Monopoly Price Dispersion Under Demand Uncertainty

James D. Dana Jr.

June, 1992

Revised: October, 1996

Primary JEL Number: D42 Monopoly
Secondary JEL Numbers: D45, L12

Keywords: Monopoly, Demand Uncertainty, Price Dispersion, Price Discrimination

Abstract

When a monopolist must choose its price before the level of demand is known, then setting dispersed prices may dominate setting a single price. This paper adds demand uncertainty to the standard monopoly model and expands the monopolist’s choice set to include distributions of prices. The result is that when the elasticity of demand is decreasing in the demand state, so that the ex post monopoly spot price is increasing in the realization of the demand state, then the monopolist’s optimal ex ante pricing strategy is to set dispersed prices. Intuitively, when the ex post monopoly price and the level of sales are positively correlated, then the monopolist is free to offer some of its units at higher prices knowing that these will sell only in the higher demand states and will not impact profits in the low demand states. This simple model explains why a monopolist would offer a product line that includes lower-priced, lower-margin units whose availability is limited while offering higher-priced, higher-margin units with unlimited availability.

† AFFILIATION: Northwestern University
CORRESPONDENCE: Department of Management and Strategy, Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, TEL: 847-467-3362, FAX: 847-467-1777, E-MAIL: j-dana@nwu.edu. This paper, its supplemental appendix, and other working papers by the author are available electronically at http://www.kellogg.nwu.edu/faculty/dana/. This paper previously circulated as Dartmouth College Working Paper 91-12 with the title “The 'Law of One Price' Does Not Hold for Monopolists Facing Uncertain Demand.” Printed on 10/21/96.

‡ I would like to thank Thomas Holmes, Carsten Kowalczyk, Olof Staffans, Kathryn E. Spier, and especially Gustaf Gripenberg for their helpful comments and the Rockefeller Grant in Economics at Dartmouth College for financial support.
When a monopolist producing a homogeneous good must choose its price before demand is known, then under a broad range of demand conditions, it will not choose to set a single price. I suppose that a monopolist has constant marginal cost, but must choose its price, or prices, before the level of demand is realized, and show that if the elasticity of demand decreases with the level of demand, so that the monopolist's *ex post* optimal price increases with demand, then the monopolist’s *ex ante* optimal pricing strategy exhibits price dispersion. On the other hand, if the elasticity of demand weakly increases with the level of demand, so that the monopolist’s *ex post* optimal price is constant or decreases with demand, then the monopolist’s *ex ante* optimal pricing strategy is to set a single price.

Intuitively, when the *ex post* monopoly price and the level of demand are positively correlated, then the monopolist can use price dispersion to achieve some of the same benefits that it gains from the ability to adjust prices in response to demand. Though I assume that the firm cannot directly discriminate between customers in the high demand states and the low demand states (there is an implicit assumption that the transactions costs of adjusting prices are high), the firm nevertheless can partially discriminate between the high and low demand states since it can set a higher price for those units that sell only in the higher demand states. This will not affect profits in the low demand states since when demand is low the firm has adequate supply at lower prices.

A simple numerical example with unit demands and two demand states illustrates the result and the intuition. Imagine a stadium owner who must print tickets in advance for a concert or sporting event with uncertain demand. Suppose that demand will either be high or low, each with probability 1/2. Low demand consists of 50 consumers with a reservation value of $10, and high demand consists of 100 consumers with a reservation value of $12. If the stadium has unconstrained capacity (i.e., more than 100 seats) and marginal cost is zero, then the optimal pricing strategy is to print 50 tickets at a price of $10 and 50 tickets at a price of $12. Expected revenue is $800, or $50 greater than the expected revenue earned from the best uniform price strategy (set price equal to $10 for all seats). In the high demand state, the first 50 consumers (it
doesn’t matter which ones since consumers are identical) will get the lower priced seats, but the monopolist earns extra profit because the remaining 50 customers pay a higher price.

The model presented here generalizes the above numerical example by considering continuously differentiable demand functions and a continuum of demand states, however the intuition is the same. If the marginal revenue is greater in high demand states than in low demand states, then the firm will offer dispersed prices, setting a lower price on those units which sell in low demand states as well as high demand states and a higher price on units that sell only in the high demand states. Although the sufficient condition for price dispersion, i.e., that the elasticity of demand must be decreasing in the level of demand, is slightly stronger the condition that the ex post monopoly price be increasing in the level of demand, the later condition offers powerful intuition and in almost any empirical demand specification is synonymous with decreasing elasticity.

The paper relies on demand uncertainty and, more importantly, on the assumption that the firm’s prices cannot be made contingent on the realization of the demand state. Hence the paper is not a realistic model of the electric power industry where firms can employ contingent contracts such as priority service pricing (see Chao and Wilson, 1987, and Wilson, 1989) or other industries with large industrial buyers who can use long term contracts to specify allocation mechanisms to handle demand uncertainty. Instead the paper should be thought of as a model of consumer pricing where the number and identity of consumers is difficult to predict ahead of time, and hence it is impossible to write contracts with all customers. I also assume that the transactions costs of adjusting prices in response to information about demand is also high. There are many justifications for such an assumption. First, price setting is often associated with menu costs, especially when items such as tickets are preprinted with prices on them. Second, firms who engage in advertising to promote their products are extremely reluctant to change advertised prices for fear of damaging their reputations, while they appear to be more willing to advertise limited availability at lower prices and to offer consumers similar, but higher priced products once the lower priced units have stocked out. Finally, firms understand that their consumers often make
reliance expenditures, such as travel arrangements or search costs, which can give the firm an incentive to commit to a fixed prices *ex ante*, and thus not appropriating consumers relationship-specific investments, even when their own transactions costs of changing prices are small (see Gilbert and Klemperer, 1993).

Several earlier papers have offered alternative theories of monopoly price dispersion. One such paper is Wilson (1988) which demonstrates that monopoly price dispersion can result from increasing marginal cost and a non-concave revenue function, such as would arise if demand were a step function. Wilson's model is best understood in a simple example. Suppose a firm has 100 units to sell, customers purchase in random order, and demand consists of 50 consumers who are willing to pay $12 and 100 customers who are willing to pay $10, then the profit maximizing pricing strategy is to offer 50 units at $10 and fifty units at $12. High valuation customers get half of the $10 seats and all of the $12 seats even though the monopolist cannot explicitly discriminate between them. Wilson's model gives the same prediction as this paper – the firm limits the number of low priced units and increases its profit by charging higher prices for its remaining units – however the result is driven by non-concavities and increasing marginal costs (or capacity constraints) and not price rigidities or demand uncertainty. Unlike Wilson's paper, I find that price dispersion may result even when costs are constant and the revenue function is strictly concave.

Stiglitz (1982), in a paper on optimal taxation, suggests that a monopolist might want to commit to a random pricing strategy. Salop (1977) suggests that a monopolist with multiple retail outlets may charge different prices across its outlets, since it allows the monopolist to extract higher average prices from consumers with higher search costs. Of course price discrimination itself is an important form of monopoly price dispersion; see Varian (1989) for a useful survey of this literature.

Prescott (1975), Eden (1990), and Dana (1992) show that when capacity is costly (or constrained) then demand uncertainty can explain price dispersion. In their models, a firm facing uncertain demand will set dispersed prices if it must set its prices in advance and it cannot sign
forward contracts with its customers. Some of the firm's units carry a lower price and are sold with very high probability, while other units carry a higher price and are less likely to be sold. Price dispersion arises because the firm must earn a higher return on units that are less likely to be sold in order to generate enough expected revenue to cover its capacity costs.

Finally, Gale and Holmes (1992, 1993) consider a model in which a monopolist faces unknown demand (it is unknown which period will be peak and which will be off peak) and uses advanced purchase discounts to allocate its scarce capacity efficiently and increase its expected profits.

**THE MODEL**

I consider a monopolist with constant marginal costs of production, $c$, facing uncertain demand. Demand uncertainty is parameterized by a random variable, $e$, drawn from a continuously differentiable cumulative probability distribution, $F(e)$, with strictly positive, and bounded probability density, $f(e)$, on a support $[e, \bar{e}]$.

Let $D(p, e)$ denote the demand function in state $e$. I assume $D(p, e)$, $D: \mathbb{R}^2_+ \to \mathbb{R}_+$, is continuous, non-increasing in $p$, non-decreasing in $e$, and is thrice continuously differentiable with derivatives $D_p < 0$ and $D_e > 0$ on the restricted domain, $\{(p, e)|D(p, e) > 0\}$. Furthermore, there are no income effects, and $D_{pp}D - 2(D_p)^2 < 0$, or equivalently, the industry marginal revenue function is strictly decreasing. Let $\eta(p, e)$ denote the elasticity of demand, i.e., $\eta(p, e) = -p D_p(p, e)/D(p, e)$. Under these assumptions, the *ex post* monopoly spot price, denoted $p^m(e)$, is well-defined, unique, and continuous in $e$.

Let $\varphi(e) = \lim_{u \to e} \left\{ \inf \{p|D(p, u) = 0\} \right\} \forall e \in [e, \bar{e}]$. So $D(\varphi(e), e) = 0$, and, loosely speaking, $\varphi(e)$ is the intercept of the inverse demand curve in state $e$. The limit is included in the

---

1 The use of forward contracts under demand and supply uncertainty is known as priority service pricing (see Robert Wilson, 1989, Chao and Robert Wilson, 1987, and Harris and Raviv, 1981). The forward contracts specify both a price and a reliability (or quality) of service, and at the optimum, the forward contract price is increasing in reliability. Hence, priority service pricing may also be thought of as a model of price dispersion, although in this case the products are clearly differentiated.

2 Letting $P(q, e)$ denote the inverse demand function, this condition is equivalent to $P_{qq} + 2P_q < 0$. 
definition of $\varphi(e)$ since the demand curve may be degenerate at $e$, in particular when $e = \varepsilon$, as in
the case of multiplicatively separable demand where $D(p, e) = d(p)e$.

I assume that one of three conditions holds: (i) $D(p, e) > 0, \forall e > \varepsilon$, and $\forall p \in (0, \infty)$;
(ii) $\varphi(e) = \bar{\varphi}$ is constant; or (iii) $\varphi(e)$ is continuous, monotonic, and increasing in $e$. System of
demand functions in the first case (e.g., constant elasticity demand) have no price intercept so $\varphi(e)$
is undefined. In Cases (ii) and (iii) I require that a finite price intercept exist for all demand states.
By construction $\varphi(e)$ is single-valued, however additional assumptions imposed above in Cases
(ii) and (iii) are required to guarantee that $\varphi(e)$ is sufficiently well-behaved. Case (ii) defines
systems of demand functions which have a common price intercept in each demand state (e.g.,
multiplicatively separable demand). Case (iii) is more general, allowing the price intercept to vary
continuously with $e$ (e.g., additively separable demand). The assumptions in Case (iii) are
sufficient to guarantee that $\varphi(e)$ is invertible. Note that in every case the demand function may still
be degenerate at $e = \varepsilon$, that is $D(p, e) = 0, \forall p \in [0, \varphi(e)]$, and $\varphi(e) > 0$.

There are no constraints on the firm’s level of production (i.e., capacity is unlimited),
however the firm must assign a price to each of its units of production before its demand is
realized. This assumption can be motivated by transactions costs, such as the importance of
maintaining reputation for offering prices as advertised or the difficulty of training sales personnel
to adjust prices, or by menu costs, such as the costs of printing tickets or price schedules.

