The Cost of Unsold Output, the Elasticity of Demand, and Prices

Michael A. Salinger*

January 2010

Abstract: When a firm must choose price and output before observing demand and therefore risks unsold output, the standard mark-up rule applies with the elasticity being the elasticity of the average quantity sold with respect to price and marginal cost being the marginal cost of an expected unit sold, computed as the marginal cost of a unit produced divided by the expected fraction of the marginal unit produced that the firm sells. The rule resolves the longstanding puzzle of why increased demand uncertainty causes output to increase with additive uncertainty and decrease with multiplicative uncertainty.

Keywords: newsvendor problem, demand uncertainty, mark-ups, pass through

Michael A. Salinger
Boston University
School of Management
595 Commonwealth Ave.
Boston, MA 02215
salinger@bu.edu
(617) 353-4408

* Professor of Economics, Boston University School of Management. I thank Miguel Ampudia, Mike Baye, Jeremy Bulow, Jim Dana, Dennis Carlton, Dan Hosken, Harish Krishnan, Makoto Hanazono, Nicholas Petruzzi, and Glen Weyl for helpful comments on an earlier version, (which was entitled “Mark-up Pricing Under Demand Uncertainty”).
I. Introduction

How does the cost of unsold output affect the price a firm sells? Under certainty, of course, the profit-maximizing level of output is where marginal revenue (accounting for the price reduction needed to increase demand) equals marginal cost. The assumption of certainty simplifies the problem in a fundamental respect, however. Marginal cost typically means the additional cost of producing an extra unit. Marginal revenue is the additional revenue from selling an additional unit. With certain demand, the distinction is unimportant because the quantity produced equals the quantity sold. If, however, demand is uncertain and the firm must choose price and output before observing demand, the firm risks producing output it cannot sell, a risk that it trades off with the risk of not being able to satisfy all demand.

The mark-up rule relating the price to the elasticity of demand and marginal cost follows directly from the condition that the profit-maximizing output is where marginal revenue equals marginal cost. Under certainty, the rule states:

\[ p = c \frac{\eta(p)}{1 + \eta(p)} \]

where \( p \) is price, \( q(p) \) is the demand relationship, \( c \) is marginal cost, and \( \eta(p) = pq'(p)/q(p) \) is the elasticity of demand facing the firm.\(^1\) The deep insight from the rule is that the optimal simple price rests on just two factors, marginal cost and the elasticity of demand (which determines the mark up over marginal cost). Again, the assumption of certainty glosses over a key economic issue. Does the elasticity of demand determine a mark-up over the marginal cost of producing an extra unit of output even if it is inevitable?

\(^1\) For a monopolist, the elasticity of demand facing the firm is also the market elasticity of demand. Interpreting the elasticity as the elasticity of residual demand, the rule applies to all market structures.
that the firm will not sell all its output, or does marginal cost include the cost of unsold output? Moreover, with demand uncertainty, the elasticity of demand can vary across states. If so, how does one measure the elasticity of demand for a formula like (1)?

This last question presumes that such a formula exists. Remarkably, however, the literature does not contain a formula like equation (1) for general forms of demand uncertainty. This paper derives such a formula.

How uncertainty affects pricing and output decisions is, of course, a broad issue about which an extensive literature already exists. The answer depends on the precise sequencing of decisions and the resolution of uncertainty. The problem addressed below is when the firm must choose both price and (completely perishable) output before knowing the state of demand, the problem called the “newsvendor problem with endogenous pricing” or the “price-setting newsvendor problem.” According to Petruzzi and Dada (1999), the literature on the problem dates back at least to Edgeworth, who addressed it in 1888 with respect to banks. The assumption of perfectly perishable output might create the appearance that it is a special case, but similar results arise if unsold goods can be held in inventory with a carrying cost. A stronger assumption is that the firm must charge a linear price that it cannot adjust after observing demand. Both non-linear and state-dependent prices fall into the realm of price discrimination, and one of the themes of the literature is that uncertainty can make such strategies profitable.

---

2 Notable contributions in the economics literature include Nevins (1966), Sandmo (1971), Leland (1972), Carlton (1978), and Pindyck (1982).

3 In the “simple” newsvendor problem, the price is given exogenously and the firm must choose its output before knowing demand. See Whitin (1955) and Karlin and Carr (1962).

4 Khouja (1999) and Petruzzi and Dada (1999) are excellent reviews of the literature.

5 For example, Dana (1999) shows that it is better to offer multiple prices with limited quantities available at the lower prices. This point is analogous to the point that, under certainty, second degree price discrimination is optimal. Stole (2007) provides such an interpretation. Carlton and Dana’s (2004)
Because the mark-up rule is based on the assumption of simple pricing, however, the price-setting newsvendor problem is the natural case to consider for extending it to uncertainty.

The standard solution to the price-setting newsvendor problem is to model the firm as having two choice variables, price and output, to take the partial derivative of the expected profit function with respect to each, and to solve the two first order conditions simultaneously. That approach, while valid, does not easily reveal an analog or extension of the mark-up rule. Under certainty, the single first order condition captures how simultaneous changes in price and output affect profits. In the standard solution to the newsvendor problem, each partial derivative by itself misses the interaction between the price and output decisions. That interaction is present in the simultaneous solution, but one cannot derive a characterization of the solution that corresponds to the mark-up rule from just one of these first order conditions.

