Deep Discounts and Comparison Shopping

S. Nuray Akin* and Brennan C. Platt†‡

Abstract

We analyze a continuous-time sequential search model where buyers face a deadline, after which search becomes more costly. A seller cannot observe a potential buyer's remaining time until deadline when quoting a price. We compare settings where quotes are lost if not used immediately, versus quotes that can be recalled at any time.

Either setting produces a unique equilibrium, taking one of two forms. In a late equilibrium, all prices induce comparison shopping, meaning some buyers pass these quotes in hope of a better price. An early equilibrium also includes a deep discount, meaning a price that is accepted immediately by any buyer.

We find that recall is not always beneficial. When a frictionless market would serve buyers upon entry, a market without recall can provide more consumer surplus and total welfare. When a frictionless market would serve buyers only at their deadline, total welfare is higher under perfect recall.

Keywords: Equilibrium search, price formation, deadline, recall, discounts

JEL Classification: D40, D83

*Department of Economics, Ozyegin University, Cekmekoy, Istanbul, Turkey (216) 594-0000, nuray.akin@ozyegin.edu.tr

†Department of Economics, Brigham Young University, 2111 WVB, Provo, UT 84602, (801) 422-8904, brennan_platt@byu.edu

‡We thank Alli Wilkinson for her excellent support as a research assistant. We are grateful for comments from many participants and discussants at the 2022 Consumer Search and Switching Costs Workshop at NYU Stern, 2022 International Industrial Organization Conference in Boston, Econometric Society European Meetings in Milan, and Econometric Society Asian Meetings in Beijing. Akin thanks the University of Iowa Department of Economics for their hospitality during her sabbatical visit and acknowledges financial support from the Technological and Scientific Research Council of Türkiye under grant 2219. All remaining errors are our own.

1 Introduction

Buyers often face a time horizon when making a purchase, needing to buy the product by a specific time to avoid extra costs. These buyer deadlines and their impact on search behavior have been documented in a growing empirical literature. Lemieux and Peterson (2011) find that shoppers for rental truck reservations become less sensitive to price and more likely to purchase as their deadline approaches. Akin, *et al* (2013) show that real estate investment trusts, which face regulatory deadlines to deploy new capital within a year, consequently end up paying more than individual buyers of identical properties. Coey, *et al* (2020) document buyer deadlines across a wide variety of consumer products, both in self-reported survey data from online shoppers and by following each bidder across multiple auctions of the same item, finding that their bids increased after each failure to win.

These scenarios are non-trivial because finding the best option for a truck, property, or product requires search. If the right item at the lowest price were perfectly known, one could acquire it just before the deadline without worry. More likely, though, one will only come across an acceptable item infrequently. Moreover, that item could be offered at a variety of prices; indeed, the highest prices might only become acceptable as the deadline looms near.

We consider equilibrium price formation in a setting with deadlines. A continuous flow of individuals seek to buy one unit of a homogenous good. These buyers enter the market with a *grace period* of length T during which they enjoy a flow of high utility; after passing the deadline, however, their flow of utility drops until the purchase is completed. Upon meeting, the buyer learns the seller's quoted price and must decide whether to make the purchase, or continue searching. With recall, the buyer can return anytime to exercise the quoted price; without recall, the quote is lost if not exercised immediately. We consider both settings.

Sellers are aware that buyers face a deadline, but are unsure how close any particular buyer is to that deadline. Thus, they engage in probabilistic price discrimination: quoting a higher price will generate greater profit if accepted, but will also limit the pool of buyers who would accept the offer. In equilibrium, these two effects may exactly cancel each other, allowing identical sellers to ask different prices and yet have equal expected profit.[1]

We find that under mild conditions, the equilibrium is unique, falling into one of two broad categories. In a late equilibrium, sellers only offer prices that will be accepted by

[1]This balanced tradeoff between markup and volume of sales is common to many models of price posting that generate dispersed prices, such as Burdett and Judd (1983), Rob (1985), Diamond (1987), etc. The unique approach of our model is that the impending deadline endogenously determines the volume of buyers willing to accept a given price.

buyers later in their search spell. We refer to this as *comparison shopping*, because early buyers turn down some quotes in order to search for better prices. In an early equilibrium, some sellers offer a *deep discount* — a price low enough that any buyer offered it is willing to skip further search. An early equilibrium can simultaneously include other prices that induce comparison shopping.

Second, our model generates rich price dispersion. In an early equilibrium, a mass of firms will offer the deep-discount price. Prices in the comparison-shopping range are higher and are continuously distributed. When quotes cannot be recalled, there can be two more atoms in the distribution. One is at the highest price, which will only be accepted by the unlucky buyers whose grace period has expired. Another is at lowest price in the comparison-shopping range — the best price besides the deep discount, though it is still turned down by newly-entering buyers.

Our equilibrium results rule out a variety of price distributions that might seem plausible at first. For instance, there are never multiple prices offered as deep discounts, even when a mass of sellers competes for early buyers. Moreover, prices offered in the comparison-shopping range have a connected support, though the deep discount is strictly below this support. An extreme version of this generates only two offered prices, the deep discount accepted by anyone and a single high price accepted only by those who pass their deadline; we refer to this as an early bimodal equilibrium.

Third, price dynamics (analyzed through comparative statics) are largely determined by two factors: deadline concentration and urgency to buy. Concentration refers to the portion of the steady-state population of buyers that are near their deadline. Greater concentration encourages sellers to target these desperate buyers more heavily. Urgency refers to the utility drop after the deadline. A larger drop causes a steeper increase in reservation prices, as buyers are willing to pay more to avoid this painful utility reduction.

The interaction of these factors leads to surprising consequences in equilibrium. In a setting without recall, a larger penalty or shorter deadline pushes firms to target buyers *earlier* in their search, eventually with all firms offering the deep discount. Buyers have greater urgency at every point in their search, but less concentration at the deadline, making it more profitable to offer a price that is immediately accepted. In perfect recall, however, the deep discount is unaffected by the deadline penalty or grace period. Thus, when urgency increases, comparison-shopping offers become more profitable while deep-discount profits are unchanged, eventually leading sellers to a late equilibrium with no deep discounts offered.

Fourth, we find that recall can be harmful to buyers and helpful to sellers. This occurs when sellers only offer the deep discount in the non-recall setting, but offer dispersed prices in the recall setting, which yields higher prices. When grace-period benefits are low, a frictionless market would have buyers consume immediately on entering the market. This makes comparison shopping induced by recall inefficient because it delays consumption; a non-recall world avoids this by only offering deep discounts. When grace-period benefits are high, a frictionless market would have buyers enjoy their full grace period before consuming at the deadline. In that scenario, the recall setting generates only comparison shopping, getting closer to the first-best than a non-recall setting, where some buyers consume earlier than is socially optimal and others consume well after their deadline.

The technical challenge of this model in either setting is the non-stationarity of the search process. Buyers continuously revise their reservation price throughout the grace period, and with recall, the best quote received at each point in the search spell also becomes a state variable. Furthermore, the distributions of buyer types and of offered prices affect and are affected by the buyer strategies in equilibrium. We overcome this by translating equilibrium conditions into a set of differential equations that yield a unique analytic solution in both versions of our model. van den Berg (1990) implemented this approach for unemployment search with an exogenous wage distribution.[2] Its first application with an endogenous wage distribution appears in Akin and Platt (2012), which only characterized late equilibria and without recall. There, unemployed workers receive unemployment insurance benefits for a finite duration, after which they are cut off. This creates ex-post differences among the workers' reservation wages, which allows firms to offer these otherwise identical workers different wages.

We proceed with a review of related work. Section 2 then presents the model without recall and defines equilibrium. In Section 3, we walk through the process of translating the equilibrium conditions, and present the equilibrium solution. Section 4 presents and solves the model with recall. Section 5 compares welfare results in both settings. We summarize potential applications of our framework and conclude in Section 6.

1.1 Related Literature

Stigler (1961) initiated the formal modeling of consumer search behavior, taking as given the distribution of prices offered by sellers. The goal of equilibrium search theory has been to complete the model, so that sellers also behave optimally, given the search strategy of buyers.

[2] A large literature has explored equilibrium search in labor markets with the intent to study wage formation and the effects of unemployment insurance. This is well surveyed in Rogerson, *et al* (2005).

Diamond (1971) highlights the difficulty in sustaining more than one price in equilibrium. Hence, successful price dispersion models typically rely on heterogeneity or uncertainty in search costs (e.g. Salop and Stiglitz, 1976; Butters, 1977; Wilde and Schwartz, 1979; Rob, 1985; Janssen and Moraga-Gonzalez, 2004), production costs (e.g Reinganum, 1979; Janssen, *et al*, 2011), quantities demanded (e.g Reinganum, 1979), or valuations (e.g Diamond, 1987; Choi, *et al*, 2018).

In our model, all buyers are ex-ante identical in their valuation of and unit-demand for the good and their cost of search. Of course, heterogeneity is necessary to justify sellers offering different prices, but, here, this arises ex-post as some buyers experience longer search spells than others. Luck plays a role in creating these different experiences, as a buyer may not encounter an opportunity to buy; but choices also contribute, as sellers may ignore segments of the market or buyers may turn down high prices early in search. The model in Coey, *et al* (2020) also generates price dispersion from buyer deadlines, but applied to an auction environment. There, sellers passively offer auctions rather than choosing a price to post, and buyers submit bids based on their remaining search time, but only the bidder with the least time remaining wins.

Our non-recall model can sustain a continuum of prices, which contrasts with Curtis and Wright (2004) where (in a monetary search setting) price posting generically results in no more than two prices that maximize profits, despite many types of buyers. Our setting overcomes this limited price dispersion because un-targeted buyer types will build up in steady state, eventually making a range of them equally profitable to target in the dispersed equilibria. A similar build up cannot happen in Curtis and Wright (2004) because buyers differ only in their idiosyncratic match values, so the proportions of each type will equal the exogenous probabilities of that type, regardless of who is targeted.

We also can get a positive mass of sellers at the deep discount and at the lowest comparison shopping price, which is surprising given there are a continuum of ex-post buyer types at those points.[3] To our knowledge, this does not occur elsewhere in the equilibrium search literature, yet it may help explain jumps in the hazard rate of accepting an offer over the search spell. An extreme version of these atoms occurs in our bimodal equilibrium, where the lowest and highest price are offered without targeting any of the buyers in between. This is reminiscent of Salop and Stiglitz (1976), except it endogenously arises with an ex-post continuum of buyer types, rather than only two types ex-ante differing in search cost.

Our recall model shares much in common with simultaneous search models (e.g Burdett

[3] There is a mass of identical buyers past the deadline, which is the third possible atom of seller offers.

and Judd, 1983; Stahl, 1989; Janssen and Moraga-Gonzalez, 2004). These each generate a continuum of prices (as in our comparison-shopping range) as sellers balance a higher markup against a lower probability of being the best price among received quotes. This mechanism is also embedded in our perfect recall setting, although the number of quotes received is random over the span of the grace period. However, our model also introduces the deep discount — a price sufficiently low that buyers are willing to skip all other quotes. This does not occur in Stahl (1989) or similar models because shoppers have no time discounting or other cost of search; even if they get the best price, they might as well receive all possible quotes. In this sense, our recall model embeds a flavor of simultaneous search that is produced in a sequential search framework.

A similar deep discount paired with a continuum of comparison-shopping prices occurs in the three-period model of Akin and Platt (2014), where consumers can partly recall past prices. We generalize these results here into continuous time and add discounting; this facilitates the comparison to the no-recall sequential search setting. Buyers can recall past prices for an extra cost in Janssen and Parakhonyak (2014) and Armstrong and Zhou (2016), where the cost is an exogenous friction in the former and an endogenous part of the price to the seller in the latter. Neither of these operate in the same way as our deep discount, though. Costly revisits in Janssen and Parakhonyak (2014) result in the same continuous price distribution as in Stahl (1989). Exploding offers in Armstrong and Zhou (2016) are not accepted by all buyers, but are used to segment those with a low outside option. Choi, *et al* (2022) also has recall in sequential search for exogenous distributions of prices and match values; some matches are good enough to immediately accept, while others are only exercised eventually if no better options arise. Each match is idiosycratic, so again no match is accepted by all buyers as with our deep discount.

Several of these papers explore issues relevant to our results on the welfare effect of recall. We find here that when recall generates an early dispersed equilibrium, buyers and total welfare can be worse off than if no recall was possible because buyers take longer to make their purchase. Similarly, greater certainty of recall can harm buyers in Akin and Platt (2014) during the analogous full equilibrium. In contrast, full recall is welfare improving in Armstrong and Zhou (2016) relative to exploding offers or other attempts to restrict recall, which inefficiently reduce search and produce lower quality matches.