Let $Q(p)$ denote the price distribution or pricing strategy of the firm. The price distribution
$Q(p)$ is defined as the total number of units at offered at prices less than or equal to $p$, and is a
cumulative distribution function. I will restrict the set of admissible pricing strategies to functions
$Q(p)$ of the form:

\[ Q(p) = Q + \sum_{p_k \leq p} Q_k + \int_p^\infty q(r)dr, \tag{1} \]

where $q(p)$, which denotes the derivative, $dQ(p)/dp$, when it is defined, is a strictly positive
continuous function on $[p, \bar{p}] \subset \mathbb{R}^+$; $(p, Q) \cup \{(p_k, Q_k); k = 2, \ldots, n\}$ is a finite set of nonnegative
mass points with prices satisfying $p_k \in [p, \bar{p}]$ for all $k$; and $p < \varphi(\bar{e})$, since otherwise even the
lowest priced units would never be sold. So, \( Q(p) \) is almost everywhere continuous and differentiable. The restriction that \( q(p) \) be continuous can be relaxed and replaced by the weaker restriction that \( q(p) \) be Lesbesgue integrable if the demand is strictly positive (i.e., the intercept is large and \( \epsilon \) is sufficiently large).

At the optimum, the monopolist's optimal pricing policy will contain at most one mass point and that mass point will be at the price \( p \), which is why I have used the separate notation, \( Q \), for the mass point at \( p \). When that mass point represents the firm’s entire price distribution, i.e., the firm sets only one price, then I will also refer to it using the notation \( (p^*,Q^*) \). The intuition for the absence of mass points at higher prices is that except for units sold in the lowest demand state, each unit that is offered by the monopolist sells in a different set of demand states and so is associated with a different optimal price. The distribution \( F(e) \) is continuous, so in equilibrium the price distribution \( Q(p) \) is also continuous.

**RATIONING**

As in other models of price-setting with either strategic or non-strategic uncertainty, this paper appeals to an exogenous rationing rule to decide how units are allocated to consumers. As in Wilson's (1988) model of monopoly price dispersion and in many other papers in oligopoly theory (see for example Beckman, 1965, and Davidson and Deneckere, 1986) I assume that goods are allocated proportionally, subject to consumer's willingness to pay. The proportional rationing rule has the advantage that it is consistent with a variety of queuing models. For example, suppose that consumers are small and arrive in random order, so that any positive measure of consumers has a demand which is proportional to the aggregate demand. The firm's units are sold in order of price, from cheapest to most expensive, and when it is their turn to buy, consumers buy as many units as they want at the lowest prices still available until their marginal willingness to pay equals or exceeds the available price. Consumers who are fortunate to arrive first pay lower prices. Of course such a queuing model assumes that consumers purchase only for their own needs and that resale is infeasible.
The assumption of proportional rationing is obviously strong, however in this model it is more compelling then the obvious alternative, the parallel or efficient rationing rule, which supposes that consumers queue in decreasing order of their valuations (this is easiest to imagine when consumers have unit demands) or that resale is permitted, so it is as if all units trade at a common price *ex post*. Parallel rationing has some advantages, but none that are compelling in this context. In oligopoly models, economists have argued that firms might prefer the parallel rationing rule to the proportional rationing rule since lower priced firms can reduce their competitors' profits by implementing parallel rationing. Since by construction the parallel rationing minimizes producer surplus, and maximizes consumer surplus, it certainly has this effect, however I have always found it difficult to imagine that such a practice is really in the long run interests of oligopolists. More importantly, in a monopoly model with price dispersion the opposite intuition is true – parallel rationing minimizes the sales of the monopolist and hence minimizes the monopolist's *ex post* profits (given its prices). The monopolist would prefer to sell its low priced units to low valuation consumers in order maximize the number of consumers remaining who will buy its higher priced goods and maximize its sales, thus maximizing its profits. Second, parallel rationing is plausible when there is a threat of a stock out, since consumers who value consumption the most (those with the highest consumer surplus) are more likely to arrive early in order to make sure that they are not stocked out. Again, this intuition does not apply in this context because here rationing order affects only the price paid and not the probability of being stocked out. If anything one might imagine that low valuation consumers would also have a lower cost of time and hence be more willing to spend time in a queue in order to obtain a lower price. Finally, if resale is allowed then the parallel rationing rule is clearly more plausible. However in any model in which the transactions costs of adjusting price in response to changes in market demand are high, the transactions costs of operating a resale market are also going to be high.

---

3 The terms *proportional* and *parallel* rationing take their names from the effect that each rule has on the shape of the residual demand curve (see Davidson and Deneckere, 1986).
The distinction between these two common rationing rules, and others that might be employed, is important because the specification of the firm’s pricing strategy and the sufficient conditions for price dispersion would be quite different under alternative rationing rules, and in some cases the sufficient conditions would be substantially stronger. Since the precise rationing rule ought to vary from industry to industry, the results of this paper should be interpreted cautiously, particularly with respect to the optimality of the particular dispersed pricing strategy presented.

The residual demand function is defined to be the demand at price $p$ given that all of the available units with a price less than or equal to $\bar{p}$ have already been sold. Under the proportional rationing rule, the residual demand function for $p \geq \bar{p} \geq p$ is given by

$$RD(p, e; \bar{p}, Q) = D(p, e) - \frac{Q}{Q_{\bar{p}, e}} \sum_{p_k \leq \bar{p}} \frac{D(p, e)}{D(p_k, e)} - \int_{\bar{p}}^{p} \frac{D(p, e)}{D(r, e)} q(r) dr.$$  \hspace{1cm} (2)

The ratio $D(p, e)/D(r, e)$ is the fraction of units sold at price $r$ which go to consumers who would also have been willing to pay $p$, and $Q$ denotes an admissible function $Q(p)$ defined by Eq. (1).

Given the demand state, $e$, and the price distribution, $Q(p)$, consumers will continue to purchase the lowest priced units still available from the monopolist as long as $RD(p, e; p, Q) > 0$. Clearly $RD(p, e; p, Q)$ is weakly decreasing in $p$. Let $\rho(e, Q)$ denote the market clearing price, defined as the price of the highest priced unit sold given the firm’s pricing strategy $Q(p)$ and the demand state $e$. So a price $p$ is the market clearing price if $RD(p, e; p, Q) \leq 0$ and $RD(\bar{p}, e; \bar{p}, Q) > 0, \forall \bar{p} < p$. More generally I define $\rho(e, Q)$ as $\sup \{ p; RD(r, e; r, Q) > 0, \forall r < p \}$.

Using this definition, I will henceforth use $\bar{p}$ to refer to $\rho(e, Q)$ since any higher priced units are never sold and are irrelevant to the analysis. Note that it is possible that $\rho(e, Q)$ exceeds $\bar{p}$, so more than the lowest priced units may sell in the lowest demand state, and it is possible that $\rho(e, Q) = \varphi(e) < \bar{p}$, so no sales occur in the lowest demand states. Let $\epsilon(p, Q)$ be the value of $e$ such that all units at prices less than or equal to $p$ are sold, so $\epsilon(p, Q)$ is the inverse of $\rho(e, Q)$,
when it is invertible, on \( \max\{p, \rho(e, Q)\}, \bar{p} \). I define \( \varepsilon(p, Q) = \varepsilon \) on \( [p, \rho(e, Q)] \) when \( p < \rho(e, Q) \).

Using (2) there are two possibilities when \( RD(p, e; p, Q) = 0 \): either \( D(p, e) = 0 \), or

\[
g(p, e, Q) = 1 - \frac{Q}{D(p, e)} - \sum_{p_k < p} \frac{Q_k}{D(p_k, e)} - \int_p^p q(r) \frac{D(r, e)}{D(p, e)} dr = 0.
\]  

(3)

For some systems of demand functions \( D(p, e) = 0 \) can be ruled out since \( D(p, e) > 0 \), \( \forall p \), however I show more generally in the appendix that for any admissible price distribution \( Q(p) \), \( D(p, \rho(e, Q)) > 0 \) for all \( p \in \left( \max\{p, \rho(e, Q)\}, \bar{p}\right) \) and so \( \varepsilon(p, Q) \) is always defined by \( g(p, \varepsilon(p, Q), Q) = 0 \) on \( \left[ \max\{p, \rho(e, Q)\}, \bar{p}\right] \). Note that \( \varepsilon(p, Q) \) is strictly increasing in \( p \) and is discontinuous at any price \( p_k \in \left( \max\{p, \rho(e, Q)\}, \bar{p}\right) \) where the firm has a mass point in its price distribution.

**MONOPOLY PRICING**

The monopolist’s profit function, given an admissible pricing strategy \( Q(p) \), is

\[
\pi(Q) = \int_{\varepsilon}^{\varepsilon(p, Q)} D(p, e)(p - c)f(e) de + \int_{\varepsilon}^{\rho(e, Q)} \left( Q(p - c) + \int_{\varepsilon}^{p} (p - c)q(p) dp \right)f(e) de + \sum_{k} \int_{\lim_{p \downarrow p_k} \varepsilon(p_k, Q)}^{\varepsilon(p_k, Q)} D(p_k, e) \left[ 1 - \frac{Q}{D(p_k, e)} - \sum_{p_j < p_k} \frac{Q_j}{D(p_j, e)} - \int_{p_k}^{p_k} q(p) \frac{D(p, e)}{D(p_k, e)} dp \right] (p_k - c)f(e) de. \]

(4)

The first term gives the portion of the firm’s profit that are earned when demand is so low that only some of the units offered at price \( p \) are sold. The second term gives the portion of the firm’s profit when demand is sufficiently large that all the units offered at price \( p \) are sold, as well as the other units priced below \( \rho(e, Q) \) which are part of the continuous portion of the firm’s price distribution. The third and fourth terms give the portions of the firm’s profits that are earned from
sales of units offered at the firm’s other mass points, \( \{p_k, Q_k\}; k = 1, \ldots, n \}. The third term measures the portion of the firm’s profit that is earned in demand states in which only some of the units at price \( p_k \) are sold, while the fourth term measures the portion of the firm’s profit that is earned when demand is sufficient to assure that all of the firm’s units at price \( p_k \) are sold. The expression for sales in the third term is from the residual demand function, and is equal to

\[
\lim_{p \uparrow p_k} RD(p_k, e; p, Q).
\]

The first order condition obtained from (4) and derived in an appendix is

\[
\left[ 1 - F(e(p,Q)) \right] (p - c) - \int_{e(p,Q)}^{\bar{e}} (\rho(e,Q) - c) \frac{D(\rho(e,Q),e)}{D(p,e)} f(e) de = 0, \ \forall p. \tag{5}
\]

The first term in (5) is the profit from an additional unit sold at price \( p \) times the probability that the unit is sold. The second term in (5) is the expected profit lost when an additional unit is sold at price \( p \), since selling the additional unit will reduce the demand for the firm’s units at the price \( \rho(e,Q) \). Note that the demand at the price \( \rho(e,Q) \) is reduced by less than one unit since only a fraction \( D(\rho(e,Q),e)/D(p,e) \) of the consumers who are willing to pay \( p \) would also willing to pay \( \rho(e,Q) \).