A long-standing puzzle in the literature concerns how the effect of increased uncertainty affects price and output. Mills (1959) analyzed the price-setting newsvendor problem with additive uncertainty and showed that an increase in the degree of uncertainty causes a firm to lower price and expand output. The result might seem counterintuitive because, by increasing the expected fraction of unsold output (for any price-output combination), the increased uncertainty would seem to increase expected costs. One might expect the firm to respond by increasing price and cutting output. Although they did not resolve the paradox, Karlin and Carr (1962) showed that with multiplicative uncertainty, one gets the more intuitive solution that increased uncertainty
results in a higher price and lower output. The sensitivity of the comparative statics to an apparently technical assumption might have created the perception that with uncertainty, the optimal price does not depend just on marginal cost and the elasticity of demand.

But, with suitable generalizations of the meaning of marginal cost and the elasticity, the mark-up rule characterizes the optimal price for the price-setting newsvendor problem. Specifically, marginal cost is the marginal cost of an expected unit sold computed as the marginal cost of an extra unit produced scaled by a factor that reflects the fraction of the marginal unit produced that the firm expects to sell.\(^6\) The relevant elasticity is the elasticity of the average quantity sold with respect to price when the output change associated with a price change maintains the probability of selling out.

These adjustments rest on three key ideas. The first is the distinction between the marginal cost of a unit produced and the marginal cost of an expected unit sold. By definition, marginal revenue is the additional revenue from *selling* an additional unit. Marginal cost usually refers to the marginal cost of *producing* an additional unit. Profit-maximization entails setting marginal revenue (from an extra unit sold) to the marginal cost of an extra unit *sold*, which requires scaling the marginal cost of an extra unit produced by a factor that reflects the fraction of an extra unit produced that the firm sells on average.

The second key concept is the “highest profitable state for production” (HPSP). With uncertain demand and any given price, the firm must decide the states in which its production will be sufficient to satisfy all demand.\(^7\) Deriving the mark-up rule requires

---

\(^6\) Petruzzi and Dada (2009) point out this distinction for multiplicative uncertainty.

\(^7\) Whether one uses the traditional formulation (with output as a choice variable) or the alternative used in this paper (with the HPSP as a choice variable), one of the partial first order conditions characterizes the HPSP.
analyzing the effect of a change in price along with the associated change in output. A key to the solution is that the sensitivity of demand to price in the HPSP drives the interrelationship between price and output changes.

The third key concept is the elasticity of the average quantity sold with respect to price when the firm changes price and output so as to maintain the probability of selling out at the optimal HPSP. It differs from the elasticity of demand under certainty in two respects. First, because the sensitivity of demand to price varies across states, the relevant elasticity must reflect some averaging. Second, because the quantity sold does not always equal the quantity demanded, the elasticity of average demand is different from the elasticity of the average quantity sold. It is the latter that enters the formula for the optimal mark-up.

Remarkably, the extended mark-up rule explains the qualitative difference in the effect of increased uncertainty in the additive and multiplicative cases ceases to be a paradox. The difference comes down to how increased uncertainty affects the marginal cost of an expected unit sold and the elasticity of the average quantity sold in the two cases. Even though increased additive uncertainty increases average cost, it does not affect the marginal cost of an expected unit sold. It does, however, increase the elasticity of the average quantity sold and therefore reduces the mark-up. In contrast, increased multiplicative uncertainty increases the marginal cost of an expected unit sold but has no affect on the elasticity of the average quantity sold or, therefore, the mark-up factor.

---

8 Because of the distinction between the quantity demanded and the quantity sold, the elasticity of the quantity sold with respect to price depends on how the firm changes output when it changes price. Thus, the elasticity of the average quantity sold with respect to price when the firm holds output constant is different from the elasticity in the mark-up rule derived below.
What is remarkable is that the formula resolves the paradox that might have caused doubt as to whether such a formula exists.

Under certainty when demand has a constant elasticity form, the ratio of price to marginal cost implied by equation (1) is also the pass-through rate, i.e., the derivative of price with respect to marginal cost. As pointed out by Bulow and Pfleiderer (1983), however, the pass-through rate is sensitive to the functional form of the demand curve. It is 0.5 for linear demand and 1 for log-linear demand. More recently, Weyl and Fabinger (2009) have derived a general principle relating the pass-through rate to the curvature of the demand curve. Specifically, they show the pass-through rate is greater than 1 when the demand curve is log-convex and less than one when it is log-concave. This paper will derive formulas for pass-through rates in the price-setting newsvendor problem. With uncertainty, the pass-through rate is not as a general matter a function of just the curvature of the demand curve around the optimum.9

The remainder of the paper is organized as follows. Section II analyzes the price and output decision with only two states of demand (which easily generalizes to any number of discrete states). The discrete case provides the intuition that underlies the solution for a continuum of demand states, which is covered in Section III. Section IV contains concluding comments.

II. Two Demand States

The simplest form of uncertainty, just two discrete demand states, provides most of the insight into how demand uncertainty affects the mark-up rule and the pass-through rate in the more general version of the price-setting newsvendor problem. Let $q(p)$

---

9 As Weyl and Fabinger recognize, their result that pass-through depends just on demand curvature rests on the strong technological assumption of constant marginal costs. Relaxing that assumption would likely lead to the conclusion that pass-through depends on cost-parameters as well.
represent demand in the high demand state, \( r(p) \) be the amount by which demand is reduced in the low demand state, and \( \phi \) be the probability the high demand state occurs. Assume \( 0 \leq r(p) \leq q(p) \) for all \( p \). Additive uncertainty implies \( r(p) \) is a constant.\(^{10}\) Multiplicative uncertainty means that \( r(p)/q(p) \) is constant. For simplicity, assume that marginal production cost, \( c \), is constant.