Welfare is also affected by the number of quote opportunities given to a consumer. Janssen and Moraga-Gonzalez (2004) show that in a low-intensity equilibrium (in which some searchers do not participate), having more firms can harm consumers as it discourages participation. Since we have a continuum of firms, the limit on quotes is the grace period length. With perfect recall, additional time is always beneficial as it induces more competition in prices and reduces the chance of hitting the deadline penalties. Without recall, though, more time is harmful in a bimodal equilibrium, as sellers focus more on buyers who have passed their deadline which, ironically, is hit by more buyers.

2 Model without Recall

Consider a continuous time environment, with infinitesimal buyers and sellers each entering the market at rate δ. All agents discount future utility at rate ρ and demand one unit of the homogenous good being sold. This good provides value x to any buyer, but because sellers may ask different prices for the good, buyers may find it worthwhile to search. Buyers encounter a seller at Poisson rate μ. Upon encounter, the buyer draws an asking price p from the distribution of offered prices, $F(p)$. The buyer can either make the purchase, obtaining $x - p$ surplus and exiting the market, or continue searching with no recall of past offers.

2.1 The Buyer's Search Problem

The buyer has T units of time to search without penalty, which we refer to as the *grace period*. During this time, she receives utility b each instant. After the grace period expires, the instantaneous utility falls to $d < b$ until a purchase is made.[4] We assume throughout that $\mu > \rho e^{\frac{\rho T}{2}}$. This ensures that buyers encounter a seller with sufficient frequency that continued search is a viable option.

We characterize this search problem using the remaining time until the grace period expires, z, as the state. At each state z, the buyer chooses a reservation price $R(z)$.

For instance, once the grace period expires ($z = 0$), the buyer's problem can be recursively formulated as follows:

$$\rho V(0) = \max_{R(0)} \ d + \mu \int_{-\infty}^{R(0)} (x - p - V(0))\, dF(p). \tag{1}$$

[4] If a buyer needs new housing before a job starts in a new city, b would reflect the buyer's consumer surplus in her current housing, while d reflects the lower surplus of switching to short-term housing or long-distance commuting. In search for a gift for a special occasion, b would be the stream of utility from a relationship, while d reflects the lower utility when expectations of a gift are not met. In labor search, b would be the flow of unemployment benefits while d would be zero or the cost of self-financing after exhausting unemployment benefits.

Here, $V(0)$ represents the expected net present utility of a buyer searching after the grace period. Each instant, she receives utility d. She encounters purchase opportunities at rate μ, and will accept any price at or below $R(0)$. If a transaction occurs at price p, her utility changes from $V(0)$ to $x - p$.

During the grace period, the recursive problem takes the following form:

$$\rho V(z) = \max_{R(z)} \; b + \mu \int_{-\infty}^{R(z)} (x - p - V(z))\, dF(p) - V'(z). \tag{2}$$

Note three changes in this Bellman equation, compared to decisions after the grace period. First, the instantaneous utility is b. Second, a buyer with grace time remaining holds out for a lower price $R(z)$. Finally, the state variable z deterministically falls as the grace period ticks down, which is reflected in the term $-V'(z)$.

By defining the Bellman equation in this way, we are assuming that both $V(z)$ and $V'(z)$ are continuous and differentiable; thus we do not examine possible equilibria with discontinuous value functions. Even though the instantaneous utility abruptly falls once $z = 0$, the present expected cost of these penalties grows smoothly as expiration approaches.

In formulating a reservation price at each instant, the buyer should make a purchase as long as it weakly increases her utility. Thus, for all $z \in [0, T]$:

$$R(z) = x - V(z). \tag{3}$$

2.2 Steady State Conditions

For sellers to choose a pricing strategy, it will be critical to know how many buyers there are at each state of the search process. We consider a steady state equilibrium, where the measure of buyers in each state stays constant over time. Let $H(z)$ denote the measure of buyers with z or less time remaining in their grace period. Note that $H(0)$ includes all whose grace period has expired. Then $H'(z)$ indicates the relative density of buyers in state z.

Buyers enter the market at rate δ; thus,

$$H'(T) = \delta. \tag{4}$$

At state $z > 0$ in the grace period, buyers exit the market only when they find an acceptable price, which happens at rate $\mu F(R(z))$. Thus, the density of buyers at z must fall at that rate:

$$H''(z) = \mu F(R(z))H'(z). \tag{5}$$

Finally, among those who have exceeded the grace period ($z = 0$), all prices offered in equilibrium are acceptable. Thus, they exit whenever they encounter a seller, *i.e.* at rate μ. At the same time, this population of expired buyers is replenished by the flow of buyers whose grace period has just expired, $H'(0)$:

$$H'(0) = \mu H(0). \tag{6}$$

2.3 The Seller's Problem

Sellers produce their good at cost $c < x$, at the time of the transaction. They are unable to observe the state of the buyer with whom they have been paired. Thus, asking a higher price bears the risk of a lower likelihood of being accepted. At the same time, it would result in higher realized profits if accepted. If a seller offers price $R(z)$, we say they are *targeting* buyers with z time until expiration, though all buyers with less time will also accept. Thus, seller expected profit from targeting type z is represented as follows:

$$\pi(z) = \frac{H(z)}{H(T)}(R(z) - c). \tag{7}$$

Since the measure of buyers and Bellman equation (and hence reservation values) are continuously differentiable, the expected profit function is also continuously differentiable.

If multiple prices produce the same maximal expected profit, sellers can randomize over these prices, which would be represented in the cumulative price distribution $F(p)$. One can interpret this as each seller using the same mixed strategy, randomizing anew for each potential buyer. Alternatively, each seller could stick with a particular price, with $F(p)$ representing the aggregate distribution of sellers' choices. Since there is no repeated interaction between any given buyer or seller, either interpretation is equally valid.

2.4 Equilibrium Definition

A steady state search equilibrium consists of seller profit π, a reservation price function $R(z)$, the measure of buyers $H(z)$, and the distribution of sellers' offered prices $F(p)$, such that:

1. $R(z)$ maximizes the utility of a buyer with z time until the expiration of the grace

period, given $F(p)$.

2. All prices in the support of F produce the same maximal profit π, while all other prices produce no more than π.

3. $H(z)$ satisfies the steady state conditions in Eqs. 4 through 6.

3 Equilibrium Characterization without Recall

Equilibria can be categorized by two features. First, the price distribution can be *degenerate*, where all sellers offer the same price, or *dispersed*, where a variety of prices are offered. If exactly two prices are offered, we refer to this dispersion as *bimodal*.

Second, the equilibrium pricing may be focused on *late* or *early* buyers. In a *late equilibrium*, sellers target buyers who are late in their search process (close to expiration). In particular, there is a critical state $Z^* \in [0, T]$ such that $R(Z^*)$ is the lowest price offered, while $R(0)$ is the highest price offered. Buyers with $z > Z^*$ have too low a reservation price, making it unprofitable for sellers to target them; these buyers reject all offers until at least time Z^*. Even buyers with $z < Z^*$ will reject some offers, preferring to continue searching; thus, all prices from $R(Z^*)$ to $R(0)$ are subject to *comparison shopping*.

An *early equilibrium* adds in an atom α^* at $R(T)$, indicating that a fraction of firms are targeting the buyers who have just entered the market with the lowest willingness to pay. We refer to $R(T)$ as a *deep discount* — a price which every buyer will immediately accept. This deep discount is almost always strictly less than $R(Z^*)$, leaving a gap in the support of the price offer distribution. For $z \in (Z^*, T)$, the relative density of buyers $H(z)$ is initially falling faster than the reservation price rises, making them unprofitable to target.

Below, we first present the possible equilibria, then demonstrate that these are the only equilibria that can occur (in Proposition 1), and show that only one will occur for a given set of parameters (in Proposition 2).

3.1 Calculation of Equilibrium

We first sketch the process to reach the equilibrium solution, which is formalized in Proposition 1. We start by translating the equilibrium conditions into functions of the reservation prices $R(z)$, using Eq. 3 to substitute for the value function and its derivatives. For brevity, here we only report the translated equilibrium conditions used to solve for reservation prices and the price distribution; the steady state conditions on population and profit follow from reservation prices, and are relegated to the Technical Appendix, along with the algebraic manipulations used to obtain the following solutions. The equilibrium conditions are equivalent to the following system of equations. When $Z > 0$, for $z \in [0, Z)$, we have:

$$
\begin{aligned}
R''(z) &= -(\rho + \mu F(R(z)))R'(z) && (8)\\
R'(0) &= d - b && (9)\\
F(R(z)) &= \frac{R''(z)}{\mu R'(z)} - \frac{2R'(z)}{R(z) - c} && (10)\\
R'(0) &= -\mu(R(0) - c). && (11)
\end{aligned}
$$

Eqs. 8 and 10 are derived from buyer optimization (Eq. 2) and equal profits (Eq. 7), respectively, for interior values of z. Eqs. 9 and 11 ensure that buyer optimization and profits at $z = 0$ are continuous (*i.e.* $V(\epsilon)$ approaches $V(0)$ and $\Pi(\epsilon) = \Pi(0)$).

When $Z = 0$, there is no interior of the comparison shopping range; however, the post-expiration Bellman equation (1) can be translated and solved directly as:

$$R(0) = x - \frac{\alpha\mu(x - R(T)) + d}{\rho + \alpha\mu}. \tag{12}$$

Whether $Z = 0$ or $Z > 0$, in the range of $z \in [Z, T]$, buyer optimization translates to:

$$R'(z) = \rho(x - R(z)) + \alpha\mu(R(T) - R(z)) - b. \tag{13}$$

This system of equations solves as follows. In the late region $z \in [0, Z^*]$, we substitute for $F(R(z))$ from Eq. 10 into Eq. 8 and get a second-order differential equation of the reservation prices:

$$\rho R'(z) + 2R''(z) + \frac{2R'(z)^2}{c - R(z)} = 0. \tag{14}$$

This differential equation has a unique solution, up to two constants which are pinned down using Eqs. 9 and 11 as boundary conditions at $z = 0$. This provides the equilibrium reservation prices.

In the early region $z \in (Z^*, T]$, the first-order differential Eq. 13 yields a unique solution as well, up to one constant determined by the boundary condition that $R(z)$ is continuous at Z^*. Indeed, the same Eq. 13 pins down the critical state Z^* by requiring that $R'(z)$ is also continuous at Z^* (relative to the Eq. 14 solution).

We then use these reservation price solutions to compute the distribution of buyers and expected profits. Finally, the atom α^* is pinned down by comparing the profits from targeting buyers with $R(Z^*)$ versus $R(T)$; if these can be equated for an $\alpha \in (0,1)$, it generates an early dispersed or bimodal equilibrium.

First, we report the solution for reservation prices. The reservation price at the deadline is always the highest, and its solution depends on whether there is a continuous portion of the distribution ($Z^* > 0$) or not:

$$R(0) = \begin{cases} x - \frac{b}{\rho} + \frac{b-d}{\rho} \cdot \frac{\rho + \alpha^* \mu e^{-(\alpha^* \mu + \rho)T}}{\rho + \alpha^* \mu} & \text{if } Z^* = 0 \\ c + \frac{b-d}{\mu} & \text{if } Z^* > 0. \end{cases} \tag{15}$$

Leading up to the deadline, reservation prices differ depending on whether they fall in the continuously dispersed range $z < Z^*$ or not.

$$R(z) = \begin{cases} c + \frac{b-d}{\mu} e^{-\frac{2\mu}{\rho}\left(1 - e^{-\frac{\rho z}{2}}\right)} & \text{if } z \in (0, Z^*] \\ x - \frac{b}{\rho} - \frac{\alpha^* \mu + \rho e^{(\alpha^* \mu + \rho)(T-z)}}{\alpha^* \mu + \rho e^{(\alpha^* \mu + \rho)(T - Z^*)}} \left(x - \frac{b}{\rho} - R(Z^*) \right) & \text{if } z \in (Z^*, T]. \end{cases} \tag{16}$$

In equilibrium, buyers will only encounter comparison-shopping prices between $R(Z^*)$ and $R(0)$ (and possibly the deep discount $R(T)$). Yet we can still compute what buyers would be *willing* to pay at any point in their search process. Sellers can then consider (but reject, in equilibrium) the option of making offers to these untargeted buyers.