Differentiating (5) with respect to \( p \) yields

\[
\left[ 1 - F(e(p,Q)) \right] + \int_{e(p,Q)}^{\bar{e}} (\rho(e,Q) - c) \frac{D(\rho(e,Q),e)D_p(p,e)}{D(p,e)^2} f(e) de = 0, \ \forall p, \tag{6}
\]

which, when combined with (5), yields

\[
\frac{p - c}{p} = \left[ \frac{\int_{e(p,Q)}^{\bar{e}} w(p,e,Q) \eta(p,e) de}{\int_{e(p,Q)}^{\bar{e}} w(p,e,Q) de} \right]^{-1}, \ \forall p \in [p, \bar{p}], \tag{7}
\]

where

\[
w(p,e,Q) = (\rho(e,Q) - c) \frac{D(\rho(e,Q),e)}{D(p,e)} f(e). \tag{8}
\]
Eq. (7) offers some useful intuition for the monopolist’s incentive to set dispersed prices. It implies that the price of each unit will be set so that the price markup is equal to the inverse of a weighted average of the elasticities of demand in the demand states in which that unit will be sold. The first unit sold by the monopolist is sold whether demand is high or low, so it carries a lower price only if the elasticity of demand in low demand states is higher than the elasticity of demand in high demand states. Recall that the profit maximizing \textit{ex post} price (if the firm could set its spot price in response to demand shocks) satisfies \( \left( p''(e) - c \right) / p''(e) = 1 / \eta(p''(e), e) \), so (7) also suggests that if the \textit{ex post} spot monopoly price is increasing in the demand state, then the monopolist’s pricing strategy will be dispersed.

While Eq. (7) gives some useful intuition, it is difficult to interpret (7) as a definition of the firm’s pricing strategy since \( Q(p) \) does not appear directly in the equation. An easier interpretation of (7) is that it describes the relationship between the market clearing price and the realized demand state, \( e \). In other words, Eq. (7) defines \( \varepsilon(p,Q) \), and then Eq. (3) defines \( Q(p) \) in terms of \( \varepsilon(p,Q) \). From (7), the price markup is equal to the inverse of a weighted average of the elasticity of demand. As \( e \) increases, the weighted average is taken over smaller ranges and higher values of \( e \). So, intuitively, if \( \eta(p,e) \) is decreasing in \( e \), then the market clearing price should also be increasing in \( e \), and hence the firm’s pricing strategy is clearly dispersed.

**THEOREM 1:**\textit{ If the elasticity of demand is strictly decreasing in }\( e \textit{, so that the ex post monopoly spot price is increasing in the realization of the demand state, then the monopolist’s optimal pricing strategy is dispersed. The unique profit maximizing pricing strategy, }Q(p)\textit{, is continuous and differentiable on the interval }\left[ p, \bar{p} \right]\textit{ with at most one mass point at }\bar{p}\textit{, satisfying }Q = D(p,e)\textit{. where }\bar{p}\textit{ is defined by}

\[
\frac{\bar{p} - c}{\bar{p}} = \frac{1}{\eta(\bar{p}, e)},
\]

\textit{and }Q(p)\textit{ satisfies Eq. (7).}

Proof: See Appendix.
In the proof of Theorem 1 I show that the first order condition (5) has a unique non-degenerate solution and that the solution has at most one mass point and then only at the price \( p \). The proof that (5) has a unique solution is complicated by the fact that (5) is an integral equation and that the functions \( \varepsilon(p,Q) \) and \( \rho(e,Q) \) appear non-linearly and in the limits of integration, however in the proof of Theorem 1 I show that (5) can be converted into a system of non-linear Volterra integral equations of the second kind for which standard existence and uniqueness theorems apply.

Of course the solution must solve both (5) and (3) and Eq. (3) is also a Volterra integral equation (linear and of the first kind). The proof that (5) and (3) have a unique solution is also complicated by the possibility that \( D(p,e,Q) = 0 \) at the optimum since in this case the integral equation defined by (3) is singular at the price \( p \). Although this problem could have been avoided by restricting the analysis to systems of demand functions for which demand is always positive for relevant price levels (i.e., by assuming that \( e \) and \( \varphi(e) \) are sufficiently large), the proof of Theorem 1 establishes the existence and optimality of a dispersed pricing strategy more generally, including this more complicated but nevertheless important case. Many simple and important examples of demand systems fall into this category, including all multiplicatively separable demand systems, i.e., \( D(p,e) = d(p)e \) when \( e = 0 \).

Eq. (9) follows from (7), taking limits appropriately, and establishes that the price of the units that are sold only in state \( \bar{e} \) is \( p^m(\bar{e}) \), the \textit{ex post} monopoly price that would have prevailed in state \( \bar{e} \). This is clearly a consequence of the proportional rationing assumption. Because the monopoly price is invariant to a proportional reduction in demand (i.e., the monopoly price given demand \( d(p) \) is the same as the monopoly price given demand \( (1/2)d(p) \)) and the price in state \( \bar{e} \) is set to maximize profits given the residual demand function, it follows that the \textit{ex ante} and the \textit{ex post} prices must be the same.

Eq. (5) also has a degenerate solution in which all of the firm’s units at sold at the price \( p \), however I show in the appendix that this solution is not a profit maximum under the conditions of Theorem 1, and instead represents an inflection point of the firm’s profit function. Not
surprisingly, the conditions under which (5) does not have a non-degenerate, dispersed-price solution are also the conditions under which the single price solution is a profit maximum.

If the elasticity of demand is increasing in the demand state, then the monopolist’s optimal pricing strategy is to offer all of its units at a single price. The monopolist would like the units sold in all demand states to be higher than the price of units sold in only the high demand states, but this is not feasible because consumers will always purchase the lower priced units first. Thus, as I prove below in Theorem 2, if the elasticity of demand is weakly increasing in the demand state then the monopolist will set a single price for all of its units.

**Theorem 2:** If the elasticity of demand is weakly increasing in $e$ then the monopolist’s optimal pricing strategy is to set a single price, $p^*$, given by

\[
\frac{p^* - c}{p^*} = \left[ \frac{-\int_{e}^{\infty} D(p^*,e) \eta(p^*,e) f(e) de}{\int_{e}^{\infty} D(p^*,e) f(e) de} \right]^{-1},
\]

with maximum sales $Q^* = D(p^*,e)$.

Proof: See Appendix.

The intuition that yielded price distribution in Theorem 1 fails here because the firm does not want the units that are only sold in the highest demand states to be sold at higher prices. In fact, if the elasticity of demand is strictly decreasing, then the monopolist would like the units sold in the highest demand states to sell at strictly lower prices. Of course, if the monopolist were to
offer any of its units at a lower price they would be purchased in every state, not just the highest demand states. So in this case, the best the monopolist can do is to set a single price.

The elasticity of demand may be neither weakly increasing nor strictly decreasing in $e$, in which case it is possible that the monopolist's optimal pricing strategy would have mass points at prices other than $p$. In the case of weakly decreasing (but not weakly increasing) elasticity of demand, bunching or mass points might occur in states for which the elasticity of demand is constant. However a formal analysis of the case when the elasticity of demand is not monotonic in $e$ would be difficult, because of the resulting discontinuities in the residual demand function, and in $e$ and $\rho$.

**Sufficient Conditions**

From the standard first-order condition for a monopolist if the elasticity of demand is decreasing in $e$ then the monopolist’s *ex post* spot price is increasing in $e$, however the converse is not necessarily true. If the monopolist’s *ex post* spot price is increasing in $e$, then
\[
\frac{\partial \eta(p^\ast(e), e)}{\partial e} < 0,
\]
however $\eta(p, e)$ may be increasing in $e$ for other values of $p$ and $e$. So while the most powerful intuition for the results in the paper arise from thinking about the correlation between the *ex post* monopoly spot price and the level of the demand, in fact this is not a sufficient condition. Nevertheless in almost any empirical specification of a system of demand functions, the condition that the elasticity of demand be decreasing with the demand state is synonymous with the condition that the *ex post* monopoly spot price be increasing in the demand state, so the intuition is remarkably robust even though is not entirely accurate.

The following corollary gives an alternative sufficient conditions for price dispersion.

**COROLLARY:** *If the monopolist's marginal revenue function is decreasing in $e$, or*

\[
D_{pe} > D_e D_p / D,
\]
*then the profit maximizing *ex ante* pricing strategy is dispersed.*
The corollary follows immediately from differentiating $\eta(p,e)$, which implies that
\[ \frac{\partial \eta(p,e)}{\partial e} < 0 \] if and only if $D_{pe}D - D_pD_e > 0$, and from the definition of marginal revenue. Note that $D_{pe} > 0$ satisfies the sufficient condition in Corollary 2, but is not necessary.

In Figures 1, 2, and 3, I illustrate the monopolist’s optimal pricing strategy for the following three different systems of demand functions

(i) \[ D(p,e) = eh(p) \] (11)

(ii) \[ D(p,e) = (e-1)+h(p) \] (12)

(iii) \[ D(p,e) = h\left(\frac{p}{e}\right) \] (13)

where $e \in [1,2]$, and the function $h(p)$ is assumed to be linear. For ease of comparison I have constructed the examples so that the demand in the low state is the same in each case, i.e., $D(p,1) = h(p)$. Note that in the first example the elasticity of demand is constant in $e$, and in the second and third examples the elasticity of demand is decreasing in $e$.

As is shown in the Figures, the *ex post* monopoly spot price is increasing in $e$ in examples (ii) and (iii) and is constant in example (i). Accordingly, the *ex ante* optimal pricing strategy is dispersed in (ii) and (iii), and is degenerate in (i).
Example (i): \( D(p, e) = eh(p) \)

**FIGURE 1**

Example (ii): \( D(p, e) = (e - 1) + h(p) \)

**FIGURE 2**

Example (iii): \( D(p, e) = h\left(\frac{p}{e}\right) \)

**FIGURE 3**
CONCLUSION

Price dispersion of some form is relatively common among firms with market power and is often attributed to price discrimination. This paper presents an alternative explanation for the failure of the law of one price when firms have market power. I have shown that if the elasticity of demand is decreasing in the realization of demand, so that the spot market monopoly price is increasing in the realization of demand, then a profit maximizing monopolist facing uncertain demand will choose a disperse pricing strategy rather than set a single price. This sufficient condition holds for a broad class of demand functions including additively separable demand, and seems likely to be satisfied in many empirical applications, especially those in which a linear model of demand estimation fits well.

While the assumptions of the paper may be restrictive, the main result is consistent with a variety of monopoly pricing practices. In particular many firms offer sale items or "loss-leaders" in restricted quantities, while similar, if not identical, products are offered at higher prices and in much greater availability. The monopolist often stocks out of the lower priced units, leaving only higher priced units available for consumers to buy. This type of pricing is common in automobiles, appliances, and other household items.

This result may also explain the existence and nature of price dispersion in markets where firms print tickets for an event in advance, such as at a football stadium or theater. When demand is uncertain (and capacity is not constrained), then a profit maximizing seller may want to have its tickets printed at different prices, knowing that the lower priced tickets will be sold first, but that if demand turns out to be high, then the firm will earn a higher profit from the consumers who purchase their tickets later. Hence this result explains why stadiums frequently have different classes of seats at very different prices.