**A. Solution**

We first establish the following lemma.

**Lemma:** Let \( \pi \) be profit, \( x \) be the quantity produced, and \((x^*, p^*)\) be the production level and price that maximize expected profits. Then, \( x^* = q(p^*) \) or \( x^* = q(p^*) - r(p^*) \).

**Proof:** The expected profit function is:

\[
E[\pi] = \phi p \min[q(p), x] + (1-\phi) p \min[q(p)-r(p), x] - cx
\]

The first order condition with respect to \( x \) has three regions to it:

\[
\frac{\partial E[\pi]}{\partial x} = p - c \quad x < q(p) - r(p)
\]

\[
\frac{\partial E[\pi]}{\partial x} = \phi p - c \quad q(p) - r(p) \leq x < q(p)
\]

\[
\frac{\partial E[\pi]}{\partial x} = -c \quad q(p) \leq x
\]

If \( \phi < 1 \), then \( p^* > c \). Otherwise, expected profits would be negative. The first of the three derivatives is therefore positive, so \( x^* \geq q(p^*) - r(p^*) \). Since \( -c < 0 \), \( x^* \leq q(p^*) \).\(^{11}\)

---

\(^{10}\) The difference would only be constant in the range where \( q(p) \) exceeds the constant. For higher prices, low demand would be 0.

\(^{11}\) These two results meant that the firm never produces less than the minimum demand could be nor more than the maximum it could be.
Within the middle region, the derivative is constant. If it is positive, then \( x^* = q(p^*) \) maximizes expected profits. If it is negative, \( x^* = q(p^*) - r(p^*) \) does.\(^{12}\) qed

The lemma captures a key feature of the economics of the newsvendor problem discussed in the introduction, and it provides an essential insight into how to extend the mark-up rule to account for uncertainty. Deriving the mark-up rule requires capturing how much the firm expands output when it lowers its price. Since, with uncertainty, the sensitivity of demand to price can vary across states, the relationship between price and output is less obvious than when demand is certain. The lemma indicates that with two demand states, production is geared toward one of them, i.e., the HPSP.\(^{13}\) As a result, the output change associated with a price change that does not alter the HPSP is driven by the sensitivity of demand to price in that state.

When the low demand state is the HPSP, the problem reduces to one of certainty. Thus, the interesting case to consider is when the HPSP is the high demand state, in which case \( x = q(p) \). A fundamental aspect of the economics of the price-setting newsvendor is that, given price, demand in the HPSP -- \( q(p) \) in this case -- determines cost but the average quantity sold -- \( \bar{q}(p) = q(p) - (1 - \phi) r(p) \) -- determines expected revenue.\(^{14}\)

Substituting for \( x \) in equation (2) implies:

\(^{12}\) If, by chance, the derivative is 0 throughout the range, production at an end point of the region maximizes expected profits, although other outputs would as well.

\(^{13}\) The insight that the firm chooses output to meet demand in one particular state applies more generally to any number of discrete states.

\(^{14}\) One point obscured by the assumption of just two demand states is the distinction between average demand and the average quantity sold. As Section III will explain, the distinction is important with continuous demand. It would also arise with three discrete states when the middle state is the HPSP. In that case, the difference between high demand and medium demand as well as differences in the sensitivity of demand to price in the two states are not important.
(4) \[ E[\pi] = \phi pq(p) + (1-\phi)p[q(p) - r(p)] - cq(p) = p\bar{q}(p) - cq(p) \]

The first order condition for maximizing (4) is:

(5) \[ \frac{dE[\pi]}{dp} = \bar{q}(p) + p\bar{q}'(p) - cq'(p) = 0 \]

The second order condition (conditional on the high demand state remaining the HPSP) is:

(6) \[ \frac{d^2E[\pi]}{dp^2} = 2\bar{q}'(p) + p\bar{q}''(p) - cq''(p) < 0 \]

Equations (5) and (6) are virtually identical to the conditions under certainty, the only difference arising from the distinction between the average quantity sold and the quantity produced. In equation (5), the first two terms are price marginal expected revenue, which depends on \( \bar{q}(p) \); and the last term is price marginal cost, which depends on \( q(p) \). The first two terms in equation (6) are the second derivative of the expected revenue function with respect to price (which, again, depends on \( \bar{q}(p) \)) while the last term is the second derivative of costs (which, again, depends on \( q(p) \)).

**B. Mark-up rule**

To derive the mark-up rule, divide the entire equation by the middle term and rearrange terms to get:

(7) \[ p = c \frac{q'(p)}{\bar{q}'(p)} \]
The fraction in brackets on the left-hand side is the inverse of the elasticity of the average quantity sold, which we can denote as $\eta_A$. Thus, letting $s = \left( \frac{q'(p)}{q'\prime(p)} \right)$, we can rewrite (7) as:

\[
(8) \quad p \left(1 + \frac{1}{\eta_A}\right) = \frac{c}{s}
\]

The left-hand side of (7) is quantity marginal revenue, i.e. the marginal revenue from an additional expected unit sold. When a company sells an extra unit of output, it gets the price it charges for the marginal unit, but the fraction of the price that it realizes as marginal revenue is determined by the elasticity of demand.\(^{15}\)

The right hand side is the marginal cost of an expected unit sold. When producing for the high demand state, the price reduction associated with a one unit increase in output increases demand in the high demand state by 1.\(^{16}\) It also increases demand in the low demand state but, except in the special case where $r'(p) = 0$ (i.e., additive uncertainty), one extra unit of output leads to something other than one additional unit sold in the low demand state. With multiplicative uncertainty, $r'(p)$ is negative, so the denominator of the right hand side is a fraction less than 1 (because $q'(p) < 0$).