Two properties of reservation prices in the comparison-shopping range are worth emphasizing. First, these reservation prices always cover the cost of production. Any additional willingness to pay comes from the impending change in utility at expiration, $b - d$, which is moderated by the frequency of price offers μ and exacerbated by the impatience of buyers ρ. Note that the objective value x to the buyer does not matter in this range. Effectively, the search friction allows sellers to extract surplus based on the idiosyncratic time until the deadline penalty, but not based on the commonly-shared underlying value x, which would be readily undercut in Bertrand-like competition among sellers.

Second, $R'(z) < 0$ and $R''(z) > 0$. That is, buyers are willing to accept higher prices (and the increase become more pronounced) as their deadline approaches.[5] This acceleration (or

[5]Discounting is one essential ingredient for this result. The present value of the penalty, $(b-d)e^{-\rho z}$ is bigger and grows faster as the deadline approaches. At the same time, this is more than just a mechanical effect of discounting, because the reservation price solution also anticipates the buyer's own willingness to accept more offers as time runs out.

curvature) in reservation prices is directly proportional to the change in utility at expiration.

Next, we report the solution for the equilibrium distribution of seller asking prices:

$$F(p) = \begin{cases} 0 & \text{if } p < R(T) \\ \alpha^* & \text{if } R(T) < p < R(Z^*) \\ 1 - \frac{\rho}{2\mu}\left(1 - \ln \frac{\mu(p-c)}{b-d}\right) & \text{if } R(Z^*) < p < R(0) \\ 1 & \text{if } p \geq R(0). \end{cases} \tag{17}$$

This solution can also be reframed as the distribution of offers targeting a specific type of buyer, $F(R(z))$. This distribution generates positive masses targeting up to three different types: a mass of α^* at the deep discount $R(T)$, a mass of $e^{-\frac{\rho Z^*}{2}} - \frac{\rho}{2\mu} - \alpha^*$ at the lowest comparison-shopping price $R(Z^*)$, and a mass of $\frac{\rho}{2\mu}$ at the highest price $R(0)$. In the continuous, comparison-shopping range of the distribution (the third case of Eq. 17), we see that $F''(p) < 0$, meaning that higher prices are relatively less frequent.

The population of buyers in equilibrium is:

$$H(z) = \begin{cases} H(Z^*)e^{-\frac{2\mu}{\rho}\left(e^{-\frac{\rho z}{2}} - e^{-\frac{\rho Z^*}{2}}\right)} & \text{if } 0 \leq z < Z^* \\ H(Z^*) + \delta(z - Z^*) & \text{if } Z^* \leq z \leq T \text{ and } \alpha^* = 0 \\ H(Z^*) + \frac{\delta}{\alpha^*\mu}\left(e^{\alpha^*\mu(z-T)} - e^{\alpha^*\mu(Z^*-T)}\right) & \text{if } Z^* \leq z \leq T \text{ and } \alpha^* > 0, \end{cases} \tag{18}$$

where

$$H(Z^*) = \frac{\delta}{\mu} e^{-\alpha^*\mu(T-Z^*) + \frac{\rho Z^*}{2}}. \tag{19}$$

This solution accounts for the rate at which buyers make make successful purchases, $\mu F(R(z))$, across all possible types.

The only remaining equilibrium variables are the critical time Z^* and the atom α^*. The solution for Z^* ensures that the reservation price solution in Eq. 16 (and hence, the buyer's Bellman equations) is continuous at $z = Z^*$. Effectively, this requires that newly-entering buyers correctly anticipate the expected benefit of search, though this is an equilibrium condition, not a choice by buyers. For this, we define:

$$\zeta(Z, \alpha) \equiv (x - c)\rho - b + (b - d)\left(\frac{\alpha\mu\, e^{(\alpha\mu+\rho)(Z-T)} + \rho}{\alpha\mu + \rho} e^{-\frac{\rho Z}{2}} - \frac{\rho}{\mu}\right) e^{-\frac{2\mu}{\rho}\left(1 - e^{-\frac{\rho Z}{2}}\right)}. \tag{20}$$

This equation compares the net surplus of selling the good, $(x-c)\rho - b$, to the expected net present loss from exceeding the deadline, $b-d$. In doing so, it accounts for the expected search duration, including that the deadline may not be reached at all. This is computed while anticipating the changing rate at which offers will be acceptable. Buyers have correct expectations when $\zeta(Z^*, \alpha^*) = 0$.

The size of the atom α^* is determined by the relative profitability of offering $R(T)$ compared to $R(Z)$, which is:

$$\phi(Z,\alpha) \equiv \frac{R(Z)-c}{\rho(x-c)-b}\left(\frac{(\alpha\mu+\rho)\left(e^{\alpha\mu(T-Z)}-1\right)}{e^{(T-Z)(\alpha\mu+\rho)}-1} - \alpha\rho e^{\frac{\rho Z}{2}}\right) + e^{\alpha\mu(T-Z)} + \alpha e^{\frac{\rho Z}{2}} - 1. \quad (21)$$

Thus, when $\phi(Z^*, \alpha^*) = 0$, these prices are equally profitable, but when $\phi(0,\alpha) > 0$ for all $\alpha \in [0,1]$, only $R(T)$ will be offered.

3.2 Equilibrium Properties

We now show that this proposed solution uniquely characterizes the equilibrium.

Proposition 1. *Assuming $\mu > \rho e^{\frac{\rho T}{2}}$, a solution $R(z)$, $F(p)$, $H(z)$, and $G(z)$ is an equilibrium if and only if it satisfies Eqs. 15 through 19 with one of the following cases:*

- *(late degenerate) $Z^* = \alpha^* = 0$ if $\zeta(0,0) \leq 0$*
- *(late dispersed) $Z^* \in (0,T]$ and $\alpha^* = 0$ if $\zeta(Z^*,0) = 0$*
- *(early dispersed) $Z^* \in (0,T]$ and $\alpha^* \in (0,1)$ if $\zeta(Z^*,\alpha^*) = 0$ and $\phi(Z^*,\alpha^*) = 0$*
- *(early bimodal) $Z^* = 0$ and $\alpha^* \in (0,1)$ if $\zeta(0,\alpha^*) \geq 0$ and $\phi(0,\alpha^*) = 0$*
- *(early degenerate) $Z^* = 0$ and $\alpha^* = 1$ if $\phi(0,1) \geq 0$.*

Which equilibrium type occurs depends on parameter values, of course, as we discuss in the next subsection. It is worth noting each equilibrium type coincides with the neighboring row in the limit. For instance, if a late dispersed equilibrium generates a Z^* approaching 0, the solution coincides with a late degenerate equilibrium. If instead it generates a Z^* approaching T, the equilibrium solution coincides with an early dispersed equilibrium.

While our primary interest is in the dispersed equilibria, the degenerate equilibria are also intriguing, especially since the market targets only one of a continuum of buyer types.

When a late degenerate equilibrium occurs, sellers offer $R(0) = c + \frac{b-d}{\mu}$. Indeed, $\zeta(0,0) \leq 0$ indicates that earlier buyers are not willing to pay enough to warrant targeting them. When an early degenerate equilibrium occurs, sellers offer $R(T) = c + \frac{b-d}{\mu}e^{-(\rho+\mu)T}$. Even though later buyers are willing to pay more, $\phi(0,1) \geq 0$ indicates that there are too few of them to make targeting them profitable.

Under the same mild assumption, we can also guarantee the uniqueness of equilibrium: only one of the five cases can occur, and only one equilibrium in that case can occur.

Proposition 2. *Assuming $\mu > \rho e^{\frac{\rho T}{2}}$, the equilibrium (Z^*, α^*) pair is unique.*

It is surprising that a degenerate equilibrium does not always exist, perhaps as one of multiple equilibria. In many search models, this occurs because if buyers expect a monopoly price to be offered (such as $R(0)$), they should accept that price whenever it is encountered. Here that is not true; buyers prefer to enjoy their remaining grace period utility before accepting $R(0)$. In the meantime, these buyers are potential targets for other sellers, and when $\zeta(0,0) > 0$, they are too profitable to pass up. Thus, the late degenerate equilibrium does not always exist, but only when parameters lead to $\zeta(0,0) \leq 0$.

Similar reasoning drives all of the uniqueness result. While buyers are ex-ante identical when they enter the market, they differ ex-post as some remain in the market longer than others. Those differences pin down the unique pricing strategy; any other price distribution would lead to an accumulation of buyers that are insufficiently targeted and thus too profitable to ignore.

It is noteworthy that, by Proposition 1, no equilibrium can exist where sellers target buyers with types $z \in [Z, T]$ while ignoring those $z \in [0, z)$. This is because buyers' reservation prices accelerate near the deadline — which would go untargeted in such an equilibrium. Thus, the rising willingness to pay at $Z - \epsilon$ more than compensates for the declining buyer population, so sellers will deviate to offer higher prices.

In the next subsection, we provide an illustration of our equilibrium variables.

3.3 Illustration of Equilibrium

We illustrate $R(z)$, $H(z)$, $F(p)$, and $\pi(z)$ for each possible equilibrium in Figure 1. Across the rows, we increase buyers' expiration penalty (holding all other parameters fixed), thereby motivating them to accept offers earlier in their search (*i.e.*, pushing them towards an early equilibrium).

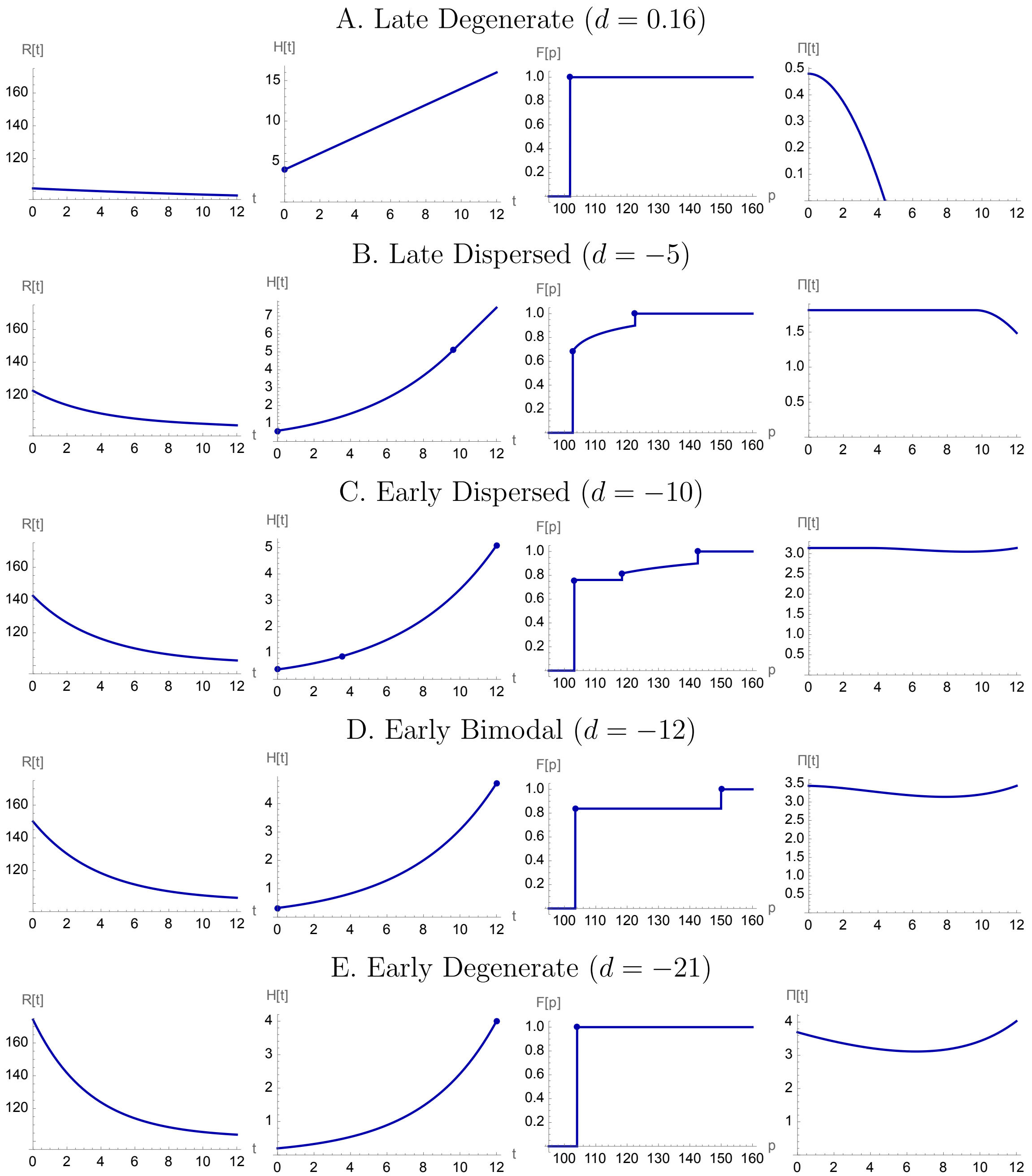

Figure 1: **Equilibrium Variables:** Reservation prices (left), population of buyers (left center), price offer distribution (right center), and profits (right), varying the expiration penalty d, while holding other parameters at $x = 105$, $c = 100$, $b = 0.64$, $\rho = 0.05$, $\mu = 0.25$, $\delta = 0.7$, and $T = 12$.