Of course seats in a stadium are typically heterogeneous goods while the model presents a theory of price dispersion with homogenous, not heterogeneous, goods. However the model can easily be interpreted to explain pricing of different quality seats. Suppose that some of the price variation that we observe between different types of seats is due to difference in the quality of seats
and that the monopolist is trying to set prices for each type of seat in advance of realizing its demand. The firm may choose to set different markups for different classes of seats in order to maximize its profit given the demand uncertainty. In low demand states, consumers buy only the lower mark-up seats while in high demand states consumers end up also buying the higher mark-up seats. It is natural that the higher quality seats would also have higher mark-ups since consumers who are willing to pay higher prices probably also have a greater willingness to pay for quality. Moreover, if firms are constrained by law or social custom not to sell identical goods at different prices, then the firm may actually have an incentive to artificially differentiate their goods in order to assign them different prices. Alternatively, the fact that some stadiums sell *different quality seats at the same price* may in fact be an example of price dispersion, since the highest quality seats will be sold first (smaller mark-up) and if demand is sufficiently large, the lower quality seats will be sold at the same price (higher mark-up). In each case, the stylized fact that is consistent with the model is that the availability of units with lower mark-ups is restricted, so that customers who buy late in the higher demand states end up paying higher mark-ups on their purchases.

The model has many limitations. First, the result applies only to monopoly pricing. If the model were generalized to allow competition with a homogeneous good, then the unique equilibrium would yield marginal cost pricing, even if demand were uncertain. However, the results presented here for monopoly could obviously be extended to a model of monopolistic competition.

Second, even if demand is uncertain, price dispersion may be due to a shortage of capacity and not to the relationship between realized demand and the elasticity of demand. Prescott (1975) and Eden (1990) show that in a perfectly competitive market when demand is uncertain and capacity is scarce or costly, then price dispersion is the unique competitive equilibrium, and in another paper, Dana (1993), I show that the same result holds for a monopolist facing uncertain demand and for a symmetric oligopoly, in the later case only if capacity costs are sufficiently large. However that result applies most readily to industries such as hotels and airlines since it requires
that firms have limited capacity or significant inventory holding costs. If firms can costlessly smooth production through inventories, then this explanation for price dispersion is not valid.

Finally, the result relies on the assumption of proportional, or random, rationing. Other rationing rules might yield different results. For example, if the monopolist were able to choose its rationing rule, then it would want to allocate the highest priced units to the consumers with the highest valuations. In other words, the monopolist would want to perfectly price discriminate. Alternatively, if a secondary market arose in which goods were purchased from the monopolist and then resold at a market clearing price, then the parallel rationing rule would be more appropriate, and would be equivalent to giving the lowest priced units to the consumers with the highest valuations. Under this rationing rule the monopolist is much less likely to have an incentive to charge dispersed prices. The reason is that the residual demand will generally contain only consumers with lower valuations, so it is unlikely that the firm will want to charge these consumers a higher price. Nevertheless, in the discrete example given in the introduction, the price dispersion result does not depend on the rationing rule.
APPENDIX 1

Proof of Theorem 1:

Let $\bar{e}(p,Q)$ be defined by

$$g(p,\bar{e}(p,Q),Q) = 1 - \frac{Q}{D(p,\bar{e}(p,Q))} - \sum_{p_i \leq p} \frac{Q_i}{D(p_i,\bar{e}(p,Q))} - \int_{\bar{Q}}^{p} \frac{q(r)}{D(r,\bar{e}(p,Q))} dr = 0, \tag{A1}$$

and let $\varepsilon(p,Q)$ be defined by $\varepsilon(p,Q) = \lim_{r \uparrow p} \bar{e}(p,Q)$.

CLAIM 1: For all admissible $Q(p)$ the function $\varepsilon(p,Q)$ is uniquely defined by

$g(p,\varepsilon(p,Q),Q) = 0$ on $\max\{\rho(\varepsilon,Q),p\}$ and by $\varepsilon(p,Q) = \varepsilon$ on $[p,\rho(e,Q)]$ when $p < \rho(e,Q)$; and satisfies $D(p,\varepsilon(p,Q)) > 0, \forall p \in \{\max\{\rho(e,Q),p\},\rho(e,Q)\}$. The function $\rho(e,Q)$ is the inverse of $\bar{e}(p,Q)$ whenever $\bar{e}(p,Q) = \varepsilon(p,Q)$ and is constant on $[\varepsilon(p,Q),\bar{e}(p,Q)]$ otherwise.

PROOF: Since $g(p,e,Q)$ has a derivative $g_e > 0$, is strictly increasing in $p$, and has a derivative $g_p < 0$ except at a finite number of prices $p_i$ where the firm has mass points, it follows that as long as the integrand in (A1) is bounded then any solution $\varepsilon(p,Q)$ to $g(p,e,Q) = 0$ at a price $p$ can be extended uniquely to the interval $[\rho(p),\bar{p}]$, where $\bar{p}$ is the price at which $\bar{e} = \varepsilon(p,q)$.

From the definition of $\varepsilon(p,Q)$, at $p$ either (A1) holds or $D(p,\varepsilon(p,Q)) = 0$. First consider the case in which $Q > 0$ and $D(p,e) \leq Q$, so (A1) holds with $D(p,\varepsilon(p,Q)) = Q$, $\varepsilon(p,Q)$ is well-defined, and $\varepsilon(p,Q) \geq \varepsilon$. By the Implicit Function Theorem, $g(p,e,Q) = 0$ has a unique solution in a local neighborhood of $\bar{p}$ and this solution can be extend to uniquely define $\varepsilon(p,Q)$ on $[p,\bar{p}]$. Since $q$ is strictly positive, it follows that for any $p$, there exists an $e > \varphi^{-1}(p)$ but sufficiently small such that $\int_{p}^{\bar{p}} q(r)/D(r,\varepsilon) dr > 1$ and $D(p,\varepsilon) > 0$ so $D(p,\varepsilon(p,Q)) > 0, \forall p \in [p,\bar{p}]$ and the integrand in (A1) will always remain bounded.

Next suppose that $Q \geq 0$ and $D(p,e) > Q$, so $\rho(e,Q) > p$. At $\varepsilon(\rho(e,Q),Q)$,
(A1) holds since \(D(\rho(\varepsilon, Q), \varepsilon(\rho(\varepsilon, Q), Q)) > 0\), so \(\varepsilon(\rho(\varepsilon, Q), Q)\) is well-defined and
\(\varepsilon(\rho(\varepsilon, Q), Q) \geq \varepsilon\) (which holds with equality if there is no mass point at \(\rho(\varepsilon, Q)\)). As above, by the Implicit Function Theorem, \(g(p, \varepsilon, Q) = 0\) has a unique solution in a local neighborhood of \((\rho(\varepsilon, Q), \varepsilon(\rho(\varepsilon, Q), Q))\) and this solution can be extend to uniquely define \(\varepsilon(p, Q)\) on \([\rho(\varepsilon, Q), \bar{p}]\).

Finally consider the case in which \(Q = 0\) and \(D(p, \varepsilon) = 0\). Clearly demand is positive for some price \(p \in [\underline{p}, \bar{p}]\) in demand state \(\varepsilon(p, Q)\), since \(p < \varphi(\varepsilon)\), so let \(\rho'\) and \(\varepsilon(\rho(\varepsilon, Q))\) be some point such that \(g(\rho(\varepsilon(\rho(\varepsilon, Q)), Q) = 0\). By the Implicit Function Theorem, \(g(p, \varepsilon, Q) = 0\) has a unique solution in a local neighborhood of \((\rho(\varepsilon(\rho(\varepsilon, Q))\)). And as above, this solution can be extend to uniquely define \(\varepsilon(p, Q)\) on \([\rho(\varepsilon, Q), \bar{p}]\) satisfying \(D(\rho(\varepsilon(\rho(\varepsilon, Q))), > 0, \forall p \in [\rho(\varepsilon, Q), \bar{p}]\).

Similarly, the solution in this case can be uniquely extended downwards to \(p\). For any \(p > \underline{p}\), there exists an \(\delta > \varphi^{-1}(p)\) but sufficiently small such that \(\int_{\underline{p}}^{p} q(r) / D(r, \varepsilon) dr > 1\), so \(D(p, \varepsilon(p, Q)) > 0\) and the integrand in (A1) will remain bounded. There exists a \(\delta > 0\) and an interval \([\underline{p}, \underline{p} + \delta]\) that contains no mass points and on which \(g(p, \varepsilon, Q)\) is continuous in \(p\) and \(\varepsilon\), so the solution \(\varepsilon(p, Q)\) is continuous in \(p\) by the Implicit Function Theorem. Since \(Q = 0\), \(\varepsilon(p, Q)\) must satisfy \(\lim_{p \downarrow \underline{p}} \int_{\underline{p}}^{p} q(r) / D(r, \varepsilon(p, q)) dr = 1\); that is, the integrand must be singular at \(\underline{p}\). So \(D(p, \varepsilon(p, Q)) = 0\). Using Lemma 2 in Appendix 2, at \(p\) the solution \(\varepsilon(p, Q)\) solves

\[
\lim_{p \downarrow \underline{p}} \int_{\underline{p}}^{p} \frac{q(r)}{D(r, \varepsilon(p, Q))} dr = \frac{q(\underline{p})}{D(\underline{p}, \varepsilon(\underline{p}, Q))} \ln \left( \frac{D(\underline{p}, \varepsilon(\underline{p}, Q)) + D_{\varepsilon}(D(\underline{p}, \varepsilon(\underline{p}, Q)) \varepsilon(\underline{p}, Q))}{D(\underline{p}, \varepsilon(\underline{p}, Q))} \right) = 1
\]
since
\[ D_p \left(p, \varepsilon(p, Q)\right) (r - p) + D_e \left(p, \varepsilon(p, Q)\right) \frac{d \varepsilon(p, Q)}{dp} (p - p) > 0 \] (A3)

and since
\[ \frac{d \varepsilon(p, Q)}{dp} > \frac{d \varphi^{-1}(p)}{dp} = -\frac{D_p \left(p, \varepsilon(p, Q)\right)}{D_e \left(p, \varepsilon(p, Q)\right)}. \] (A4)

Hence \( \varepsilon(p, Q) \) is defined by \( g(p, \varepsilon(p, Q), Q) = 0 \) on \( \left[ \max \{\rho(e, Q), p\}, \rho(p, Q)\right] \).

Recall that there are two types of demand systems considered for which it is possible that \( D(p, \varepsilon(p, Q)) = 0 \). In case (ii), \( \varphi(e) = \overline{\varphi} \) and all demand functions have the same price-intercept. By assumption, \( p < \overline{\varphi} \), so \( D(p, e) > 0 \) for all \( e > e \). Therefore \( D(p, \varepsilon(p, Q)) = 0 \) implies that \( \varepsilon(p, Q) = e \). Next consider case (iii), in which \( \varphi(e) \) is invertible. By assumption \( p < \varphi(\overline{e}) \). If \( p \in [\varphi(e), \varphi(\overline{e})] \), then \( \varepsilon(p, Q) = \varphi^{-1}(p) \), or, if \( p \leq \varphi(e) \), then, \( \varepsilon(p, Q) = e \). So in Case (iii) \( D(p, \varepsilon(p, Q)) = 0 \) implies that \( \varepsilon(p, Q) = \varphi^{-1}\left(\max\{p, \varphi(e)\}\right) \).

So \( \varepsilon(p, Q) = \overline{\varepsilon}(p, Q) \) and \( \varepsilon(p, Q) \) is uniquely defined by \( g(p, \varepsilon(p, Q), Q) = 0 \).