A simple numerical example illustrates the point. Suppose low demand is always half high demand, that the two states occur with equal probability, and that the marginal

\(^{15}\) The mark-up rule under certainty rests on two underlying results. First, the profit-maximizing quantity is where marginal revenue equals marginal cost. Second, the ratio of marginal revenue to price (i.e., the fraction of the price the firm realizes as marginal revenue) is a function of the elasticity of demand. With uncertainty, this latter relationship holds in each state. The left-hand side of (6) reveals how the relationship can be averaged across states.

\(^{16}\) More generally with discrete states, the price reduction associated with a one-unit increase in output increases demand in the HPSP by 1. As a result, the quantity sold in the HPSP and all higher demand states increases by 1.
cost of a unit of output is $3. If the firm expands output by 1, it lowers its price to increase demand in the high demand state by 1. Because demand in the low demand state is half the demand in the high demand state, the price reduction increases demand in the low demand state by only 0.5. Thus, even though output increases by 1, expected sales increase by only 0.75. Yet, the firm incurs the full production cost of $3. Per expected unit sold, which is the appropriate basis of comparison with marginal revenue, marginal cost is $3/0.75 = $4.

Rearranging (8) gives the extension of (1) when the firm produces for the high demand state given two-state uncertainty:

\[ p = \frac{c}{s} \eta_A \frac{1}{1 + \eta_A} \]

Equation (9) is the discrete version of the main result of this paper. As will be shown in section III below, a closely analogous equation applies with continuous demand.

**C. Resolving the Paradox**

Equation (9) implies that the resolution of the paradox about the qualitative difference in the effects of increased additive and multiplicative uncertainty on price must lie in differences in how they affect \( s \) and \( \eta_A \).

Under additive uncertainty, \( r'(p) = 0 \) so \( q'(p) = \bar{q}'(p) \). In words, the effect of a price change on demand is constant across states, so the price reduction needed to increase demand by 1 in the high demand state increases average demand by 1 as well. As a result, \( s = 1 \), meaning that the marginal cost of an additional unit sold equals the marginal cost of a unit produced. An increase in uncertainty, interpreted as a reduction in
\( \phi \),\(^{17}\) does not affect the marginal cost of an expected unit sold.\(^{18}\) It does, however, lower the mark-up by increasing the elasticity of demand. With additive uncertainty, the slope of the demand relationship for a given price is constant across states, so increased uncertainty does not affect the slope of average demand with respect to price. But it lowers average demand at any price. The combination of a constant derivative of demand with respect to price and lower average demand implies a greater elasticity of average demand at any price.\(^{19}\)

With additive uncertainty, increased uncertainty lowers the optimal price because it lowers the mark-up and has no effect on marginal cost.

Multiplicative uncertainty implies \( r(p) = k q(p) \), so \( \bar{q}(p) = q(p) \left[ 1 - (1 - \phi) k \right] \). As a result,

\[
(10) \quad s = \frac{\bar{q}'(p)}{q'(p)} = 1 - (1 - \phi) k = \frac{\bar{q}(p)}{q(p)}, 
\]

\[
(11) \quad \eta_s \equiv \frac{d\bar{q}(p)}{dp} \frac{p}{\bar{q}(p)} = \frac{dq(p)}{dp} \frac{p}{q(p)} 
\]

Equation (10) says that \( s \), the ratio of the derivatives of the average quantity sold to the average quantity produced, equals the fraction of output the firm expects to sell. Because that fraction is generally less than 1, the marginal cost of an expected unit sold exceeds

---

\(^{17}\) One could also interpret increased risk as an increase in \( r(p) \). Under either interpretation, the risk is that the firm is unable to sell output that it produces so that an increase in risk implies an increase in the expected cost of unsold output.

\(^{18}\) An increase in \( \phi \) does increase the total cost of unsold output. However, with additive uncertainty, the amount of unsold output does not increase when the firm decides to lower price and increase its output. Thus, despite the effect on average cost, an increase in uncertainty has no effect on marginal cost.

\(^{19}\) There is another way of seeing the same point. When a firm lowers its price by $1, it sacrifices $1/unit on the units sold but it gets additional sales. With additive uncertainty, a greater probability of the low demand state does not alter the increase in sales due to a price reduction, but it does lower what the firm sacrifices by charging the lower price.
the marginal cost of a unit produced. Moreover, a reduction in $\phi$, which implies a greater fraction of unsold output on average, implies a reduction in $s$ and an increase in the marginal cost of a unit sold. Equation (11) says that the elasticity of the average quantity sold with respect to output (and, therefore, the mark-up) is independent of $\phi$.

With multiplicative uncertainty, an increase in the risk of unsold output increases the optimal price because it increases marginal cost and has no effect on the percentage mark-up of price over marginal cost.

In both the additive and multiplicative cases, the qualitative effect of increased uncertainty on price only applies when the high demand state is the HPSP. Figure 1 depicts how optimal prices vary with demand uncertainty for additive uncertainty, linear demand, and constant unit costs. It shows the optimal price as a function of $\phi$ for three different levels of the difference between high and low demand ($a$). Starting from the point where demand is high with probability 1, decreases in $\phi$ - that is, increases in the probability of low demand – cause a linear reduction in price. In this region, a firm produces for the high demand state, so a price reduction implies an increase in output. The rate of decrease is greater for larger values of $a$. Thus, within this region, the greater the amount of output that goes unsold in the low-demand state, the greater the increase in output as the probability of low demand increases. For all three values of $a$, however, there is a critical value of $\phi$ below which the firm produces for the low demand state and charges the price that would maximize profits if that state occurred with certainty. Once a firm reaches that break, it ignores the opportunity to sell extra in the high demand state

---

20 Because the quantity demanded in the low demand state is proportional to demand in the high demand state with multiplicative uncertainty, the quantity of unsold output must also necessarily be proportional to demand in the high demand state. As a result, a decision to lower price and expand supply to meet demand in the high demand state results in more unsold output on average.
so variations in the probability of high demand do not cause it to change its price or output.