The left panels of Figure 1 show that buyers' willingness to pay rises as they get closer to the deadline and the price increase becomes more pronounced as the deadline approaches. In addition, reservation prices have greater curvature when the penalty is greater.

As we compare across equilibria, the expiration penalty affects who the sellers target. When expiration has trivial consequences, all sellers charge the highest price $R(0)$, and buyers wait until after expiration to accpet it (Panel A). As the consequences become more serious, price dispersion emerges as sellers are willing to target a continuous range of earlier (but still the most desperate) buyers, some of whom will successfully purchase before their deadline (Panel B).

As the deadline becomes even more consequential, a mass of firms will offer the deep discount $R(T)$. At first, this also includes a continuous distribution from $R(Z^*)$ to $R(0)$, meaning that the remaining firms are targeting those closest to their deadline, ignoring those between Z^* and T (Panel C). As the expiration penalty increases, this comparison shopping range shrinks, so that eventually sellers are either pricing for those who have hit their deadline, or for those who have just entered (Panel D).

With the largest penalties, only $R(T)$ is offered; no buyer does comparison shopping once they have found their product (Panel E). Indeed, across all rows, as the expiration penalty increases, sellers shift their targets more toward early buyers.

Expected profits from offering price $R(z)$ are displayed in the right panels of Figure 1, including for prices outside the support that are strictly worse. Indeed, in the early equilibria, expected profit is U-shaped between $R(Z^*)$ and $R(T)$. If sellers tried targeting buyers at $T-\epsilon$, they would find that they are not paying much more than $R(T)$, but because the atom α^* has induced many buyers to transact, fewer of them survive to state $T-\epsilon$. This draining of the buyers continues as buyers get more desperate, but eventually the reservation price accelerates enough that it more than compensates for the dwindling population of desperate buyers, generating the early dispersed or bimodal equilibrium.

As seen in Figure 1, changes in underlying parameters will affect which equilibrium emerges. We further illustrate this in Figure 2, indicating which equilibrium arises for various combinations of penalty d and grace period T. Consistent with Proposition 2, there is no overlap; each equilibrium is unique. Also, each type of equilibrium occurs across a generic parameter space, not simply for knife-edge conditions.

As the deadline penalty becomes larger or the grace period is shorter, an early equilibrium emerges, eventually collapsing to the early degenerate equilibrium. The reverse will eventually lead to a late dispersed equilibrium. The late degenerate equilibrium requires

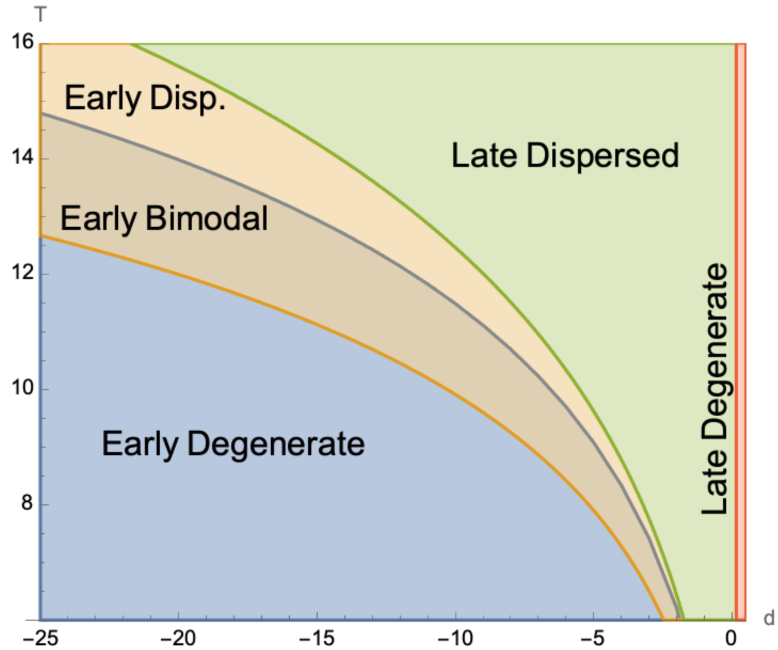

Figure 2: **Equilibrium Regions:** Each region indicates the type of equilibrium that occurs for combinations of d and T. Other parameters remain as in Figure 1.

a trivially small penalty, but does not depend on T, since all buyers are inactive before reaching the deadline anyhow. In the next subsection, we explain the forces driving these results.

3.4 Comparative Statics

To understand the importance of the equilibrium search process, we examine comparative statics. The interaction between changes in the buyers' reservation prices and the sellers' price distribution often generates counterintuitive consequences.

Specifically, we investigate the equilibrium behavior of the minimum and the maximum price, expected price, and the measure of consumers at the deadline, asking how each responds to a change in the deadline penalty d and the length of the grace period T. The signs of the comparative statics are analytically derived in most cases; whenever an analytical sign is not possible (due to the need to solve for Z^* or α^* numerically), we show the sign of the numerical derivative (shown in the Technical Appendix).

The comparative statics can largely be explained in terms of two deadline effects: *urgency* (a direct effect) and *concentration* (an indirect effect). If buyers experience a large penalty

		Early			Late	
		degenerate	bimodal	dispersed	dispersed	degenerate
$R(T)$	$\partial/\partial d$	$-$	$-^*$	$-$	$-$	$-$
	$\partial/\partial T$	$-$	$+^*$	$-$	$-$	$-$
$R(0)$	$\partial/\partial d$	$-$	$-^*$	$-$	$-$	$-$
	$\partial/\partial T$	$-$	$+^*$	0	0	0
$Mean(p)$	$\partial/\partial d$	$-$	$+^*$	$-$	$-$	$-$
	$\partial/\partial T$	$-$	$+^*$	0	0	0
$H(0)$	$\partial/\partial d$	0	$+^*$	$+^*$	$+$	0
	$\partial/\partial T$	$-$	$+^*$	$+^*$	0	0

Table 1: **Comparative Statics**. Each cell reports the sign of the analytically-derived partial derivative of a key endogenous variable with respect to d or T for a given type of equilibrium (columns), assuming parameter paths are initially set to produce a given equilibrium. Whenever an analytical derivative is not possible, the numerical derivative is presented and depicted with a *.

at the deadline (as the gap $b - d$), they are willing to pay more to avoid it. This urgency is most cleanly expressed by the willingness to pay at the deadline, $R(0)$, in Eq. 15, but can also be seen in the properties of $R(z)$ as buyers approach their deadline.

As buyers are concentrated more heavily near their deadline, sellers will offer with greater frequency the higher prices that target these late buyers. This can be seen in clearest form by looking specifically at the concentration of those who have passed their deadline, $H(0)$, expressed below.

$$H(0) = \begin{cases} \frac{\delta}{\mu} e^{-\alpha^* \mu T} & \text{if } Z^* = 0 \\ \frac{\delta}{\mu} e^{\frac{\rho Z^*}{2} - \frac{2\mu}{\rho}\left(1 - e^{-\frac{\rho Z^*}{2}}\right)} & \text{if } Z^* > 0 \text{ and } \alpha^* = 0 \\ \frac{\delta}{\mu} \frac{1 - e^{\alpha^* \mu (Z^* - T)}}{\alpha^* \mu} \left(\frac{(\alpha^* \mu + \rho) e^{\frac{\rho Z^*}{2}}}{1 - e^{(\alpha^* \mu + \rho)(Z^* - T)}} - \mu \right) e^{-\frac{2\mu}{\rho}\left(1 - e^{-\frac{\rho Z^*}{2}}\right)} & \text{if } Z^* > 0 \text{ and } \alpha^* > 0. \end{cases} \tag{22}$$

These two effects jointly determine expected profit. That is, $\pi(0) = (R(0) - c) * H(0)$; and (in all but the early degenerate equilibrium) all other prices offered in equilibrium must generate that same profit. Thus, changes in urgency and concentration indicate changes in profit that drive sellers to adjust who they target.

3.4.1 An Increase in Post-Expiration Utility, d

Suppose that the instantaneous utility after the grace period were to increase. To help visualize the effects, in Figure 2, this would be like moving horizontally across the graph at a given T. The direct effect is decreased urgency, which reduces willingness to pay: $R(0)$ and $R(T)$ fall.

One of the sharpest implications of an increase in post-expiration utility is the resulting increase in the concentration of buyers at the deadline. Lower reservation prices indirectly concentrate buyers closer to their deadline: $H(0)$ rises (except in degenerate equilibrium where it is unchanged). The reason for this concentration is that lower reservation prices reduce profitability, but disproportionately for those targeting people early in their search.[6] Thus, sellers adjust their pricing strategy, increasing the proportion of comparison-shopping offers that target buyers late in their search. This also means that buyers who would have accepted a deep discount are now less likely to find one; thus, more of them reach their deadline.

In a nutshell, the concentration effect worsens the offer distribution (resulting in higher offered prices for consumers in expectation), while the urgency effect tends to improve it. In all of the equilibria except for early bimodal equilibrium, the urgency effect on prices dominates. In an early bimodal equilibrium, there is an extreme shift in concentration, as some fraction of sellers shift from offering the deep discount to offering the highest price, causing an increase in expected price.

3.4.2 A Longer Grace Period, T

We next consider how key equilibrium variables react when buyers are given a longer grace period for their search. Visually, in Figure 2, this would be like moving vertically up the graph at a particular d.

In the late dispersed equilibrium, the equilibrium solution is unaffected by an increase in T. The reservation price path is unchanged, except that it is lengthened from $R(T)$ to $R(T+\epsilon)$. That is, buyers feel no change in their urgency in the range from 0 to T. Moreover, the sellers were already unwilling to offer any price below $R(Z^*)$, so the fact that some buyers are willing to pay even lower prices is irrelevant to them. Hence, firms leave their pricing strategy the same as profitability is unchanged: willingness to pay at the deadline, $R(0)$,

[6] For instance, suppose $R(t)$ fell by \$1 at every t. Since all buyers are willing to accept $R(T)$, the expected profit $H(T)(R(T)-c)$ from offering $R(T)$ would now be \$$H(T)$ lower. But the expected profit from offering $R(0)$ would fall by \$$H(0) < H(T)$, since there is a lower probability of it being accepted.

stays constant; buyer concentration $H(0)$ also remains as it was. Thus, the expected price offer stays the same. The same logic applies to the late degenerate equilibrium.

The concentration of consumers at the deadline exhibits interesting features. One would expect mechanically that a longer grace period results in fewer buyers reaching expiration because of more opportunities to receive a price quote — indeed this is always true in an early degenerate equilibrium. However, in the early bimodal and the early dispersed equilibria, this is more than offset by changes in the price distribution. Sellers target buyers closer to their deadline (α^* decreases), so buyers reject more of the offers they receive early in search — enough to more than reverse the mechanical direct effect. In the late equilibria, the time before Z^* is irrelevant since buyers reject all offers early on. Since Z^* does not change with T, the expected number of acceptable quotes remains unchanged.

This concentration effect interacts with the urgency effect in equilibrium, of course, since H and R are jointly determined. One would also expect willingness to pay to fall with T, as it does in the early degenerate equilibrium. But the deep discount $R(T)$ falls faster than the rest of $R(z)$, eventually making it profitable to target expired buyers with $R(0)$. This continues in the bimodal equilibrium, but that worsening of the price distribution counteracts the direct effect of more time, surprisingly making buyers willing to pay more. The effects cancel out in the late equilibria, having no net effect on the average price offer.

4 Model with Perfect Recall

So far, we have assumed that consumers cannot recall prior price quotes. This assumption may be appropriate for fast-moving consumer goods or goods that are in limited supply. For instance, in our setting, there are infinitely many firms that produce a homogenous good, however each firm produces a single unit and exits the market when it sells. In this context, the consumer who passes on a quote today cannot be sure that the seller who gave the quote will still be available later when the price offer becomes acceptable.

However, if sellers have no capacity constraints, sellers might commit to honoring a past price if the consumer returns. Intuitively, one might expect this would increase competition among firms and result in prices closer to marginal cost; however, we find that these dynamics do not always hold, and under certain parameters, firms will earn more profit in a perfect recall environment than with no recall.