Since the function \( \varepsilon(p, Q) \) is strictly increasing in \( p \) and is continuous in \( p \) on intervals that do not contain a mass point, \( RD(p, \varepsilon(p, Q); p, Q) = 0 \), and the market clearing price \( \rho(e, Q) \) is the inverse of \( \varepsilon(p, Q) \) whenever \( \overline{\varepsilon}(p, Q) = \varepsilon(p, Q) \). On the interval \( \left[ \varepsilon(p_k, Q), \overline{\varepsilon}(p_k, Q)\right] \) when the firm’s price distribution contains a mass point at \( p_k \), \( \sup\{p; RD(r, e; r, Q) > 0, \forall r < p\} \) is clearly constant, so \( \rho(e, Q) = p_k \) for all \( e \in \left[ \varepsilon(p_k, Q), \overline{\varepsilon}(p_k, Q)\right] \). Thus \( \rho(e, Q) \) is also uniquely defined and weakly increasing in \( e \). Q.E.D.
CLAIM 2: For all strictly increasing differentiable functions $\mathcal{E}(p)$ on $[\underline{p}, \bar{p}]$, the function $Q(p)$ is uniquely defined by $g(p, \mathcal{E}(p), Q) = 0$.

PROOF: Given $\mathcal{E}(p)$, $g(p, \mathcal{E}(p), Q) = 0$ and (A1) imply that either $Q = D\left(p, \mathcal{E}(p)\right) > 0$ and that $q(r)$ is defined by

$$1 - \frac{Q}{D(p, \mathcal{E}(p))} = \int_{\underline{p}}^{p} \frac{q(r)}{D(r, \mathcal{E}(p))} dr,$$

(A5)

or that $Q = D\left(p, \mathcal{E}(p)\right) = 0$ and $q(r)$ is defined by

$$1 = \int_{\underline{p}}^{p} \frac{q(r)}{D(r, \mathcal{E}(p))} dr,$$

(A6)

where (A5) and (A6) are Volterra integral equations of the first kind. Note that because $\mathcal{E}(p)$ is continuous, $Q(p)$ cannot contain mass points except at $p$.

Differentiation of (A5) with respect to $p$ using $Q = D\left(p, \mathcal{E}(p)\right) > 0$ yields

$$q(p) = D\left(p, \mathcal{E}(p)\right)k\left(p, p\right) + \int_{\underline{p}}^{p} k(p, r) q(r) dr,$$

(A7)

where:

$$k(p, r) = D\left(p, \mathcal{E}(p)\right) \frac{D\left(r, \mathcal{E}(p)\right) \mathcal{E}'(p)}{D\left(r, \mathcal{E}(p)\right)^2}.$$

(A8)

Eq. (A7) is a Volterra integral equation of the second kind, and since $D\left(r, \mathcal{E}(p)\right) > 0$,

$\forall p, r; \underline{p} \leq r \leq p \leq \bar{p}$, the kernel, $k(p, r)$, and forcing function, $D\left(p, \mathcal{E}(p)\right)k\left(p, p\right)$, satisfy all of the standard conditions for existence and uniqueness of a continuous solution $q(p)$, as given in Miller (1971) and Gripenberg, et. al. (1990). Note that the solution to (A7) is unique in a broader class a functions (e.g., Lesbegue integrable) then is the solution to (A5), since any function that is identical to $q(p)$ except on a set of measure zero is a solution to (A5) but not to (A7).
Eq. (A6) also has a unique continuous solution, however the proof is more complicated since the integrand of (A6) is unbounded when \( p = p_\ast \).

Any continuous solution to (A6) must satisfy

\[
q(p) = \frac{1}{\left[ \lim_{p \downarrow p_\ast} \int_p^\infty \frac{1}{D(r, \vartheta(p))} dr \right]}.
\]  

(A9)

In Proposition of Appendix 2, I establish that (A6) has a unique continuous solution, since

\[
\lim_{p \downarrow p_\ast} D_r(p, \vartheta(p)) \leq 0, \quad \lim_{p \downarrow p_\ast} D_r(p, \vartheta(p))\vartheta'(p) > 0 \quad \text{and} \quad \lim_{p \downarrow p_\ast} \left( D_r(p, \vartheta(p)) + D_e(p, \vartheta(p))\vartheta'(p) \right) > 0 \quad \text{(see below)}.
\]

To show that \( q(p) \) is finite, I consider two cases. The first is when \( \lim_{p \downarrow p_\ast} D_r(p, \vartheta(p)) < 0 \) and the second is when \( \lim_{p \downarrow p_\ast} D_r(p, \vartheta(p)) = 0 \). The first case can arise only when the system of demand functions satisfies Case (ii), i.e., the price intercept varies continuously in \( \epsilon \). In this case it is possible that \( \vartheta(p) = \varphi^{-1}(p) \) and so \( D(p, \vartheta(p)) = 0 \). The second case can arise for any of the three cases since I have allowed the demand curve to be degenerate at \( \epsilon \), however when \( \lim_{p \downarrow p_\ast} D_r(p, \vartheta(p)) = 0 \) it must be the case that \( \vartheta(p) = \epsilon \).

When \( \lim_{p \downarrow p_\ast} D_r(p, \vartheta(p)) < 0 \), using Lemma 2 in Appendix 2,

\[
\lim_{p \downarrow p_\ast} \int_p^\infty \frac{q(r)}{D(r, \vartheta(p))} dr = q(p) \lim_{p \downarrow p_\ast} \int_p^\infty \frac{1}{D_r(p, \vartheta(p))(r - p) + D_e(p, \vartheta(p))\vartheta'(p)(p - p)} dr
\]

\[
= \frac{q(p)}{D_r(p, \vartheta(p))} \ln \left( \frac{D_r(p, \vartheta(p)) + D_e(p, \vartheta(p))\vartheta'(p)}{D_r(p, \vartheta(p))\vartheta'(p)} \right) = 1
\]

(A10)

so

\[
q(p) = \frac{D_r(p, \vartheta(p))}{\ln \left( \frac{D_r(p, \vartheta(p)) + D_e(p, \vartheta(p))\vartheta'(p)}{D_r(p, \vartheta(p))\vartheta'(p)} \right)}
\]

(A11)

which is strictly positive as long as \( D_r(p, \vartheta(p)) + D_e(p, \vartheta(p))\vartheta'(p) > 0 \), or

\[
\vartheta'(p) > -\frac{D_r(p, \vartheta(p))}{D_r(p, \vartheta(p))} = -\frac{D_r(p, \varphi^{-1}(p))}{D_r(p, \varphi^{-1}(p))} = \frac{d\varphi^{-1}(p)}{dp}.
\]

(A12)
And since $\xi(p)$ is continuous and satisfies $D(p, \xi(p)) > 0$ for all $p \in [\underline{p}, \overline{p}]$, it follows that $\xi'(p) > d\phi^{-1}(p)/dp$, so $D_p \left(p, \xi(p)\right) + D_e \left(p, \xi(p)\right)\xi'(p) > 0$.

When $\lim_{p \times \underline{p}} D_p \left(p, \xi(p)\right) = 0$,

$$\lim_{p \times \underline{p}} \int_{\underline{p}}^{\overline{p}} \frac{q(r)}{D(r, \xi(p))} dr = q(p) \lim_{p \times \underline{p}} \int_{\underline{p}}^{\overline{p}} \frac{1}{D_e \left(p, \xi(p)\right)\xi'(p)} (p-r) dr = \frac{q(p)}{D_e \left(p, \xi(p)\right)\xi'(p)} = 1$$

so

$$q(p) = \frac{1}{D_e \left(p, \xi(p)\right)\xi'(p)}.$$  \hfill (A14)

In both cases it is clear that $q(p)$ is strictly positive and finite. Q.E.D.

Using Claim 1 the monopolist's optimization can be rewritten as follows: the monopolist chooses $p$, $Q$, $\{(p_k, Q_k): k = 2, \ldots, n\}$, and $q(p)$ to maximize its expected profit function:

$$\pi(Q) = \int_{\underline{p}}^{\overline{p}} D(p, e) (p-c) f(e) de$$

$$+ \int_{\underline{p}}^{\overline{p}} \left[ Q(p-c) + \int_{\underline{p}}^{\overline{p}} (p-c)q(p) dp \right] f(e) de$$

$$+ \sum_{k} D(p_k, e) \left[ Q_k - \sum_{\rho_k < p_k} \frac{Q_j}{D(p_j, e)} - \frac{Q_k}{D(p_k, e)} \int_{\underline{p}}^{\overline{p}} \frac{q(p)}{D(p, e)} dp \right] (p_k-c) f(e) de$$

$$+ \sum_{k} \int_{\rho_k}^{\overline{p}} Q_k (p_k-c) f(e) de.$$  \hfill (A15)

where the limits of integration are stated in terms of $\overline{e}(p, Q)$ and $\underline{\xi}(p, Q)$ instead of $\xi(p, Q)$.

The first order condition is derived from the profit function using first principles. If $Q(p)$ is the optimal pricing strategy of the firm, and $\pi(Q)$ is the expected profit function, given above, then $d\pi(Q + aH)/da|_{a=0} = 0$ for all functions $H(p)$ such that $Q(p) + aH(p)$ satisfies (1). In a supplemental appendix, I use this methodology to derive the necessary first order conditions using
(A1) to define the derivatives of \(\varepsilon(p,Q)\) and \(\rho(e,Q)\) with respect to \(a\). I treat \(p\) as a variable endpoint and allow the problem to have an infinite upper endpoint since any solution \(Q(p)\) on \([p,\infty)\) becomes a solution on \([p,\bar{p}]\) by setting \(\bar{p} = \rho(\bar{e},Q)\). The first of two necessary first order conditions is

\[
[1 - F(\varepsilon(p,Q))](p-c) - \int_{\varepsilon(p,Q)}^r (\rho(e,Q) - c) \frac{D(\rho(e,Q),e)}{D(p,e)} f(e) \, de = 0, \tag{A16}
\]

for all \(p \in [p,\bar{p}]\). Note that Eq. (A16) can also obtained directly by differentiating (A15) with respect to \(Q_k\) and by differentiating (A15) with respect to \(Q\), using (A1) to define the derivatives of \(\varepsilon(p,Q)\) and \(\rho(e,Q)\) with respect to \(Q_k\) and \(Q\). So (5), (6) and (7) hold for all \(p \in [p,\bar{p}]\).