Figure 2 depicts the optimal price as a function of $\phi$ for three different values of $k$. As in Figure 1, each value of the demand shift parameter has two ranges, one where the firm produces for high state and one where it produces only for low demand. In contrast to Figure 1, the prices increase as $\phi$ decreases, implying that output also decreases. In the region where the firm produces for high demand, the relationship between the price and $\phi$ is steeper for lower values of $k$, which means that the bigger the demand reduction in the low demand state, the greater is the sensitivity of price to the probability that the low demand state occurs.

**D. Comparative Statics**

Using equation (6) and noting that the partial derivative of (5) with respect to $c$ is $-q'(p)$, the pass through rate for this discrete version of the price-setting newsvendor problem is:

$$
\frac{dp}{dc} = -\frac{\frac{\partial^2 E[\pi]}{\partial p \partial c}}{\frac{\partial^2 E[\pi]}{\partial p^2}} = \frac{q'(p)}{2\bar{q}'(p) + p \bar{q}''(p) - cq''(p)}
$$

Under certain demand, average demand equals high demand, so (9) reduces to:

$$
\frac{dp}{dc} = \frac{1}{2 + \frac{(p-c)q''(p)}{q'(p)}} = \frac{1}{2 - \frac{q(p)q''(p)}{[q'(p)]^2}},
$$

14
where the second equality uses the first order condition to substitute for \((p-c)\). With uncertainty \((0 < \phi < 1)\), \(q'(p)\) equals \(q'(p)\) only if \(r'(p)\) is 0, which implies additive uncertainty.\(^{21}\)

As noted in the introduction, pass-through for a linear demand curve is 0.5 under certainty. Equation (12) implies that linearity in both states \((r'(p)\text{ being any constant})\) is not sufficient for this to be case under uncertainty. Rather, the 0.5 pass-through for linear demand arises only with additive uncertainty.

III. Continuous Distribution of Demand

To extend these results to the case of continuous demand, let \(Q(\epsilon,p)\) be the inverse cumulative distribution function for the quantity demanded conditional on \(p\), where \(p\) is price and \(\epsilon\) is a random variable uniformly distributed between 0 and 1 with \(\partial Q/\partial \epsilon > 0\) and \(\partial Q/\partial p < 0\). Again letting \(x\) be output, the firm sells \(x\) if \(Q(\epsilon,p) \geq x\) and \(Q(\epsilon,p)\) otherwise.

A. Solution

A standard way to find the optimum is to make price and output the choice variables. With this approach, the first order condition with respect to price captures the effect on expected profits of changing price holding output (and therefore cost) constant. The first order condition with respect to output captures the effect on expected profits of changing output holding price (and therefore demand) constant. Neither corresponds to the single first order condition under certainty which analyzes the effect on profits of

\(^{21}\) \(q''(p)\) equals \(q''(p)\) only if \(r''(p)\) is 0.
simultaneously changing price and output to reflect the response of demand to price. Since the mark-up rule under certainty is simply a form of that first order condition, it is not surprising that neither partial first order condition by itself implies a mark-up rule similar to the one that applies under uncertainty.

Assuming that, given price, the probability that the firm can satisfy all demand is an invertible function of output, one can substitute this probability (i.e., the HPSP) for output as a choice variable in the model of the firm’s optimization. With this transformation, the first order condition with respect to price captures the effect on expected profits of simultaneously changing price and output to hold constant the probability of being able to meet all demand. By holding that probability (and, therefore, the demand state) constant, the change in output associated with a change in price reflects a pure response of the quantity demanded to price. This transformation is a generalization of the “stocking factor” transformation introduced by Petruzzi and Dada (1999) and subsequently used by Petruzzi, Wee, and Dada (2009).

Whether demand uncertainty is discrete or continuous, a key aspect of the economics of the price-setting newsvendor problem is that, given price, the average quantity purchased determines revenue whereas demand in the HPSP determines cost. Let \( \epsilon^* \) be the probability that the firm can satisfy all demand, i.e., the HPSP. The average quantity purchased, which depends on both price and the quantity produced (or, equivalently, \( \epsilon^* \)), is

\[
(13) \quad \overline{Q}(\epsilon^*, p) = Q(\epsilon^*, p)(1-\epsilon^*) + \int_0^{\epsilon^*} Q(\epsilon; p) d\epsilon
\]

16
The distinction between $\overline{Q}(\varepsilon^*, p)$ and $Q(\varepsilon, p)$ plays the same role in the continuous case as the distinction between $\overline{q}(p)$ and $q(p)$ plays in the two-demand state case.