Allowing for recall poses a significant technical challenge in our non-stationary search environment, because it adds a second state variable to the problem. We need to find the equilibrium solution as a function of time remaining until the deadline as well as the consumer's best price offer in-hand.

Here we find that, unlike the no-recall model, the equilibrium price distribution features at most one atom α, occurring at the deep discount $R(T)$, which any buyer is willing to immediately accept and forego further search. Comparison-shopping prices are distributed continuously on a range we denote $[\underline{p}, \bar{p}]$, where $\underline{p} > R(T)$ and $\bar{p} = R(0)$. Buyers delay accepting these offers until hitting their deadline, at which point the best offer is accepted.

4.1 Bellman Equations with Recall

With perfect recall, the value of future search includes holding the current best offer for potential future use. As a consequence, there is a unique price that can persuade buyers to accept immediately.

Lemma 1. *The highest equilibrium price that can persuade buyers to immediately accept an offer is* $R(T) = x - \frac{b}{\rho}$.

This price $R(T)$ leaves buyers indifferent between immediate acceptance (giving consumer surplus of $\frac{b}{\rho}$) and delayed acceptance (giving a flow of utility b until the deadline, plus the eventual consumer surplus $\frac{b}{\rho}$ after). In fact, this same indifference applies throughout the grace period — any price above $R(T)$ will be deferred until the deadline. The proof of Lemma 1, presented in the appendix, closely follows this intuition.

There is no incentive for sellers to offer any price less than $R(T)$, since buyers will accept this higher price. Unlike the higher prices in the comparison-shopping region, the quote $R(T)$ is executed immediately and thus never compared against competitor prices. Thus, there is no reason to attempt to undercut other sellers[7] who might offer $R(T)$.

Next, we solve for $\bar{p}$, the maximum price offered in equilibrium. If a buyer reaches her deadline with only $\bar{p}$ in hand, she should be indifferent between making the purchase versus continuing search for a better price. Indeed, any other draws are at least as good and possibly better, though waiting comes at a cost d after the deadline. The Bellman equation for such a search is:

$$\rho V(0) = d + \mu \left(\alpha(x - R(T) - V(0)) + \int_{\underline{p}}^{\bar{p}} (x - p - V(0)) dF(p) dp \right).$$

[7] This reasoning also prevents an equilibrium in which buyers use a mixed strategy to accept the deep discount. Doing so would create a mass of buyers waiting to use the deep discount at the deadline, and thus sellers would strictly benefit by offering a slightly deeper discount.

By applying indifference $V(0) = x - \bar{p}$, we can solve for $\bar{p}$ for a given $F(p)$:

$$\bar{p} = \frac{\rho x - d + \mu E[p]}{\rho + \mu}. \tag{23}$$

We postpone the solution for $\underline{p}$, as it requires equal-profit conditions derived below.

4.2 Steady State Population with Recall

The population of buyers in the market is composed of three sets of buyers. First, we must track how many buyers go through their entire grace period without receiving a quote. These buyers will then accept the first offer they receive after their deadline (rather than wait to compare against other offers). Second, we must track how many buyers are offered the deep discount and thus leave the market prior to the end of their grace period. The remaining buyers will reach their deadline and compare whatever offers they have in hand.

For firms to evaluate expected profits, it is sufficient to track the density of all buyers still in the market ($H(z)$, as in the no-recall model) and the subset of those buyers who have no quotes by time z remaining (which we label $H_u(z)$).

First consider buyers in the market that have not yet received a quote. Buyers enter without a quote, so $H_u'(T) = \delta$. They receive quotes at rate μ, so the population of quote-less buyers falls at rate $H_u''(z) = \mu H_u'(z)$. Finally, in steady state, the flow of quote-less buyers newly reaching their deadline must equal the flow of quote-less buyers at their deadline who receive a quote: $H_u'(0) = \mu H_u(0)$. This system of differential equations yields a cumulative quote-less population of:

$$H_u(z) = \frac{\delta}{\mu} e^{-\mu(T-z)}. \tag{24}$$

We must also track how many buyers are in the market at all, with their cumulative density denoted $H(z)$. Again, all buyers start in the market ($H'(T) = \delta$) but they only exit if they are offered $R(T)$, so $H''(z) = \alpha\mu H'(z)$. At the deadline, however, all buyers with quotes immediately exit, so only those without quotes remain: $H(0) = H_u(0) = \frac{\delta}{\mu} e^{-\mu T}$. This solves as:

$$H(z) = \frac{\delta}{\mu} \left(\frac{e^{\alpha\mu(z-T)} - e^{-\alpha\mu T}}{\alpha} + e^{-\mu T} \right). \tag{25}$$

The fractional term becomes μz as $\alpha \to 0$ (i.e. in a late equilibrium). Note that all buyers still in the market are relevant to sellers because they are still considering price offers; even those with a quote can be lured with a better offer.

4.3 Expected Profits with Recall

As before, sellers are unaware of the remaining time of a given buyer who requests a quote; likewise, the buyer's prior quotes are private information. If the seller offers the deep discount, the buyer's timeline and best price are irrelevant; any buyer will immediately accept, generating profit $R(T) - c$ each time a quote is generated.

If the seller offers price $p \in [\underline{p}, \bar{p}]$, he must be concerned with the possibility of being undercut by another offer. Competing offers arrive at a Poisson rate μ, and are lower than p with probability $F(p)$. Thus, if a buyer has z periods until expiration, the probability that she gets no better offers *after* this one is $e^{-\mu F(p) z}$. On the other hand, the buyer may already have a better offer in hand at time z, though we know that offer is not $R(T)$ because she would have already accepted it and exited the market. The probability that a buyer received no better price in the $T - z$ time *before* this offer p is $e^{-\mu(F(p)-\alpha)(T-z)}$.

Even if there were no better offers before or after, the sale will not occur for another z periods, and thus profit must be discounted accordingly. Thus, upon making an offer to a buyer of type z, the profit from sale is discounted by $e^{-\rho z - \mu F(p) z - \mu(F(p)-\alpha)(T-z)}$. Finally, there are those buyers already past their deadline who accept any offer immediately, so no discounting is necessary.

Combining the components of the preceding paragraphs, the expected profit upon making the offer p, averaged over all types z (including those at the deadline) is:

$$\Pi(p) = \frac{H(0) + \int_0^T e^{-\rho z - \mu F(p) z - \mu(F(p)-\alpha)(T-z)} H'(z) dz}{H(T)} (p - c). \tag{26}$$

This evaluates to $\frac{\left(\mu\left(1-e^{-\rho T}\right) e^{\mu T(1-F(p))} + \rho\right) \alpha e^{\alpha \mu T}}{\left(\alpha e^{\alpha \mu T} + e^{(\alpha+1)\mu T} - e^{\mu T}\right)\rho} (p - c)$.

Since all prices need to yield equal profits, the distribution of prices can easily be solved for by taking the derivative of the profit equation with respect to p, and equating it to zero. The resulting differential equation (reported in the Technical Appendix) is solved with the boundary that $F(\bar{p}) = 1$, yielding:

$$F(p) = 1 - \frac{\ln\left(\frac{\left(\mu\left(1-e^{-\rho T}\right)+\rho\right)\frac{\bar{p}-c}{p-c} - \rho}{\mu\left(1-e^{-\rho T}\right)}\right)}{\mu T}. \tag{27}$$

The reason there are no atoms in this portion of the distribution is that all such offers are evaluated at the deadline. In the no-recall model, each offer was evaluated in isolation, only considered against future possibilities. Here, evaluation may be against other offers, so ties become relevant. If a positive fraction of sellers offer the same price, there will be a coin flip to see which actually makes the sale at the time of the deadline (if it is the best offer). But then any seller could lower their offer by ϵ to avoid the coin flip, strictly increasing their chance of sale with an infinitesimal reduction in price.

This also allows us to use the solution $F(p)$ in order to calculate $\underline{p}$, since $F(\underline{p}) = \alpha$:

$$\underline{p} = c + (\bar{p} - c)\frac{e^{-\mu(1-\alpha)T}\left(\rho + \mu\left(1 - e^{-\rho T}\right)\right)}{\rho e^{-\mu(1-\alpha)T} + \mu\left(1 - e^{-\rho T}\right)}. \tag{28}$$

With this solution to $F(p)$ and $\underline{p}$ as functions of $\bar{p}$, we can insert these into Eq. 23 and solve for $\bar{p}$ (with intermediate steps reported in the Technical Appendix):

$$\bar{p} = c + \frac{T(\alpha\mu(\rho(x-c) - b) + \rho(\rho(x-c) - d))}{\rho(\mu+\rho)T + (\rho + \mu\left(1 - e^{-\rho T}\right))\ln\left(1 - \frac{\rho\left(1-e^{-(1-\alpha)\mu T}\right)}{\rho+\mu\left(1-e^{-\rho T}\right)}\right)}. \tag{29}$$

We can also substitute for $F(p)$ in the profit equation to get the expected discounted profit.

$$\Pi(\alpha) = \frac{\alpha\left(\rho + \mu\left(1 - e^{-\rho T}\right)\right)}{\rho\left(\alpha + e^{\mu T} - e^{(1-\alpha)\mu T}\right)}(\bar{p} - c). \tag{30}$$

Profit across all comparison-shopping prices are now constant by construction, so with slight abuse of notation, we define Π as a function of the atom at $R(T)$. If offered in equilibrium, the deep discount must be equally profitable, so $R(T) - c = \Pi(\alpha)$ if $\alpha \in (0, 1)$.

4.4 Equilibrium with Recall

The preceding analysis pins down the population dynamics, the pricing behavior by sellers, and the willingness to pay by buyers. These necessary conditions lead to three possible equilibria depending on the atom at $R(T)$. We prove that in equilibrium, one and only one of these will occur. The price distribution and profits in each equilibrium are illustrated in Figure 3.

Proposition 3. *With perfect recall, a solution $R(z)$, $F(p)$, and $\Pi(z)$ is an equilibrium if and only if it satisfies Eqs. 27 through 29 with one of the following cases:*

- *(late dispersed)* $\alpha^* = 0$ *if* $\Pi(0) \geq x - c - \frac{b}{\rho}$
- *(early dispersed)* $\alpha^* \in (0, 1)$ *if* $\Pi(\alpha^*) = x - c - \frac{b}{\rho}$

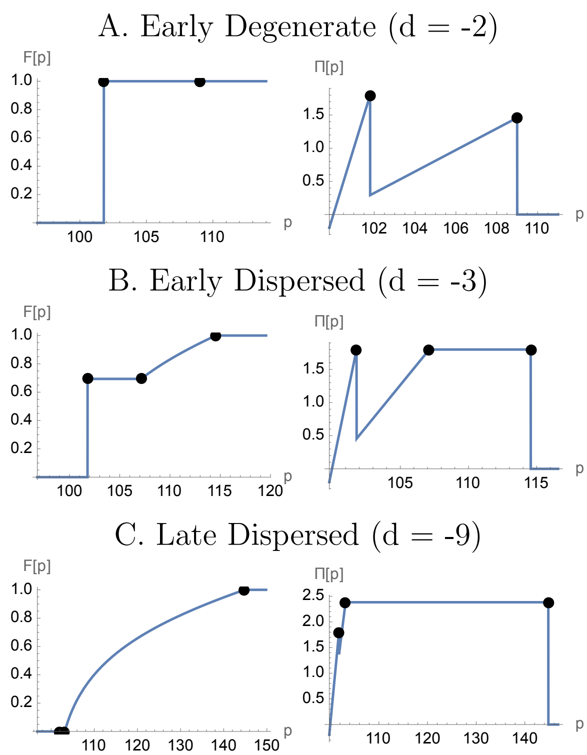

Figure 3: **Equilibrium Solutions:** price offer distribution (left), and profits (right), varying the expiration penalty d, while holding other parameters at $x = 105$, $c = 100$, $b = 0.16$, $\rho = 0.05$, $\mu = 0.25$, $\delta = 0.7$, and $T = 12$. Dots indicate $R(T)$, $\underline{p}$ and $\bar{p}$, respectively.