The first order condition, (A16), must be equal to zero, and hence must be constant, for all \(p\). After a change of variables, this first order condition can be rewritten as

\[
[1 - F(\bar{\varepsilon}(p,Q))](p-c) - \lambda^i - \int_p^\infty (r-c) \frac{D(r,\bar{\varepsilon}(r,Q))}{D(p,\bar{\varepsilon}(r,Q))} f(\bar{\varepsilon}(r,Q)) \frac{d\bar{\varepsilon}(r,Q)}{dr} \, dr
\]

\[- \sum_{p_k > p} \int_{\varepsilon(p_k,Q)}^{\varepsilon(p,Q)} (p_k - c) \frac{D(p_k,e)}{D(p,e)} f(e) \, de = 0, \tag{A17}
\]

and its derivative is

\[
[1 - F(\bar{\varepsilon}(p,Q))] - \lambda - \int_p^\infty (r-c) \frac{D_x(\rho,\bar{\varepsilon}(r,Q))D(\rho,\bar{\varepsilon}(r,Q))}{D(p,\bar{\varepsilon}(r,Q))^2} f(\bar{\varepsilon}(r,Q)) \frac{d\bar{\varepsilon}(r,Q)}{dr} \, dr
\]

\[+ \sum_{p_k > p} \int_{\varepsilon(p_k,Q)}^{\varepsilon(p,Q)} (p_k - c) \frac{D(p_k,e)D(p_k,e)}{D(p,e)^2} f(e) \, de = 0. \tag{A18}
\]

Taking the difference between this derivative evaluated at \(p_k\) and at \(\lim_{p \uparrow p_k} f(e) = 0\) yields

\[
\int_{\varepsilon(p_k,Q)}^{\varepsilon(p,Q)} 1 + (p_k - c) \frac{D_x(p_k,e)}{D(p_k,e)} f(e) \, de = 0. \tag{A19}
\]

Using (7), which holds whether or not the distribution contains mass points, \(\partial \eta(p,e)/\partial e < 0\) implies that \((p-c)/p > 1/\eta(p,e)\), for all \(e < \varepsilon(p,Q)\), which implies that

\[1 + (p-c)D_x(p,e)/D(p,e) < 0, \quad \text{for all } e < \bar{\varepsilon}(p,Q), \text{ so for (A19) to hold, it must be true that}\]
\( \bar{e}(p, Q) = \bar{e}(p, \bar{Q}) \) and that there is no mass point at \( p_k \). Since this argument applies at all prices \( p_k \in \bar{p}, \bar{p} \), the optimal pricing strategy of the monopolist has no mass points other than a single potential mass point at \( \bar{p} \).

The second first order condition derived in the supplemental appendix, which can be obtained directly from (A1) by differentiating with respect to \( p \), is

\[
\int_{\bar{p}}^{\bar{Q}} \left( D(p, e) + D_p(p, e)(p - c) \right) f(e) de + \int_{\bar{p}}^{\bar{Q}} \left( \rho(\rho(e, Q)e)q(\rho(e, Q)) + \rho(\rho(e, Q)e)q(\rho(e, Q)) \right) \frac{D(\rho(e, Q)e)}{D(\rho(e, Q))} de = 0
\]

which can be simplified using (A17) and (A18) to

\[
\int_{\bar{p}}^{\bar{Q}} \left( D(p, e) - \frac{p - c}{p} D(p, e) \eta(p, e) \right) f(e) de = 0. \tag{A21}
\]

Eq. (A21) implies that either (i) \( \bar{e}(p, Q) = \bar{e} \) and \( Q = D(p, e) \), (ii) \( D(p, e) = 0 \), \( \forall e \in \left[ e, \bar{e}(p, Q) \right] \) and \( Q = D(p, e) = D(p, e(p, Q)) = 0 \), or (iii) that \( \bar{e}(p, Q) > \bar{e} \), and, from (A21),

\[
\frac{p - c}{p} = \frac{\int_{\bar{p}}^{\bar{Q}} D(p, e) f(e) de}{\int_{\bar{p}}^{\bar{Q}} D(p, e) \eta(p, e) f(e) de}. \tag{A22}
\]

However (iii) is not consistent with the assumption that the elasticity of demand is decreasing in \( e \).

Suppose that \( \bar{e} > \bar{e}(p, Q) > e \), then (A22) implies that the markup at \( \bar{p} \) is equal to the inverse of a weighted average of the elasticity of demand between \( e \) and \( \bar{e}(p, Q) \). However, we have shown that (7) holds for \( p = p \), so we also have that the markup at \( \bar{p} \) must be equal to the inverse of a weighted average of the elasticity of demand between \( \bar{e}(p, Q) \) and \( \bar{e} \). But, if the elasticity of demand is strictly decreasing, then this is impossible.
Now suppose instead that \( \varepsilon(p, Q) = \bar{e} \), that is that the firm’s price distribution is degenerate. This is the case in which the monopolist's pricing strategy is a uniform price, that is, all of the firm's sales are at \( p \). This solution clearly satisfies both first order conditions, (A16) and (A21), however it is not a profit maximum. Consider the left-hand side derivative of the profit function with respect to \( Q \) when \( \varepsilon(p, Q) = \bar{e} \) (the right hand side-derivative is necessarily zero). The first derivative, which was stated above as (A16), is

\[
\left[ 1 - F(\bar{\varepsilon}(p, Q)) \right] (p - c) - \int_{\varepsilon(p, Q)}^{\bar{e}} (\rho(e, Q) - c) \frac{D(\rho(e, Q), e)}{D(p, e)} f(e) de = 0, \tag{A23}
\]

which can be obtained directly from (A15) by differentiating with respect to \( Q \) using (A1) to define the derivative \( d\rho(e, Q)/dQ = -\left(1/D(p, e)\right)\left(D(\rho(e, Q), e)/q(\rho(e, Q), e)\right) \). The first term of the derivative is equal to zero since \( F(\bar{\varepsilon}(p, Q)) = F(\bar{e}) = 1 \) and the second is equal to zero since \( \varepsilon(p, Q) = \bar{e} \). The second derivative of the profit function with respect to \( Q \) is

\[
\int_{\varepsilon(p, Q)}^{\bar{e}} \left[ \frac{(\rho(e, Q) - c)D_p(\rho(e, Q), e) + D(\rho(e, Q), e)}{D(p, e)} \right] \frac{D(\rho(e, Q), e)}{q(\rho(e, Q))} f(e) de = 0, \tag{A24}
\]

which is also zero since \( \varepsilon(p, Q) = \bar{e} \). The third derivative, simplified using \( \varepsilon(p, Q) = \bar{e} \), is

\[
\left[ \frac{(p - c)D_r(p, \bar{\varepsilon}(p, Q)) + D(p, \bar{\varepsilon}(p, Q))}{D(p, \bar{\varepsilon}(p, Q))D_r(p, \bar{\varepsilon}(p, Q))} \right] q(\rho(e, Q)) f(\bar{\varepsilon}(p, Q)), \tag{A25}
\]

which is negative under the conditions of Theorem 1 (using (7), \( \partial \eta(p, e)/\partial e < 0 \) implies that \( (p - c)D_r(p, \bar{\varepsilon}(p, Q)) + D(p, \bar{\varepsilon}(p, Q)) < 0 \)). Hence the profit of the monopolist increases as \( Q \) is decreased when the profit function is evaluated at the degenerate solution.

Having ruled out case (iii), we therefore have either (i) \( \varepsilon(p, Q) = \varepsilon \) and \( Q = D(p, e) \), or (ii) \( D(p, \varepsilon(p, Q)) = 0 \), so \( D(p, e) = 0 \), \( \forall e \in [\varepsilon, \varepsilon(p, Q)] \).
Recall that Eq. (6) is obtained by differentiation of (5), and (7) is obtained directly from (5) and (6), and from (7),

\[
\frac{\bar{\rho} - c}{\bar{\rho}} = \frac{1}{\eta(\bar{\rho}, \bar{\varepsilon})},
\]  

which implies that \(\bar{\rho} = \rho^m(\bar{\varepsilon})\).

The next step of the proof is to establish that \(q(p)\) is uniquely defined by the first order condition, (A16). Let \(x(p) = F(\varepsilon(p, Q))\). Making this substitution, and changing the variable of integration, (A16) can be written as:

\[
x(p) = 1 - \int_{\rho}^{\bar{\rho}} \frac{(r-c)}{(p-c)} \frac{D(r, F^{-1}(x(r)))}{D(p, F^{-1}(x(r)))} x'(r) dr,
\]

which is a nonlinear integral-differential equation. Differentiating (A27) with respect to \(p\) yields:

\[
0 = \int_{\rho}^{\bar{\rho}} \frac{(r-c)}{(p-c)^2} \frac{D(r, F^{-1}(x(r)))}{D(p, F^{-1}(x(r)))} \left[ D_p(p, F^{-1}(x(r)))(p-c) + D(p, F^{-1}(x(r))) \right] x'(r) dr,
\]

and differentiating again with respect to \(p\) yields:

\[
\left[ \frac{D_p(p, F^{-1}(x(p)))(p-c) + D(p, F^{-1}(x(p)))}{(p-c)D(p, F^{-1}(x(p)))} \right] x'(p) = \int_{\rho}^{\bar{\rho}} \frac{(r-c)}{(p-c)} D(r, F^{-1}(x(r))) \left[ \frac{D_{pp}(p, F^{-1}(x(r)))(p-c) + 2D_p(p, F^{-1}(x(r)))}{(p-c)^2 D(p, F^{-1}(x(r)))^2} \right. \\
\left. - 2 \left[ D_p(p, F^{-1}(x(r)))(p-c) + D(p, F^{-1}(x(r))) \right] \right] x'(r) dr
\]

Let \(y(p)\) be defined by:

\[
y(p) = \left[ \frac{D_p(p, F^{-1}(x(p)))(p-c) + D(p, F^{-1}(x(p)))}{(p-c)D(p, F^{-1}(x(p)))} \right] x'(p),
\]
then substituting \( y(p) \) into (A27) yields

\[
x(p) = 1 - \int_p \left( \frac{(r-c)D(r,F^{-1}(x(r)))}{D(p,F^{-1}(x(r)))} \left( \frac{(r-c)D(r,F^{-1}(x(r)))}{D(r,F^{-1}(x(r)))} (r-c)D(r,F^{-1}(x(r))) + D(r,F^{-1}(x(r))) \right) y(r) \right) dr,
\]

(A31)

and substituting \( y(p) \) into (A29) yields

\[
y(p) = \int_p \left[ \frac{D_{pp}(p,F^{-1}(x(r)))(p-c) + 2D_p(p,F^{-1}(x(r)))}{(p-c)^2D(p,F^{-1}(x(r)))^2} \right.
\]

\[
- \frac{2D_p(p,F^{-1}(x(r)))(p-c) + D(p,F^{-1}(x(r)))}{(p-c)^3D(p,F^{-1}(x(r)))^3} \left( \frac{(r-c)^2D(r,F^{-1}(x(r)))^2}{D_p(r,F^{-1}(x(r)))(r-c) + D(r,F^{-1}(x(r)))} y(r) \right) dr
\]

(A32)

The result of this extensive manipulation is that Eq. (A31) and Eq. (A32) are a system of nonlinear Volterra integral equations of the second kind. It is straight-forward to prove that the integrands of the two equations are continuous in \( x, y, r, \) and \( p, \) provided that

\[ D(p,F^{-1}(x(p))) > 0, \]

or equivalently that \( D(p,e(p,Q)) > 0. \) Hence standard theorems (see Theorems 12.2.1 and 12.2.8 in Gripenberg, et. al., 1990, and those in Miller, 1971) on the existence and uniqueness of solutions to systems of nonlinear Volterra integral equations of the second kind can be used to show that there exists a unique solution \( (x(p), y(p)) \) to (A31) and (A32).

Clearly \( D(\bar{p},F^{-1}(x(\bar{p})) > 0 \) since \( \bar{p} \) is the \textit{ex post} monopoly price in demand state \( F^{-1}(x(\bar{p})) = \bar{e}. \) So (A31) and (A32) have a unique solution near \( \bar{p}. \) Moreover from Theorem 12.2.1 in Gripenberg, et. al. (1990) this solution can be extended (downwards) until either

\( x(p) = 0, \) so \( F^{-1}(x(p)) = e, \) in which case \( \bar{p} \) is defined by \( x(\bar{p}) = 0, \) or until

\( D(p,F^{-1}(x(p))) = 0, \) in which case \( \bar{p} \) is defined by \( D(\bar{p},F^{-1}(x(\bar{p}))) = 0. \) Though the kernel of both integral equations is singular when \( D(\bar{p},F^{-1}(x(\bar{p}))) = 0, \) it is straight forward to verify that the integral is continuous in \( p, \) even at \( \bar{p}. \) The simplest case in which \( D(\bar{p},F^{-1}(x(\bar{p}))) = 0 \) would
arise is if demand were multiplicatively separable, \( D(p,e) = d(p)e \), and \( e = 0 \). In this case \( D\left[p, F^{-1}(x(p)) \right] \) must be zero since \( D(p,e) = 0 \) for all prices.