The expected profit function is:

\[ E[\Pi] = pQ(\varepsilon^*, p)(1 - \varepsilon^*) + p \int_0^{\varepsilon^*} Q(\varepsilon; p) d\varepsilon - cQ(\varepsilon^*, p) = p\overline{Q}(\varepsilon^*, p) - cQ(\varepsilon^*, p) \]

The first order conditions are:

\[ \frac{\partial E[\Pi]}{\partial p} = \left[ Q(\varepsilon^*, p) + p \frac{\partial Q(\varepsilon^*, p)}{\partial p} \right] (1 - \varepsilon^*) + \int_0^{\varepsilon^*} Q(\varepsilon; p) d\varepsilon + p \int_0^{\varepsilon^*} \frac{\partial Q(\varepsilon; p)}{\partial p} d\varepsilon \]

\[ -c \frac{\partial Q(\varepsilon^*, p)}{\partial p} = \overline{Q}(\varepsilon^*, p) + p \frac{\partial \overline{Q}(\varepsilon^*, p)}{\partial p} - c \frac{\partial Q(\varepsilon^*, p)}{\partial p} = 0 \]

The second order conditions are:

\[ \frac{\partial^2 E[\Pi]}{\partial \varepsilon^* \partial p} = [p(1 - \varepsilon^*) - c] \frac{\partial Q}{\partial \varepsilon^*} = 0, \]

\[ \frac{\partial^2 E[\Pi]}{\partial p^2} = \left[ 2 \frac{\partial Q(\varepsilon^*, p)}{\partial p} + p \frac{\partial^2 Q(\varepsilon^*, p)}{\partial p^2} \right] (1 - \varepsilon^*) + 2 \int_0^{\varepsilon^*} \frac{\partial Q(\varepsilon; p)}{\partial p} d\varepsilon \]

\[ + p \int_0^{\varepsilon^*} \frac{\partial^2 Q(\varepsilon; p)}{\partial p^2} d\varepsilon - c \frac{\partial Q(\varepsilon^*, p)}{\partial p^2} \]

\[ = 2 \frac{\partial \overline{Q}(\varepsilon^*, p)}{\partial p} + p \frac{\partial^2 \overline{Q}(\varepsilon^*, p)}{\partial p^2} - c \frac{\partial^2 Q(\varepsilon^*, p)}{\partial p^2} < 0 \]
\[
\frac{\partial^2 E[\Pi]}{\partial p^2} \frac{\partial^2 E[\Pi]}{\partial \varepsilon^*} - \left[ \frac{\partial^2 E[\Pi]}{\partial p \partial \varepsilon^*} \right]^2 = \\
- \left\{ 2 \frac{\partial^2 Q(\varepsilon^*, p)}{\partial p^2} + p \frac{\partial^2 Q(\varepsilon^*, p)}{\partial p^2} - c \frac{\partial^2 Q(\varepsilon^*, p)}{\partial p^2} \right\} \frac{\partial Q(\varepsilon^*, p)}{\partial \varepsilon} \frac{p}{\partial \varepsilon} \\
- \left[ (1 - \varepsilon^*) \frac{\partial Q(\varepsilon^*, p)}{\partial \varepsilon} \right]^2 > 0
\]

**B. Interpretation and the Mark-up Rule**

The simpler of the two first order conditions is equation (16), which characterizes \( \varepsilon^* \), the HPSP. Since \( \varepsilon^* \) is the probability of being able to satisfy all demand, \( 1 - \varepsilon^* \) is the probability of stocking out. In states where the firm can satisfy all demand, an extra unit of production does not lead to additional sales or revenue. Thus, holding price constant, the expected marginal revenue from an additional unit produced is \( p(1 - \varepsilon^*) \). Equation (16) says that the firm should produce up to the point where this expected marginal revenue equals the marginal cost of a unit produced.

Equation (15) is the condition that corresponds to the first order condition under certainty and is the source of the mark-up rule. The first two terms are price marginal expected revenue when output is adjusted to satisfy demand in the HPSP. The third term is price marginal cost with the same output adjustment. Under certainty, the derivatives in the second and third term equal each other, which makes it straightforward to solve for \( (p - c) \) (or \( p \)). With uncertainty (that is not additive), they are not generally equal.
Following steps similar to the derivation of equation (9) implies a generalization of the mark-up rule to the price-setting newsvendor problem. Along with equation (9), equation (19) is the key result in this paper:

\[
p = \frac{c}{S} \frac{E_A}{1 + E_A}
\]

where, \( E_A = \frac{\partial Q(e^*, p)}{\partial p} \frac{p}{Q(e^*, p)} \) and \( S = \frac{\partial Q(e^*, p)}{\partial p} \).

The intuition underlying (19) closely follows the intuition behind the discrete case. The mark-up rule is simply a reformulation of the condition that (quantity) marginal revenue equals (quantity) marginal cost. Quantity marginal revenue is the additional revenue from an additional unit sold. For a firm without any market power, marginal revenue is the price. For a firm with market power, the ratio of marginal revenue to price (or, in other words, the fraction of the price that the firm “keeps” as marginal revenue) is a function just of the elasticity of demand. Under uncertainty, the relationship between the ratio of marginal revenue to price and the elasticity of demand still holds with marginal expected revenue being the additional expected revenue from an additional expected unit sold and the elasticity of demand being the elasticity of the average quantity sold with respect to price.

As with discrete demand states, \( c \) is the marginal cost of an extra unit produced which, under demand uncertainty, is generally different (except in the special case of additive uncertainty) from the marginal cost of an additional expected unit sold. Thus, to equate marginal expected revenue from an additional expected unit sold to the marginal
cost of an expected additional unit sold, one must scale the marginal cost of a unit produced by a factor that reflects the fraction of the marginal unit produced that the firm sells.