- *(early degenerate)* $\alpha^* = 1$ *if* $\Pi(1) \leq x - c - \frac{b}{\rho}$

Only the first case can occur if $\frac{b}{\rho} \geq x - c$*; otherwise, exactly one of these equilibria can occur.*

With perfect recall, search deadlines can still create price dispersion, but through a different mechanism. In the no-recall model, when firms make an offer to a buyer, they are unable to observe the buyer's time remaining, and thus face a tradeoff of higher markup but lower probability of acceptance. When making an offer with perfect recall, sellers are unable to observe the set of competing offers that the buyer will hold at her deadline, and thus face a tradeoff of higher markup but lower probability of being the best offer. Indeed, this mirrors the uncertainty sellers face in traditional models of simultaneous search,[8] except in our

[8]In Burdett and Judd (1983), buyers select a fixed number of quotes which is not known to the seller. Price dispersion only occurs there because some buyers seek only one quote; otherwise, Bertrand-like competition pushes prices to marginal cost. In our setting, the search friction naturally ensures that some buyers will only get one quote.

model, receiving those competing offers takes time, which sellers account for in discounting expected future profits. In this sense, our model with perfect recall embeds the flavor of the simultaneous search models and brings those flavors to a sequential search framework.[9]

The late dispersed equilibrium is guaranteed to occur when $\frac{b}{\rho} \geq x - c$. That is, when there is positive flow of utility while searching that is better than the net welfare gain from trade, buyers will want to enjoy all of those benefits until the deadline, and only then execute the best offer in hand. Indeed, sellers earn negative profits if they offer a price that is low enough to induce buyers to forego the full sequence of search benefits. Those parameters might be applicable if the buyer is replacing an existing product, such as looking for a new apartment (in a new city, perhaps) while her current lease is still in force. If the current apartment is at least as good as the new one and rent quotes can be recalled, she is in no rush to pay for a new one earlier than the lease expiration.

With that intuitive explanation, one might expect that $b < 0$ would ensure an early degenerate equilibrium, but that is not the case. It is possible to sustain a late dispersed equilibrium with $b < 0$ so long as d is sufficiently negative as well. This occurs because as d falls, the maximum price $\bar{p}$ rises. Buyers who have reached expiration are willing to pay more, which sellers exploit. Thus, even if $b < 0$ so that price $R(T)$ would cover the seller's cost of production, it may still be less profitable than targeting those near expiration.

Indeed, the possibility of a deep discount creates added nuance between the sequential and simultaneous search behavior. Offers in the comparison-shopping region are evaluated by buyers simultaneously at the deadline, disciplined by competition to beat other offers at that time. Offers of $R(T)$ are evaluated by buyers sequentially, being acted on immediately. These offers are never compared against another offer, but they are disciplined by competition against *prospective* offers through further search. Indeed, this is the competitive force in the non-recall world, and is also the force that limits the highest price $\bar{p}$ in the recall world; buyers past their expiration still have the option to continue their search, but in equilibrium are indifferent about doing so if offered $\bar{p}$.

It is noteworthy that a late degenerate or an early bimodal equilibrium cannot occur with perfect recall. Both require an atom of firms offering $\bar{p}$, which is only accepted at the deadline. But that means slightly lower offers will undercut that price when they are later compared at the deadline, so sellers would deviate from the mass at $\bar{p}$.

[9]Stahl (1989) has a similar feel in that some buyers never search more than once while others seek quotes from all N firms. In our model, buyers are ex-ante identical, differing only ex-post in their luck as to how many quotes they received.

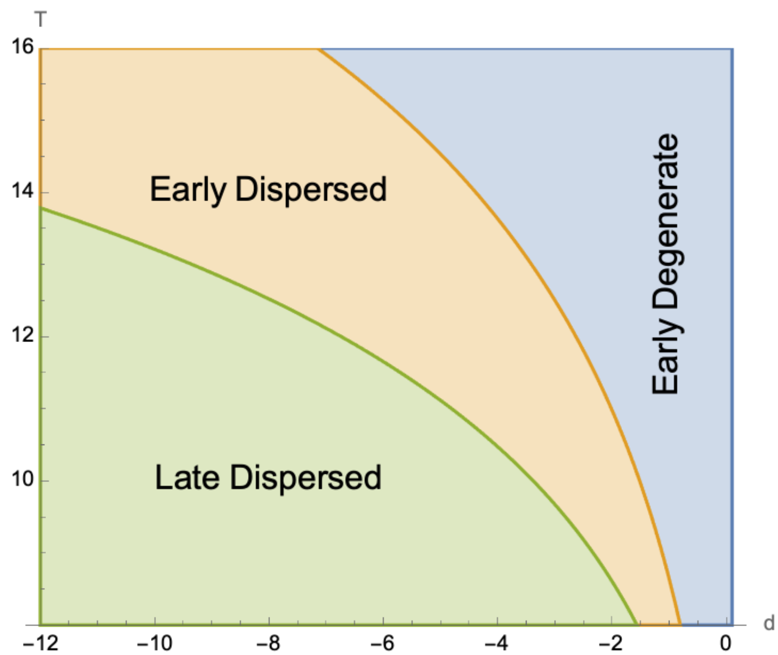

Figure 4: **Equilibrium Regions with Recall:** Each region indicates the type of equilibrium that occurs for combinations of d and T. Other parameters remain as in Figure 1.

Finally, we would like to underscore that the seller behavior is the driving force of the equilibrium features with perfect recall. Sellers are making complicated choices of whom to target and how hard to compete. Buyers, on the other hand, have a simple choice of accepting $R(T)$ if it is offered — otherwise, they hold on to all other offers and evaluate them at the end of the grace period. Hence, the order in which they are received carries no information; the recalled price is uniformly likely to have been received at any point during the grace period.

4.5 Comparative Statics with Recall

In stark contrast to Figure 2, perfect recall reverses the order in which equilibria occur as the deadline penalty changes (given a fixed T). As seen in Figure 4, small penalty levels give rise to an *early* (rather than *late*) degenerate equilibrium. This interesting reversal is simply due to recall empowering consumers: consumers' maximum willingness to pay $\bar{p}$ is low, and thus sellers would rather get the immediate purchase at the deep discount than wait for a slightly greater markup.

		Early		Late
		degenerate	dispersed	dispersed
$R(T)$	$\partial/\partial d$	0	0	0
	$\partial/\partial T$	0	0	0
$R(0)$	$\partial/\partial d$	$-$	$-^*$	$-$
	$\partial/\partial T$	0	$-^*$	$-^*$
$Mean(p)$	$\partial/\partial d$	0	$-^*$	$-$
	$\partial/\partial T$	0	$-^*$	$-^*$
$H(0)$	$\partial/\partial d$	0	0	0
	$\partial/\partial T$	$-$	$-$	$-$

Table 2: **Comparative Statics with Recall**. Each cell reports the sign of the analytically-derived partial derivative of a key endogenous variable with respect to d or T for a given type of equilibrium (columns), assuming parameter paths are initially set to produce a given equilibrium. Whenever an analytical derivative is not possible, the numerical derivative is presented and depicted with a *.

As the deadline penalty increases, those who have reached their deadline are willing to pay more, yet $R(T)$ is unchanged. Eventually, comparison-shopping quotes become equally profitable and an early dispersed equilibrium occurs. Indeed, as the penalty increases further, the higher willingness to pay $\bar{p}$ is balanced by more competition among comparison-shopping quotes. Within the early dispersed equilibrium, these forces balance out to be equally profitable as offering $R(T)$; but with a large enough penalty, even with competition, the comparison-shopping quotes will be strictly more profitable.

Comparative statics with respect to d and T (reported in Table 2) are interesting. Since the deep discount $R(T) = x - \frac{b}{\rho}$ doesn't depend on d or T, its derivatives are all zero. Economically, $R(T)$ makes the buyer indifferent about immediate acceptance, where the alternative is to wait until the deadline and exercise the same (best) price. Thus the deadline becomes irrelevant once this offer is received, removing all urgency. In contrast, $R(T)$ typically falls with d or T in the no-recall model because if the buyer were to reject the $R(T)$ offer, she runs the risk of not finding it or any other offer before expiration.

Even so, the deadline penalty and grace period are still relevant to comparison shopping prices exercised at the deadline, as illustrated by the maximum price $R(0)$. This price specifically targets those who reach their deadline with no quotes, who will incur the penalty until the next offer arrives. Thus, reducing the penalty (increasing d) will reduce the willingness to pay, which happens for the same reason in the no-recall model. A longer grace period also decreases $R(0)$ in the recall setting because more time will increase the number of offers compared at the deadline (in the dispersed equilibria). This contrasts with the no-recall model, where $R(0)$ increases with T during an early bimodal equilibrium, and is constant in a dispersed equilibrium.

The average price with recall is simply the deep discount in an early degenerate equilibrium, which thus does not depend on d or T. For the dispersed equilibria, average price declines in both d and T, again because they encourage competition at the deadline to get consumers to exercise their best option. In the no-recall model, a higher d also encourages lower prices as firms compete with the outside option of waiting longer. A longer T, however, has no effect on average prices in the dispersed equilibria.

The measure of buyers who reach their deadline in the recall setting is simply those who received no quotes, as all others exit early or right as the deadline occurs. Thus, $H(0)$ depends only on the arrival rate of offers μ and on the time given T. Thus, a longer grace period mechanically increases the chance of having at least one quote, but a change in the deadline penalty has no effect. In the no-recall model, this same mechanical effect exists, but the price distribution also shifts with T to target the buyers later in their search, sufficient to cancel out or even reverse this mechanical concentration effect.

5 Welfare

In this section, we consider the welfare implications of search frictions, with a particular focus on who benefits from the ability to recall prices and when. We first define welfare metrics in this setting, and then explore how these are affected by parameter values.

5.1 Welfare metrics

This search market generates value for buyers and sellers, which we measure from the perspective of a new entrant. For buyers, this is ex-ante consumer surplus, $V(T)$, which anticipates the discounted expected flow of grace period benefits b, penalties d, consumption utility x, and the price at which it occurs. In the no-recall model, $V(T) = x - R(T)$, which is:

$$V(T) = \frac{b}{\rho} + \frac{\alpha^*\mu + \rho e^{(\alpha^*\mu+\rho)(T-z)}}{\alpha^*\mu + \rho e^{(\alpha^*\mu+\rho)(T-Z^*)}} \left(x - \frac{b}{\rho} - R(Z^*) \right). \tag{31}$$

This includes the expected flow of benefits b, which is potentially cut short if the buyer accepts an offer before expiration (captured in the second fractional term).

In the recall model, ex-ante consumer surplus is:

$$V(T) = \frac{b}{\rho} - \frac{b}{\rho} e^{-(\alpha\mu+\rho)T} + e^{-(\rho+\mu)T} \left((x - \hat{p}) \left(e^{(1-\alpha)\mu T} - 1 \right) + \frac{\mu x + d - \mu E[p]}{\mu + \rho} \right), \tag{32}$$

where $\hat{p}$ is the expected best quote when evaluated at the deadline. This utility has three components (matching the order in which they appear in the equation): the flow of search benefits b (which an offer $R(T)$ exactly replaces), the utility from offers received in the grace period but executed at the deadline, and the utility from having no offers in the grace period but taking the first thereafter.

Total welfare includes the same benefits, but compares them relative to the discounted expected cost of production, c, rather than the purchase price. In the no-recall model, this is solved from the differential equations

$$\rho W(0) = d + \mu(x - c - W(0)) \tag{33}$$

$$\rho W(z) = b - W'(z) + \mu F(R(z))(x - c - W(z)). \tag{34}$$

This indicates that after the deadline, buyers incur the penalty d and accept any offer, which occurs at rate μ and generates net welfare $x - c$. Prior to the deadline, buyers receive utility b, have their grace period steadily decrease, and accept offers at rate $\mu F(R(z))$.

In the recall model, total welfare uses the same post-expiration equation, but pre-expiration welfare is computed by:

$$\rho W(z) = b - W'(z) + \mu\alpha(x - c - W(z)) + \mu(1-\alpha)(\omega(z) - W(z)) \tag{35}$$

$$\rho\omega(z) = b - \omega'(z) + \mu\alpha(x - c - \omega(z)) \tag{36}$$

$$\omega(0) = x - c. \tag{37}$$

Here, we distinguish between expected welfare without any retained offers, $W(z)$, and expected welfare after getting an offer that is not immediately executed, $\omega(z)$. The latter occurs at rate $\mu(1-\alpha)$, while immediately-accepted $R(T)$ offers occur at rate $\mu\alpha$. Once an offer is retained, the buyer is assured of being able to consume at the deadline; but it can still happen sooner if a deep discount offer is found in the meantime.

Discounted expected profit is necessarily $W(T) - V(T)$ in both models. This is measured in terms of how much profit some seller receives from a consumer who just entered the market with a grace period of T; in other words, this is Π discounted by the expected duration of search before purchase.

In the absence of any search friction, buyers could produce and consume the good at any moment they desired. When $b < \rho(x-c)$, consuming immediately on entering the market generates the most social welfare ($x-c$ utils per buyer). When $b > \rho(x-c)$, welfare is greatest when the buyer enjoys the full stream of grace-period benefits and then consumes right at expiration, contributing $\frac{b}{\rho}(1-e^{-\rho T})+(x-c)e^{-\rho T}$ per buyer. Note that this first-best welfare does not depend on the penalty d because a frictionless world can always avoid the penalty phase.

