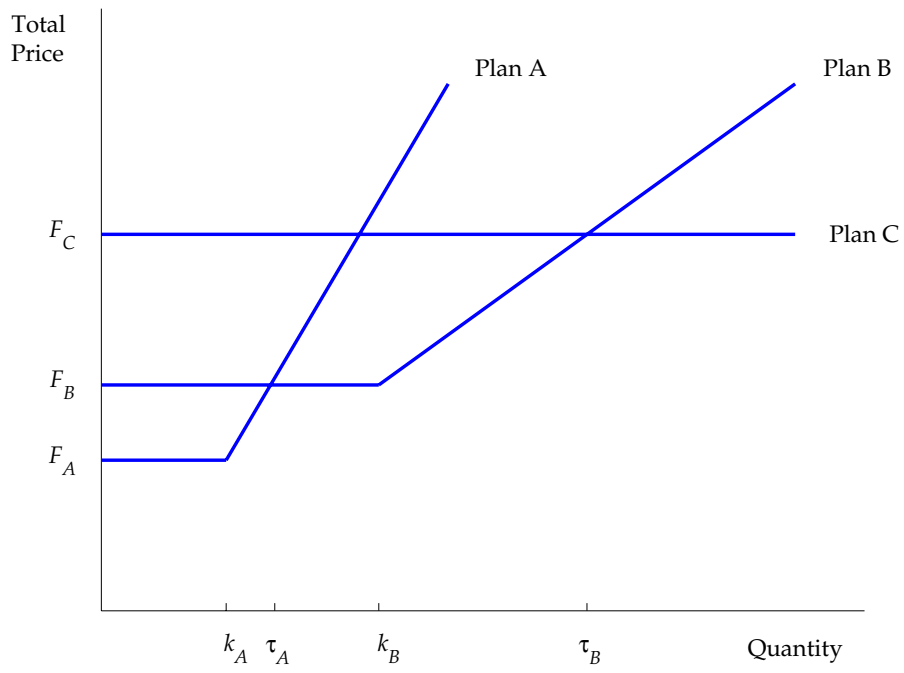


FIGURE 4