Note that the distinction between the average quantity sold and the quantity produced plays a key role in the second order conditions. Equation (17) says that the second derivative of the expected revenue function with respect to price minus the second derivative of the expected cost function must be negative. Because price expected marginal revenue depends on the average quantity sold whereas price marginal cost depends on the quantity produced, the second derivative of expected revenues with respect to price depends on \( \bar{Q}(\varepsilon^*, p) \) and its derivatives whereas the second derivative of costs with respect to price depends on \( Q(\varepsilon^*, p) \) and its derivatives.

### C. Resolving the Paradox

Armed with the general result for the mark-up, we can resolve the paradox of why additive and multiplicative uncertainty have qualitatively different effects on pricing. The analysis closely parallels the analysis with two demand states. Equation (20) implies that the difference in qualitative effects of additive and multiplicative uncertainty must come down to differences in how they affect the marginal cost of an expected unit sold and the elasticity of the average quantity sold with respect to price.

With additive uncertainty, \( \frac{\partial Q(\varepsilon, p)}{\partial p} \) is constant across states, so

\[
\frac{\partial Q(\varepsilon^*, p)}{\partial p} = \frac{\partial \bar{Q}(\varepsilon^*, p)}{dp} \quad \text{and} \quad S = 1.
\]

As when there are just two states of demand, the marginal cost of an expected unit sold equals the marginal cost of an extra unit produced.
Also, similar to the two-demand state case, a constant slope of the demand curve across states implies that an increase in the probability of low-demand states that lowers the average quantity sold increases the absolute value of the demand elasticity and lowers the optimal mark-up.

An interesting feature of the additive case is that the cost of unsold output is a fixed cost. When the firm decides to lower price and increase output with the expectation of selling more on average, the quantity (and therefore cost) of the unsold output does not change. Because the cost of unsold output is a fixed cost, additive uncertainty creates increasing returns to scale even when the underlying production process is constant returns. That is, average cost declines with the level of output and marginal cost is less than average cost.

With multiplicative uncertainty,

\[ Q(\varepsilon, p) = q(p)M(\varepsilon) \]

where \( q(p) \) is the maximum possible demand conditional on price, \( \varepsilon \) is distributed uniformly from 0 to 1, and \( M(\varepsilon) \) is an inverse cumulative distribution function for the multiplicative random factor with \( M(1) = 1 \) and \( M'(\varepsilon) > 0 \). Equation (20) implies:

\[ Q(\varepsilon^*, p) = q(p)M(\varepsilon^*) \]

\[ \overline{Q}(\varepsilon^*, p) = q(p)M(\varepsilon^*)(1-\varepsilon^*) + \int_0^{\varepsilon^*} q(p)M(\varepsilon) \, d\varepsilon \]

which in turn imply:

---

22 In their reconciliation of the additive and multiplicative cases, Petruzzi and Dada (1999) also observe that the cost of unsold items is a fixed cost with additive uncertainty.
As with two states of demand, the ratio of the derivatives of the average quantity sold and the average quantity produced with respect to price equals the average fraction of the quantity produced that the firm sells. Since that average is typically less than 1, the marginal cost of an expected unit sold is greater than the marginal cost of an extra unit produced; and an increase in the fraction of unsold output implies an increase in the marginal cost of an expected unit sold.

Substituting (21) into the expression for $E_A$ implies:

$$E_A = \frac{p \left[ dq \frac{\partial}{\partial p} M(\varepsilon^*) (1-\varepsilon^*) + \int_0^{\varepsilon^*} dq \frac{\partial}{\partial p} M(\varepsilon) \, d\varepsilon \right]}{q(p) M(\varepsilon^*) (1-\varepsilon^*) + \int_0^{\varepsilon^*} q(p) M(\varepsilon) \, d\varepsilon} = \frac{dq}{dp} \frac{p}{q}$$

With multiplicative uncertainty, the elasticity of demand at a given price is the same in all states, so the elasticity of the quantity sold is not affected by changes in the probabilities of the different demand states. Because an increase in the probability of low demand states (and therefore in the expected amount of unsold output) does not affect the elasticity of the average quantity sold, it does not affect the markup of price over the marginal cost of an expected unit sold.
In contrast to the linear uncertainty case, multiplicative uncertainty implies that the expected quantity of unsold output is proportional to the quantity of output. Thus, if the underlying production process entails constant returns with respect to the quantity produced, multiplicative uncertainty implies constant returns with respect to the quantity sold.

The results in this section demonstrate that the resolution of the paradox for continuous states of demand is essentially the same as in the case of two demand states. Increased additive uncertainty lowers the price that maximizes expected profits because it does not affect marginal cost but it does increase the elasticity of the average quantity sold with respect to price. Increased multiplicative uncertainty increases the marginal cost of an expected unit sold but does not affect the percentage mark-up.

**D. Comparative Statics**

Let the notation $E\pi_{..}$ denote second partial derivatives of the expected profit function. The increase in price due to an increase in marginal cost is, then,

$$\frac{dp}{dc} = \frac{E\pi_{pe} - E\pi_{pe^2}}{E\pi_{ee^2}}$$

As described above with respect to equation (12’), the derivative of price with respect to cost (or the pass-through rate) simplifies under certainty into a function of just the demand curve (and its first two derivatives). Given that a similar simplification is not possible even with just two states of demand, it should come as no surprise that there is not a similar simplification with continuous demand states.
There are two reasons why the simplification under certainty does not generalize to uncertainty. The first is similar to the reasoning in the two demand state case and relates to the distinction between the average quantity sold and the quantity produced. Consider what might be termed a “partial pass through effect” defined as the change in price in response to a change in the marginal cost of a unit produced holding $\varepsilon^*$ constant. Paralleling the certainty case,

\begin{equation}
\frac{\partial p}{\partial c_{\varepsilon^*}} = -\frac{E\pi_{pc}}{E\pi_{pp}} = \frac{\partial Q(\varepsilon^*, p)}{\partial p} = -\frac{\partial Q(\varepsilon^*, p)}{\partial p} + p \frac{\partial^2 Q(\varepsilon^*, p)}{\partial p^2} - c \frac{\partial^2 Q(\varepsilon^*, p)}{\partial p^2}
\end{equation}