5.2 Welfare response to deadline length or penalty

To understand the advantages posed by a market with recall versus one without, we now examine welfare under various parameters, beginning with the size of the deadline penalty d and time until it is incurred T. Both parameters have a similar, intuitive impact; Panel A of Figure 5 illustrates this effect for d. Under the given parameterization, the Recall setting produces a late dispersed equilibrium for all values of d, while the No Recall setting progresses from early degenerate through late dispersed equilibria.

With or without recall, consumers are better off with less severe penalties or a longer grace period. Both have an obvious direct impact on utility by reducing or delaying the pain after expiration. They also reduce urgency for the buyers, preventing the sellers from charging as much and reducing their profits.[10] The gain to buyers outweighs the loss to sellers, though, so that total welfare is increasing in d or T with or without recall.

For a given d and T, we ask who fares better with recall versus without. One might expect that recall is always better for buyers, but this is only true when the penalty is small or grace period is long. Likewise, one would expect that sellers earn less profit under recall, but this is not true with large penalties or short deadlines. Both unexpected results come from how recall affects prices. Recall generates intense competition and thus lower prices when deadlines have minor or far-off consequences; but if the deadlines loom large, that competition erodes faster and the average price increases faster with recall than without.

Surprisingly, which setting generates more total welfare does not depend on d or T at all. Rather, when the first-best outcome would recommend delayed consumption ($b > \rho(x-c)$,

[10] The only exception is in an early bimodal equilibrium without recall; longer grace periods actually increase prices and leave buyers worse off.

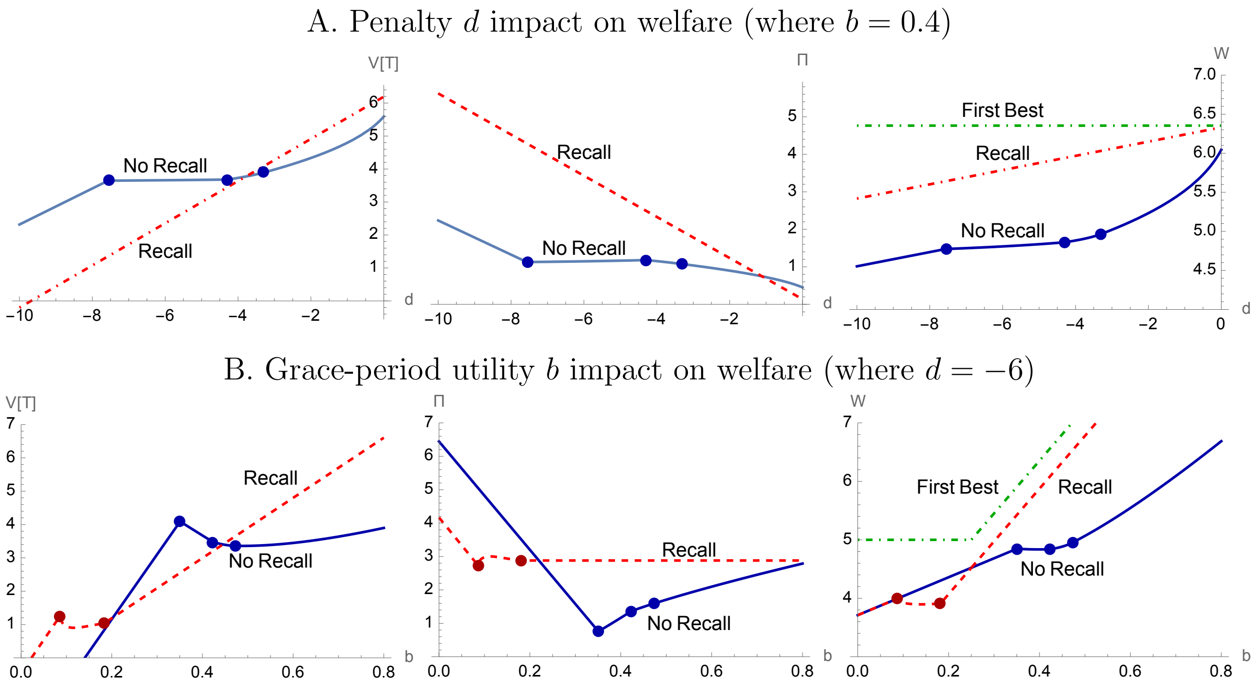

Figure 5: **Equilibrium welfare impact** in the No-recall model (solid) and Recall model (dashed), shown for ex-ante consumer surplus $V(T)$, discounted expected profits Π, and total welfare (their sum). The frictionless first-best welfare is shown with a dot-dashed line. All other parameters as in Figure 1. On each graph, dots indicate the transitions between equilibria, from early degenerate (leftmost) to late dispersed (rightmost). Only the late dispersed equilibrium occurs for Recall in Panel A.

as pictured in Panel A), the recall setting generates more total welfare, and vice versa. We examine the cause of this in the next subsection.

5.3 Welfare response to grace-period benefits

At first glance, b and d seem to play a related role; one might even expect that only their difference matters. However, b has a more subtle impact on welfare, as depicted in Panel B of Figure 5, particularly since b can change which equilibrium occurs in the Recall setting.

First, we see larger benefits b have a non-monotonic effect in both settings. Intuition suggests that buyers prefer larger benefits, but during an early bimodal or dispersed equilibrium, sellers shift to target later buyers, driving prices up enough that buyers are worse off. Similarly, one would expect sellers to be worse off as b increases, but aside from the early degenerate equilibrium, this is not necessarily true. The total welfare is also non-monotonic, falling whenever a bimodal (no-recall) or early dispersed (recall) equilibrium occur, but rising otherwise.

In both settings, the local maximum in utility and local minimum in profit occurs at the transition from the early degenerate equilibrium. But that occurs at a higher b in the no-recall model. This enables a market without recall to produce more consumer surplus than a market with recall for moderate values of b. Put another way, the disadvantage of recall is that buyers make use of it — by delaying the acceptance of some offers — even when the benefit of waiting is somewhat small. The non-recall setting continues offering the deep discount for a larger range of b. Even so, recall (particularly in a late dispersed equilibrium) has the advantage of ensuring that buyers enjoy their full flow of benefits b, which becomes more important as b becomes large.

Social welfare is best understood relative to the first-best outcome, depicted with the dot-dashed line in Panel B on the right in Figure 5. Its kink occurs when $b = \rho(x - c)$, so that the grace-period utility equals the flow of welfare after purchase. For lower values of b, total welfare is maximized by encouraging early consumption, which is exactly what an early degenerate equilibrium will do. Indeed, recall and non-recall models generate the same welfare if both are in an early degenerate equilibrium.

The recall environment can generate dispersed equilibria even when $b < \rho(x - c)$, while the non-recall environment never does. Thus, the non-recall setting provides greater welfare for low b. Indeed, in this range, one can think of the gap between first-best and no-recall utilities as the "matching friction" that occurs simply because of the random arrival of trading partners, leaving some buyers unmatched for a time. Meanwhile the gap between the no-recall and recall welfare is a "pricing friction," causing inefficient delay because a recall buyer want to comparison shop until the deadline.

In contrast, when $b > \rho(x - c)$, total welfare is maximized if buyers can enjoy their full grace period. Ideally, one can consume right at the deadline; but at least a market with recall in a late dispersed equilibrium can achieve that for everyone who got an offer prior to expiration. Indeed, in this range, the gap between first-best and recall welfare is the "matching friction," reflecting the disutility of those who received no offers before hitting their deadline.

The non-recall world performs worse in this range for two reasons. First, some buyers are accepting offers prior to expiration, which is inefficient (even if it is individually rational to avoid losing a good price). Second, some buyers reject an offer early in their search but expire before accepting another, and thus unnecessarily incur the penalty (which is also individually rational to seek a better price). Thus, the gap between recall and no-recall welfare is the "pricing friction" that encourages buyers to consume at an inefficient time.

The examination of welfare underscores the importance of a complete characterization of possible equilibria. While the ability for recall seems like an unambiguous advantage to buyers and welfare generally, this is not necessarily true. Indeed, what flips the conventional wisdom is that the type of equilibrium will determine the targeting of buyers and therefore the timing of their consumption. When recall and non-recall lead to different types of equilibria, the comparison is less obvious; but whichever more closely follows the first-best timing will generate the most welfare.

6 Conclusion

We analyze price formation and welfare in a continuous time search model where buyers face a deadline to complete their transaction. Sellers are uninformed about a potential buyer's remaining time or which quotes she has received; therefore, they engage in probabilistic price discrimination, limiting who will accept their offers based on which price is posted. We solve for the endogenous price distribution that prevails in the steady state by translating the equilibrium conditions of the dynamic search problem into a set of differential equations. The unique equilibrium is characterized by whether it includes a deep discount and/or a strictly higher continuous range of comparison-shopping prices.

The types of equilibria are pivotal in understanding welfare consequences of search. Indeed, it is surprising that recall can harm consumers and reduce total welfare. This scenario occurs when recall generates price dispersion that encourages inefficient comparison shopping, while only the deep discount is offered when there is no recall.

One virtue of our model is that search provides the consistent disciplining force throughout. Buyers' willingness to pay is endogenously determined by the value of continued search. Rather than fixing a number of quotes, buyers always have the option to gather more, even after crossing the deadline. Indeed, even the highest price is disciplined by buyers being able to seek another quote, rather than setting an exogenous outside option (as in Coey, *et al*, 2020). This unified framework makes for a cleaner comparison when recall is added, which generates a price distribution similar to those in simultaneous search, but derived from a sequential setting and with the possibility of a deep discount.

Our analysis considers the two extremes of perfect recall and no recall; an intermediate case might include some chance of losing past offers, as in Akin and Platt (2014). This would be more complicated to execute in a continuous time model — for instance, in that setting, no recall could sustain at most two prices. Even so, we anticipate that the resulting price distribution and welfare results would lie between the two extremes.

Our model also assumes homogeneity in buyer valuations, grace period length, penalties, and seller costs, which is an intentional modeling choice to isolate the pricing incentives created by the impending deadline that differs ex-post across the buyers. These incentives would still be present after adding exogenous heterogeneity, though it would affect who sellers target. If the heterogeneous buyers have reservation price ranges that overlap, we expect that the equilibrium price distribution will be qualitatively similar to those depicted here. With wider heterogeneity, it could be that the market will ignore some less-populous types.

Our model can easily be adjusted to incorporate other plausible features of search on a deadline. Service or subscription contracts sometimes give new customers an introductory rate in exchange for committing to the service for a fixed period of time, after which the monthly price could be much higher. As the contract nears its conclusion, the customer becomes interested in shopping for a new contract. Of course, while the customer wants to avoid the rate increase after the contract expires, she is already committed to service with her current provider through the end of the contract, and would like to avoid early termination fees. These features mostly just change the interpretation of pre- and post-expiration benefits; the one modeling difference is that the flow of new buyers now endogenously arises from the rate at which buyers sign new contracts.

Another adaptation would allow buyers who purchase the good to delay consuming it until their deadline, rather than at the time of purchase. For instance, the deadline might represent the start date of a new job in Minneapolis, but the buyer must continue working in Miami up until then. If so, the household enjoys b up until the deadline regardless of when the purchase is made. If the deadline is crossed, however, the household must use expensive short-term living arrangements d until the permanent housing is secured. Compared to our no-recall model, this only removes the effect of b, since the grace-period benefits are effectively sunk, obtained regardless of the timing of purchase.

A Proofs

A.1 Proposition 1

Proof. The translation process demonstrated in the technical appendix ensures that the $R(z)$ function is equivalent to conditions 1 and 3 in the equilibrium definition, and demonstrates that all prices in the support of $F(p)$ are equally profitable. Indeed, in an early dispersed or bimodal equilibrium, $R(0)$ and $R(T)$ are equally profitable by construction.

We only need to verify that any price outside the support generates weakly lower profits. Of course, any price above $R(0)$ will be rejected by all buyers and hence generates zero profit. Any price below $R(T)$ will reduce revenue per sale without increasing the number of potential buyers, and hence is strictly not preferred. Last we inspect the impact of offering a price below $R(Z^*)$, taking as given $H(z)$ and $R(z)$. We proceed assuming that this is a dispersed equilibrium; the same logic applies in a degenerate equilibrium, after appropriately substituting the equations used.