Given \( x(p) \), \( q(p) \) is uniquely determined by \( g\left[p, F^{-1}(x(p)), Q\right] = 0 \) using Claim 2. It must be verified, however that \( q(p) > 0 \).

Totally differentiating (A1), using \( Q_k = 0 \), yields

\[
\frac{d\bar{e}(p,Q)}{dp} = \frac{q(p)}{D(p,\bar{e}(p,Q))} \cdot \left[ \frac{D\left[p, \bar{e}(p,Q) \right]}{D(p,\bar{e}(p,Q))} Q + \int_{\bar{e}(p,Q)}^{r} \frac{D_e (r, \bar{e}(r,Q))}{\bar{e}} \frac{q(r)dr}{\left[D(r,\bar{e}(r,Q))\right]^2} \right].
\] (A33)

so \( d\bar{e}(p,Q)/dp \) and \( q(p) \) have the same sign. Differentiating (6) with respect to \( p \) yields:

\[
\left[ 1 + (p - c) \frac{D_r (p, \bar{e}(p,Q))}{D(p,\bar{e}(p,Q))} \right] f(\bar{e}(p,Q)) \frac{d\bar{e}(p,Q)}{dp} = \int_{\bar{e}(p,Q)}^{r} \left( \rho(e,Q) - c \right) D(\rho(e,Q),e) \left[ \frac{D_{pp}(p,e)D(p,e) - 2D_p(p,e)^2}{D(p,e)^3} \right] f(e)de.
\] (A34)

By assumption the right hand side of (A34) is negative, so \( q(p) > 0 \) if and only if

\[
1 + (p - c) D_r (p, \bar{e}(p,Q))/D(p,\bar{e}(p,Q)) < 0 \text{ for all } p \in [p, \bar{p}] \text{ or equivalently,}
\]

\[
(p - c)/p > 1/\eta(p, \bar{e}(p,Q)).
\]

Using (7), \( \partial \eta(p,e)/\partial e < 0 \) is clearly a sufficient condition for \( d\bar{e}(p,Q)/dp > 0 \) and \( q(p) > 0 \) for all \( p \in [p, \bar{p}] \).

The second order condition for the choice of \( Q(p) \) is derived in the supplemental appendix, using the same techniques used to derive the first order condition. The second-order sufficient condition is

\[
\frac{d^2 \pi^2 (q + ah)}{da^2} \bigg|_{a=\alpha} = \int_{\bar{e}(p,Q)}^{\eta(p,Q)} \left[ \left( \rho(e,Q) - c \right) D(\rho(e,Q),e)^2 + D_p(\rho(e,Q),e)D(\rho(e,Q),e) \right] \frac{q(\rho(e,Q))}{\left[ D(p,e) \right]^3} \left[ \int_{\rho(e,Q)}^{\eta(p,Q)} \left[ \frac{h(p)}{D(p,e)} \right] dp + \frac{H}{D(p,e)} \right] f(e)de < 0
\] (A35)
for all $H(p)$. Since (7) implies that
\[
\frac{(\rho(e,Q) - c)}{\rho(e,Q)} > \frac{1}{\eta(\rho(e,Q), e)} = -\frac{D(\rho(e,Q), e)}{\rho(e,Q)D_p(\rho(e,Q), e)},
\]  
(A36)
which implies that $D(\rho(e,Q), e) + (\rho(e,Q) - c)D_p(\rho(e,Q), e) < 0$, the second order condition is clearly satisfied when the elasticity of demand is strictly decreasing in $e$.

**Proof of Theorem 2:**

Eq. (A21), the first order condition obtained by differentiating (A1) with respect to $p$, implies that either: (i) $\varepsilon(p,Q) = \varepsilon$, in which case $p < \bar{p}$ and the firm’s pricing strategy is dispersed with mass $Q = D(p,e)$ at price $p$, (ii) $\varepsilon < \varepsilon(p,Q) < \bar{\varepsilon}$ in which case the firm’s pricing strategy is dispersed with mass $Q > D(p,e)$ at price $p$, or (iii) $\varepsilon(p,Q) = \bar{\varepsilon}$, in which case the optimal price distribution is degenerate. Cases (i) is easily ruled out since the second order condition, (A35), is not satisfied when the elasticity of demand is non-decreasing in $e$. And Case (ii) is ruled out by the same argument given in Theorem 1. From (6), $\partial \eta(p,e)/\partial e \geq 0$ implies $(p - c)/p < 1/\eta(p,e)$, which, using (A33), implies that $\partial \varepsilon(p,Q)/\partial p < 0$ and $q(p) < 0, \forall p \in [p, \bar{p}]$, which is a contradiction. So the monopolist’s profit maximization problem does not have a disperse price optimum.

The single price optimum, Case (iii), is the only remaining alternative. The second order condition when $\varepsilon(p,Q) = \bar{\varepsilon}$, derived in the Supplemental Appendix, is
\[
\left. \frac{d \pi^2(Q + aH)}{d a^2} \right|_{a=0} = (\delta p) \int \left[ D_p(p,e)(p - c) + 2D_p(p,e) \right] f(e)de < 0 \quad (A37)
\]
for all $H(p)$, where $\delta p$ represents the change in $p$. Clearly this is negative for all $\delta p$, and hence the single price optimum given by the first order condition (A21) is a profit maximum in the class of degenerate distributions. However, while this condition is clearly non-positive more generally, it is not strictly negative (e.g., let $h(p) > 0, \forall p \in [p, \infty)$ and $\delta p = 0$). A complete analysis of the third and fourth derivatives would reveal that this is indeed an optimum in the class of dispersed price distributions (the solution to the first order condition is unique) and it is reassuring to note
that unlike the case when elasticity is decreasing in $e$, when the elasticity is weakly decreasing in $e$ the firm’s profit decreases as $Q$ is decreased at the optimum (see Eq. (A25)). Since the solution is degenerate, the degenerate price is given by (A22),

$$\frac{p^* - c}{p^*} = \frac{\int_\varepsilon^\zeta D(p^*, e)f(e)de}{\int_\varepsilon^\zeta D(p^*, e)\eta(p^*, e)f(e)de},$$

(A38)

and the equilibrium output is $Q^* = D(p^*, \bar{e})$. 
APPENDIX 2

The following Lemmas are used in the proof of Proposition 1 below.

**LEMMA 1:** If \( m(t,s) \) is a twice continuously differentiable function of \( s \) and \( t \), and if \( \lim_{t \downarrow 0} \int_0^t h(t,s)ds \) is finite, then

\[
\lim_{t \downarrow 0} \int_0^t h(t,s)m(t,s)ds = m(0,0)\lim_{t \downarrow 0} \int_0^t h(t,s)ds .
\] (B1)

**Proof of Lemma 1:**

This result follows directly from the mean value theorem:

\[
\lim_{t \downarrow 0} \int_0^t h(t,s)m(t,s)ds = \lim_{t \downarrow 0} \int_0^t h(t,s)[m(0,0) + m_s(\alpha(t,s)t,\alpha(t,s)s)t + m_t(\alpha(t,s)t,\alpha(t,s)s)s]ds
\]

\[
= m(0,0)\lim_{t \downarrow 0} \int_0^t h(t,s)ds + \lim_{t \downarrow 0} t \int_0^t h(t,s)m_t(\alpha(t,s)t,\alpha(t,s)s)ds
\] (B2)

\[+ \lim_{t \downarrow 0} \int_0^t h(t,s)m_s(\alpha(t,s)t,\alpha(t,s)s)sds,
\]

where the second term on the right hand side of (B2) is clearly 0, and integration by parts reveals that the third term is also zero:

\[
\lim_{t \downarrow 0} \int_0^t h(t,s)m_s(\alpha(t,s)t,\alpha(t,s)s)sds
\]

\[= \lim_{t \downarrow 0} \int_0^t h(t,s)ds \int_0^t m_s(\alpha(t,t)t,\alpha(t,t)t)
\]

\[+ \lim_{t \downarrow 0} \int_0^t h(t,r)dr \int_0^t [m_s + m_t \alpha_s ts + m_s[\alpha(t,s) + \alpha_s s]]sds
\]

\[= 0 \quad \text{(B3)}
\]

**LEMMA 2:** If \( h(t,s) \) and \( g(t,s) \) are twice continuously differentiable functions of \( s \) and \( t \), \( h(0,0) = 0 \), \( h_t(0,0) > 0 \), \( h_s(0,0) \leq 0 \), and \( h_t(0,0) + h_s(0,0) > 0 \) then

\[
\lim_{t \downarrow 0} \int_0^t \frac{g(t,s)}{h(t,s)} ds = \lim_{t \downarrow 0} \int_0^t \frac{g(0,0)}{h_t(0,0)t + h_s(0,0)s} ds.
\] (B4)
More generally, if \( h(t, s) \) and \( g(t, s) \) are \( n \)-times continuously differentiable functions of \( s \) and \( t \), \( h(t, s) \) satisfies \( h_{(i)_{(i)}(j)_{(j)}}(0, 0) = 0 \) for all \( i, j \geq 0; i + j = n - 1 \), and \( g(t, s) \) satisfies \( g_{(i)_{(i)}(j)_{(j)}}(0, 0) = 0 \) for all \( i, j \geq 0; i + j \leq n - 2 \), then

\[
\lim_{t \downarrow 0} \int_0^t \frac{g(t, s)}{h(t, s)} \, ds = \lim_{t \downarrow 0} \int_0^t \frac{1}{\sum_{i=0}^{n-1} \frac{1}{i!(n-i-1)!} \sum_{j=0}^{n-i} \frac{g_{(i)_{(i)}(j)_{(j)}}(0, 0)s^i t^{n-i}}{\sum_{i=0}^{n} \frac{1}{i!(n-i)!} h_{(i)_{(i)}(j)_{(j)}}(0, 0)s^i t^{n-i}}} \, ds ,
\]

(B5)

provided that the limit exists, and where \( g_{(i)} \) and \( h_{(i)} \) denote the \( i \)-th partial derivatives of \( g \) and \( h \) with respect to \( s \), and \( g_{(i)_{(i)}(j)_{(j)}} \) and \( h_{(i)_{(i)}(j)_{(j)}} \) denote the corresponding cross-partial derivatives.