The simplification under certainty requires using the first order condition to substitute for $(p - c)$. However, because (conditional on price) revenue is a function of average demand whereas output and therefore cost are functions of demand in the HPSP, price and marginal production cost are multiplied by different derivatives in both the first and second order conditions. Since these terms multiplying $p$ and $c$ are first derivatives of the demand relationships in the first order conditions and second derivatives in second order conditions, there is no general way to use the first order condition to substitute out for $p$ and $c$ in the second order conditions and, therefore, the comparative statics.

The second reason that the simplification cannot occur is the additional terms in (24) that are not present in (25). These terms reflect the fact that when $c$ changes, the HPSP also changes with feedback effects for $p$.

IV. Concluding observations

Aided by the generalization of the stocking factor transformations suggested by Petruzzi and Dada for the additive and multiplicative cases, the derivation of a general
mark-up rule for the price setting newsvendor problem is surprisingly simple. The solution of the theoretical problem suggests, though, that a series of practical problems might be more difficult than had been appreciated.

The results indicate that in demand estimation used to understand pricing, the specification of the error term is not merely a question of getting the most precise estimates possible or accurate estimates of parameter standard errors. The properties of the residual affect the first order condition for pricing.23

The results also reveal a subtlety in the measurement and even conceptualization of marginal cost. In the standard microeconomics textbook treatment of production and cost, the production function gives the quantity produced for given levels of inputs and the cost function gives the cost of the quantity produced. The cost of output that is produced but not sold, which is an inevitable feature of many businesses, does not enter the analysis. To understand pricing and output, however, it has to.

An important application of these principles is market definition in merger review. Throughout the world, market definition rests on the hypothetical monopolist test. The key question underlying this so-called SSNIP24 test is whether a monopolist over a hypothesized market would have an incentive to raise prices over current or likely future levels. In many cases, the only available way to do so is to model the monopoly price.25 The analysis here suggests that doing that calculation properly requires far more subtlety than is generally recognized.

23 This is the case even when the factors unobserved by the econometrician are the same as the factors unobserved by managers making pricing and output decisions.
24 SSNIP stands for “Small, but Significant, Non-transitory Increase in Price.”
25 Within antitrust, the analysis underlying market definition is referred to as “critical loss” analysis. See Harris and Simons (1989), Katz and Shapiro (2003), and O’Brien and Wickelgren (2003).
Another practical application of the newsvendor problem is vertical contracts.\textsuperscript{26} Krishnan and Winter (2007) recent analysis of resale price maintenance relies on results from the price-setting newsvendor literature. With the recent decision by the United States Supreme Court to overturn the per se illegality of minimum resale price maintenance,\textsuperscript{27} increased scholarly attention to the problem is likely. To the extent that a manufacturer’s downstream sellers compete with each other, models of optimal vertical contracts might need to rest on the solution of the newsvendor problem under oligopoly. Just as the monopoly mark-up rule provides useful intuition for various aspects of the solution to oligopoly models,\textsuperscript{28} a monopoly mark-up rule might yield insights into solutions of oligopoly models with demand uncertainty.

There are many extensions of the newsvendor problem with endogenous pricing. One can allow for a depreciated value of unsold output, a reputation cost for unmet demand, and demand sensitivity to the probability of available supply.\textsuperscript{29} One can allow for decision variables other than price to affect demand.\textsuperscript{30} Based on the insights of this paper, it is likely that there are more intuitive solutions for these extensions.

\begin{footnotesize}
\begin{itemize}
\item\textsuperscript{26} See Krishnan and Winter (2007) and Petruzzi and Dada (2009).
\item\textsuperscript{27} See \textit{Leegin Creative Leather Products, Inc. v. PSKS}, 551 U.S. 877 (2007).
\item\textsuperscript{28} For example, in the Cournot equilibrium with heterogenous costs, each firm’s share is inversely related to marginal cost. The result is a natural implication of the mark-up rule with respect to the elasticity of residual demand faced by each firm. With a common price, a firm with a higher marginal cost than another necessarily has a lower mark-up and therefore a higher demand elasticity. In the Cournot model, firms with lower shares have higher elasticities of residual demand.
\item\textsuperscript{29} See Dana and Petruzzi (2001), Petruzzi, Wee, and Dada (2009), and Kocabıyıköglü and Popescu (2005).
\item\textsuperscript{30} For example the Dorfman and Steiner (1954) analysis of optimal advertising and optimal quality could be extended to incorporate uncertainty. See also Petruzzi, Wee, and Dada (2009).
\end{itemize}
\end{footnotesize}
References


Figure 1

Additive Uncertainty

Assumptions

High Demand Curve: \( Q = 2 - P \)

Costs: \( c = 1 \)

Optimal Price vs. Probability of High Demand for different values of \( a \):
- \( a = 0.15 \)
- \( a = 0.30 \)
- \( a = 0.45 \)
Figure 2

Multiplicative Uncertainty

Assumptions

High Demand Curve: $Q = 2 - P$
Costs: $c = 1$

Optimal Price

Probability of High Demand