**Proof of Lemma 2:**

Taking a first order Taylor expansion of \( h(t, s) \) around the point \((0, 0)\) yields

\[
\lim_{t \downarrow 0} \int_0^t \frac{g(t, s)}{h(t, s)} \, ds = \lim_{t \downarrow 0} \int_0^t \frac{g(t, s)}{h_t(0, 0)t + h_s(0, 0)s + o^2} \, ds ,
\]

(B6)

\[
= \lim_{t \downarrow 0} \int_0^t \frac{g(t, s)}{h_t(0, 0)t + h_s(0, 0)s} \, ds
\]

\[
- \lim_{t \downarrow 0} \int_0^t \frac{g(t, s)o^2}{(h_t(0, 0)t + h_s(0, 0)s)(h_t(0, 0)t + h_s(0, 0)s + o^2)} \, ds .
\]

Since \( h_t(0, 0)t + h_s(0, 0)s > 0 \) the integrand in the second integral is bounded, so the second integral is 0. Application of Lemma 1 to the first integral yields (B4). If \( h_s(0, 0) < 0 \) then

\[
\lim_{t \downarrow 0} \int_0^t \frac{g(0, 0)}{h_t(0, 0)t + h_s(0, 0)s} \, ds = \frac{g(0, 0)}{h_s(0, 0)} \ln \left( \frac{h_t(0, 0) + h_s(0, 0)}{h_t(0, 0)} \right) ,
\]

(B7)

and if \( h_s(0, 0) = 0 \), then

\[
\lim_{t \downarrow 0} \int_0^t \frac{g(0, 0)}{h_t(0, 0)t + h_s(0, 0)s} \, ds = \frac{g(0, 0)}{h_t(0, 0)} .
\]

(B8)

The proof can be easily generalized using higher order Taylor approximations to include (B5).
**PROPOSITION 1.** Let $g(t, s) = k(t, s)^{-1}$ and suppose that $g(0, 0) = 0$ and that the function $g(t, s)$ is non-negative, twice continuously differentiable, and strictly increasing in $t$ on $[0, T] \times [0, T]$, so $k(t, s)$ is strictly positive and strictly decreasing in $t$ on $[0, T] \times [0, T]$ and is twice continuously differentiable on $(0, T] \times (0, T]$. If $g_s(0, 0) + g_r(0, 0) > 0$ and $g_s(0, 0) \leq 0$ so that $g_r(0, 0)t + g_s(0, 0)s > 0$ for all $(t, s)$, $0 < s \leq t \leq T$, and that

$$\lim_{t \to 0} \int_0^t k(t, s) ds = \lim_{t \to 0} \int_0^t \frac{1}{g_s(0, 0)s + g_r(0, 0)t} ds \tag{B9}$$

exists and is strictly positive, then the Volterra integral equation of the first kind,

$$\int_0^t k(t, s)x(s) ds = 1, \tag{B10}$$

has a unique continuos solution $x(t)$ on $[0, T]$.

**PROOF OF PROPOSITION 1.**

Any continuous solution of (B10) must satisfy

$$\lim_{t \to 0} \int_0^t k(t, s)x(s) ds = 1, \tag{B11}$$

so

$$x(0) = x_0 = \left[\lim_{t \to 0} \int_0^t k(t, s) ds \right]^{-1}, \tag{B12}$$

which is strictly positive and finite given the assumptions made on $g(t, s)$. [If $g_r(0, 0) = 0$ then $x_0 = g_s(0, 0)$ and if $g_s(0, 0) < 0$ then $x_0 = g_s(0, 0) \ln \left(g_r(0, 0) / (g_r(0, 0) + g_s(0, 0))\right)$].

Any continuous solution to (B9) must also solve the Volterra integral equation of the second kind,

$$x(t) = -\int_0^t \frac{k(t, s)}{k(t, t)} x(s) ds, \tag{B13}$$

I am extremely grateful to Gustaf Gripenberg for suggesting this proof.
which is obtained by differentiating (B10) with respect to $t$, given the initial value condition $x(0) = x_0$. However the kernel $-k_s(t,s)/k(t,t)$ does not satisfy the usual sufficient conditions for existence of a solution to (B13).

Let $y(t) = x(t) - x_0$ so that $y(t)$ satisfies

$$y(t) = \int_{0}^{t} k(t,s)x(s)ds + f(t),$$

where $k(t,s) = -k_s(t,s)/k(t,t)$ and

$$f(t) = x_0\left[\int_{0}^{t} k(t,s)ds - 1\right].$$

Note that

$$\lim_{t \to 0^+} \int_{0}^{t} k(t,s)ds = -\lim_{t \to 0^+} \int_{0}^{t} \frac{k_s(t,s)}{k(t,t)} ds$$

$$= \lim_{t \to 0^+} \int_{0}^{t} \frac{g_s(t,s)g(t,t)}{g(t,s)^2} ds$$

$$= g_s(0,0)\lim_{t \to 0^+} \int_{0}^{t} \frac{g(t,t)}{g(t,s)^2} ds$$

$$= g_s(0,0)\lim_{t \to 0^+} \int_{0}^{t} \frac{(g_s(0,0) + g_s(0,0)t}{(g_s(0,0)t + g_s(0,0)s)^2} ds$$

so if $g_s(0,0) = 0$ then

$$\lim_{t \to 0^+} \int_{0}^{t} k(t,s)ds = \lim_{t \to 0^+} g_s(0,0)\int_{0}^{t} \frac{1}{g_s(0,0)t} ds = 1$$

and if $g_s(0,0) < 0$ then

$$\lim_{t \to 0^+} \int_{0}^{t} k(t,s)ds = g_s(0,0)(g_s(0,0) + g_s(0,0))\lim_{t \to 0^+} \int_{0}^{t} \frac{t}{(g_s(0,0)t + g_s(0,0)s)^2} ds$$

$$= g_s(0,0)(g_s(0,0) + g_s(0,0))\left[\frac{1}{g_s(0,0)} - \frac{1}{g_s(0,0)}\right] = 1$$

so $\lim_{t \to 0^+} f(t) = 0$. 

The function \( f(t)/t \) is bounded since

\[
\lim_{t \to 0} \frac{f(t)}{t} = \lim_{t \to 0} \left[ \int_0^t \frac{g(t,s)g(t,t)}{g(t,s)^2} ds - \frac{1}{t} \right] = \lim_{t \to 0} \int_0^t \frac{g(t,s)(g(t,t) - g(t,s)^2)}{g(t,s)^2} ds
\]

\[
= \lim_{t \to 0} \int_0^t \frac{(g_s - g_{ss}^2)5^2 - 2g_s s - g_{ss} s^2}{(g_t + g_s)^2 t^2} ds + \int_0^t \frac{h_{ut} t^3 + h_{ut} t^2 s + h_{ut} t^2 s^2 + h_{ss} s^3}{(g_t + g_s)^2 t^2} ds
\]

\[
= \lim_{t \to 0} \int_0^t \frac{g_t (g_s + g_{ss})^2 t^2 - 1}{t} ds + \int_0^t \frac{h_{ut} t^3 + h_{ut} t^2 s + h_{ut} t^2 s^2 + h_{ss} s^3}{(g_t + g_s)^2 t^2} ds
\]

\[
= \lim_{t \to 0} \int_0^t \frac{h_{ut} t^3 + h_{ut} t^2 s + h_{ut} t^2 s^2 + h_{ss} s^3}{(g_t + g_s)^2 t^2} ds
\]

where \( g_t = g_t(0,0) \), \( g_s = g_s(0,0) \), \( h(t,s) = g_t(t,s)g(t,t) \), and \( h_{ss} = h_{ss}(0,0) \). The calculations in Eq. (B19) differ from the more direct applications of Lemma 2 since the second order terms in the third order Taylor expansion of the numerator are not zero, but integrate to zero as shown.

The solution to (B14) is formally

\[
y(t) = f(t) + \int_0^t r(t,s) f(s) ds, \quad \text{(B20)}
\]

where

\[
r(t,s) = k1(t,s) + \int_0^t r(t,u) k1(u,s) du. \quad \text{(B21)}
\]

If we let \( q(t,s) = s r(t,s) \) then (B20) and (B21) can be written as

\[
y(t) = f(t) + \int_0^t q(t,s) \frac{f(s)}{s} ds, \quad \text{(B22)}
\]
and
\[ q(t, s) = k3(t, s) + \int_{0}^{t} q(t, u)k2(u, s)\,du. \]  \hspace{1cm} (B23)

where \( k2(t, s) = (s/t)k1(t, s) \) and \( k3(t, s) = s\,k1(t, s) \).

Using this transformation we have that \( k3(t, s) \) is continuously differentiable \((\lim_{t \to 0} k3(t, s) = 0)\) and that \( k2(t, s) \) is integrable and satisfies
\[
\lim_{t \to 0} \int_{0}^{t} k2(t, s)\,ds = \lim_{t \to 0} \int_{0}^{t} \frac{g_t(t, s)g(t, t)s}{g(t, s)^2}\,ds \\
= g_t \lim_{t \to 0} \int_{0}^{t} \frac{g(t, t)s}{g(t, s)^2}\,ds \\
= g_t \lim_{t \to 0} \int_{0}^{t} \frac{(g_t + g_s)st}{(g_t + g_s)^2}\,ds \\
\]
so if \( g_s(0, 0) = 0 \) then
\[
\lim_{t \to 0} \int_{0}^{t} k2(t, s)\,ds = g_t \lim_{t \to 0} \int_{0}^{t} \frac{s}{g_t^2}\,ds \\
= \lim_{t \to 0} \frac{1}{2} \frac{t^2}{t^2} = \frac{1}{2} \\
\]
and if \( g_s(0, 0) < 0 \) then
\[
\lim_{t \to 0} \int_{0}^{t} k2(t, s)\,ds = \frac{g_t(g_t + g_s)}{g_s} \lim_{t \to 0} \int_{0}^{t} \frac{1}{(g_t + g_s)} - \frac{g_t}{(g_t + g_s)^2}\,ds \\
= \frac{g_t(g_t + g_s)}{g_s} \left[ \ln \left( \frac{g_t + g_s}{g_t} \right) - \frac{g_t g_s}{g_t(g_t + g_s)} \right] \\
= - \frac{g_t}{g_s} + \frac{g_t(g_t + g_s)}{g_s} \ln \left( \frac{g_t + g_s}{g_t} \right) < 1 \\
\]
Since in both cases the limit is less than one, equation (B23) defines a mapping that has a fixed point. So the function \( q(t, s) \) exists (see Gripenberg, Londen and Staffans, 1990, Corollary
9.3.14 and Theorem 12.2.1). Since the function \( q(t, s) \) exists, a solution \( y(t) \) exists given by Eq. (B22).

Now suppose that there are two (or more) continuous solutions to Eq. (B10) and let \( z(t) \) be the difference between them. Any continuous solution to Eq. (B10) must satisfy

\[
x(0) = \left[ \lim_{t \searrow 0} \int_{0}^{t} k(t, s) ds \right]^{-1},
\]

so \( z(0) = 0 \) and

\[
\lim_{t \searrow 0} \int_{0}^{t} k(t, s)z(s) ds = \lim_{t \searrow 0} \int_{0}^{t} k(t, s)|z(s)| ds = 0.
\]

Differentiating Eq. (B10) gives

\[
z(t)k(t, t) + \int_{0}^{t} k_{i}(t, s)z(s) ds = 0,
\]

for all \( t \in (0, T] \), and multiplying the above by \( \text{sign}[z(t)] \) gives

\[
|z(t)|k(t, t) + \int_{0}^{t} k_{i}(t, s)|z(s)| ds \leq 0,
\]

for all \( t \in (0, T] \) since \( k_{i}(t, s) \leq 0 \). Finally, integrating again yields

\[
\int_{0}^{t} k(t, s)|z(s)| ds \leq \lim_{t \searrow 0} \int_{0}^{t} k(t, s)|z(s)| ds = 0.
\]

Since \( k(t, s) > 0 \), it must be true that \( z(t) = 0 \) and hence that the solution to Eq. (B10) is unique. Thus both existence and uniqueness is proved.
REFERENCES