We proceed by computing $\Pi(z, \alpha^*) = (R(z) - c)H(z)$ for $z \in (Z^*, T)$ from the proposed solution in Eqs. 16 and 18. See the technical appendix for details in the algebraic manipulations.

In a dispersed equilibrium (late or early), offering a price $R(z)$ where $z \in (Z^*, T)$ would generate profit:

$$\Pi(z) \equiv \left(\frac{e^{\alpha^* \mu (z - Z^*)} - 1}{\alpha^*} + e^{\frac{\rho Z^*}{2}} \right) \cdot \left((\alpha^* \mu + \rho) e^{\frac{\rho Z^*}{2}} - \mu \left(1 - e^{-(\alpha^* \mu + \rho)(z - Z^*)} \right) \right)$$
$$\cdot \frac{(b-d)\delta e^{-\alpha^* \mu T - \frac{2\mu}{\rho}\left(1 - e^{-\frac{\rho Z^*}{2}}\right) - \frac{\rho Z^*}{2}}}{\mu^2 (\alpha^* \mu + \rho)}.$$

Note that the last fraction is always positive and constant w.r.t. z and is therefore omitted in the following derivatives.

In the case of a late dispersed equilibrium (with $\alpha^* = 0$), the first fraction becomes $\mu(z - Z^*)$ as $\alpha^* \to 0$. If a firm targets a higher z, profit changes by:

$$\Pi'(z) \equiv -\mu \left(\mu \rho (z - Z^*) e^{-\rho(z - Z^*)} + \left(\mu - \rho e^{\frac{\rho Z^*}{2}} \right) \cdot \left(1 - e^{-\rho(z - Z^*)} \right) \right).$$

Because $z > Z^*$ and $\mu > \rho e^{\frac{\rho T}{2}}$ by assumption, profit is always decreasing when shifting the target earlier to some $z > Z^*$. We use the same strategy for the late degenerate equilibria (reported in the technical appendix), showing that profit strictly decreases as the targeted z increases.

In an early dispersed equilibrium, we find $\Pi'(Z^*) = 0$ and $\Pi''(Z^*) = -\mu(\alpha^*\mu + \rho) \cdot \left(2\mu - (2\alpha^*\mu + \rho)e^{\frac{\rho Z^*}{2}}\right) < 0$. The latter inequality must hold because $F(R(Z^*)) = e^{-\frac{\rho Z^*}{2}} - \frac{\rho}{2\mu}$ and must be larger than $F(R(T)) = \alpha^*$, which rearranges to yield $2\mu > (2\alpha^*\mu+\rho)e^{\frac{\rho Z^*}{2}}$. Thus, any price just below $R(Z^*)$ will be strictly less profitable.

For the remaining potential targets $z \in (Z^*, T)$, we show that $\Pi''(z)$ will change sign only once (from negative to positive). Suppose that $\Pi''(\hat{z}) = 0$ at some $\hat{z} \in (Z^*, T)$, yielding:

$$\rho^2 e^{\alpha^*\mu\hat{z}} = \alpha^{*2}\mu e^{\alpha^*\mu\hat{z}} \left(\mu - e^{\frac{\rho Z^*}{2}}(\alpha^*\mu + \rho)\right) e^{(\hat{z}-Z^*)(\alpha^*\mu+\rho)} + (\alpha^*\mu + \rho)^2 e^{\alpha^*\mu Z^*} \left(1 - \alpha^* e^{\frac{\rho Z^*}{2}}\right).$$

We then take the third derivative evaluated at $\hat{z}$, substituting for $\rho^2 e^{\alpha^*\mu\hat{z}}$ using the previous equation:

$$\Pi'''(\hat{z}) \equiv (\alpha^*\mu + \rho)\left(1 - \alpha^* e^{\frac{\rho Z^*}{2}}\right) e^{\alpha^*\mu Z^*} + \alpha e^{\alpha^*\mu\hat{z}} \left((\alpha^*\mu + \rho)e^{\frac{\rho Z^*}{2}} - \mu\right) e^{(\hat{z}-Z^*)(\alpha^*\mu+\rho)}.$$

Since $e^{-\frac{\rho Z^*}{2}} - \frac{\rho}{2\mu} > \alpha^*$ in an early dispersed equilibrium (as noted above), it is also the case that $e^{-\frac{\rho Z^*}{2}} > \alpha^*$, so $1 > \alpha e^{\frac{\rho Z^*}{2}}$. Moreover, from $\phi(Z^*, \alpha^*) = 0$, we find that $e^{\frac{\rho Z^*}{2}}(\alpha^*\mu+\rho) > \mu$, as described in the technical appendix. Thus, $\Pi'''(\hat{z}) > 0$ whenever $\Pi''(\hat{z}) = 0$. Thus, there is only one such $\hat{z}$.

Thus, as z increases, profit initially falls near Z^*, but eventually will increase (as the second derivative turns positive, then eventually the first derivative). However, it can never rise above the profit $\Pi(T)$, because if it did, then it would have to later decrease before reaching $\Pi(T)$. This would generate a point where $\Pi''(\hat{z}) = 0$ but $\Pi'''(\hat{z}) < 0$. In the technical appendix, we use this same strategy for a bimodal or an early degenerate equilibrium, showing that the third derivative is positive when the second is zero.

Thus, the proposed equilibrium solution satisfies all the necessary conditions for equilibrium. □

A.2 Proposition 2

Proof. We proceed by demonstrating that if the conditions for one of the five equilibria holds, it precludes any of the others. As a preliminary, note that the first derivative of ζ w.r.t. Z

is:

$$\zeta_Z(Z,\alpha) = \frac{(b-d)\left(\rho+\alpha\mu e^{(\rho+\alpha\mu)(Z-T)}\right)e^{-\frac{2\mu}{\rho}\left(1-e^{-\frac{\rho Z}{2}}\right)-\rho Z}}{2(\rho+\alpha\mu)}\left((\rho+2\alpha\mu)e^{\frac{\rho Z}{2}}-2\mu\right) < 0,$$

where the sign holds because all the first terms are positive, while the last term is negative because in a early dispersed or any late equilibrium, $F(R(Z^*)) = e^{-\frac{\rho Z^*}{2}} - \frac{\rho}{2\mu}$ and must be larger than $F(R(T)) = \alpha^*$ (and ζ is not relevant in an early bimodal or degenerate equilibrium). The derivative w.r.t. α is:

$$\zeta_\alpha(Z,\alpha) = \frac{\mu(b-d)e^{-\frac{2\mu}{\rho}\left(1-e^{-\frac{\rho Z}{2}}\right)-\frac{\rho Z}{2}}}{(\rho+\alpha\mu)^2\, e^{(\rho+\alpha\mu)(Z-T)}}\left(\rho-\rho e^{(\rho+\alpha\mu)(T-Z)}-\alpha\mu(\rho+\alpha\mu)(T-Z)\right) < 0,$$

where the sign holds because the first terms are positive, and the last term is negative because $T > Z$.

First, suppose a late degenerate equilibrium occurs ($\zeta(0,0) \le 0$). Since $\zeta_Z(Z,0) < 0$, there is no $Z > 0$ for which $\zeta(Z,0) = 0$. Moreover, $\zeta_\alpha(Z,\alpha) < 0$, so likewise, any $\alpha > 0$ would result in $\zeta(Z,\alpha) < 0$, precluding an early equilibrium.

Next, suppose a late dispersed equilibrium occurs ($\zeta(Z^*,0) = 0$). Since $\zeta_Z(Z,0) < 0$ for all $Z \in [0,T]$, so no other late dispersed equilibrium can exist, nor can a degenerate equilibrium since $\zeta(0,0) > 0$. Likewise, $\zeta_\alpha(Z,\alpha) < 0$, so any increase in α will ensure that $\zeta(Z,\alpha) < 0$, thereby precluding an early equilibrium.

The same approach applies to the early dispersed equilibria. However, since $\zeta_Z(Z,\alpha) < 0$ and $\zeta_\alpha(Z,\alpha) < 0$, the possibility is introduced that one can offset an increase in Z with a decrease in α so as to maintain $\zeta(Z,\alpha) = 0$. However, this will also disrupt the $\phi(Z,\alpha) = 0$ condition because $\phi_Z(Z,\alpha) < 0$ and $\phi_\alpha(Z,\alpha) > 0$, as shown below (with details in the Technical Appendix):

$$\phi_Z(Z,\alpha) \;=\; \frac{\mu e^{-\frac{\rho Z}{2}}e^{\rho(Z-T)}\left(\rho(2\alpha\mu+\rho)\left(e^{\rho(T-Z)}-1\right)-\rho^2\left(e^{(T-Z)(\alpha\mu+\rho)}-e^{-\alpha\mu(T-Z)}\right)\right)}{2(2\alpha\mu+\rho)} < 0.$$

This is negative because $\phi_Z(T,\alpha) = 0$ and each element is increasing in Z for all $Z < T$.

$$\begin{aligned}\phi_\alpha(Z,\alpha) \;=\;& \mu(T-Z)e^{\alpha\mu(T-Z)}\left(e^{\alpha\mu(T-Z)}+\alpha e^{\frac{\rho Z}{2}}-1\right)\left(\alpha\mu+\rho e^{(T-Z)(\alpha\mu+\rho)}-(\alpha\mu+\rho)e^{\rho(T-Z)}\right)+\\ &\left(e^{(T-Z)(\alpha\mu+\rho)}-1\right)\left(e^{\alpha\mu(T-Z)}-1-\alpha\mu(T-Z)e^{\alpha\mu(T-Z)}\right)\cdot\\ &\qquad\left(\rho e^{(T-Z)(\alpha\mu+\rho)+\frac{\rho Z}{2}}-\mu\left(e^{\alpha\mu(T-Z)}-1\right)\right) > 0.\end{aligned}$$

This is positive because $\phi_\alpha(Z,0) = 0$, and each parenthetical term in ϕ_α is increasing in α. Thus, if (Z^*, α^*) constitutes an early dispersed equilibrium, any alternative $(\hat{Z}, \hat{\alpha})$ will either violate $\zeta(\hat{Z}, \hat{\alpha}) = 0$ or $\phi(\hat{Z}, \hat{\alpha}) = 0$. Indeed, if followed until $\alpha = 0$, we rule out a late dispersed equilibrium, or until $Z = 0$, we rule out an early bimodal equilibrium.

Finally, note that $\phi_\alpha(0, \alpha^*) < 0$, so if an early bimodal equilibrium exists ($\phi(0, \alpha^*) = 0$), then an early degenerate cannot (since $\phi(0,1) < 0$) and vice versa.

□

A.3 Lemma 1

Proof. Suppose a buyer with z periods remaining is offered price p which is the lowest price in the support of $F(p)$. If she immediately accepts this price, she receives utility $x - p$.

If instead she continues her search, note that she already has a quote for the best price available, so she will not find a better price. Rather, she is merely delaying the purchase at that price, getting flow utility b in the meantime. This "search" will give expected present utility of:

$$\frac{b}{\rho}\left(1 - e^{-\rho z}\right) + (x - p)e^{-\rho z}.$$

The highest such price would be where the buyer is indifferent between search and acceptance. Setting these equal and solving yields $p = x - \frac{b}{\rho}$. Note that this solution does not depend on z — the same reservation price (for immediate acceptance) applies for all $z > 0$.

In light of this, buyers will never offer a price $p < x - \frac{b}{\rho}$ because all buyers will already accept the higher price. In doing so, they exit, so it is impossible for any buyer to have a tie with two such offers. □

A.4 Proposition 3

Proof. The process in the text of constructing the proposed solution from the equilibrium conditions establishes their equivalence. For uniqueness, consider that if $\frac{b}{\rho} \geq x - c$, then the price $R(T)$ is below marginal cost and would result in negative profit, even if $\bar{p}$ is profitable. This precludes $\alpha^* > 0$ in that case.

When $\frac{b}{\rho} < x - c$, recall that $b > d$ implies $\frac{d}{\rho} < x - c$ as well. We will establish uniqueness by showing that $\Pi'(\alpha) > 0$ whenever $\Pi(\alpha) = x - c - \frac{b}{\rho}$. Thus, there can be at most one such α^*.

For simplicity of notation, let $Q \equiv \rho(\mu+\rho)T + \left(\rho + \mu\left(1 - e^{-\rho T}\right)\right) \ln\left(1 - \frac{\rho\left(1-e^{-(1-\alpha)\mu T}\right)}{\rho+\mu(1-e^{-\rho T})}\right)$, so $\Pi(\alpha^*) = \frac{\alpha T\left(\rho+\mu\left(1-e^{-\rho T}\right)\right)(\alpha\mu(\rho(x-c)-b)+\rho(\rho(x-c)-d))}{\rho\left(\alpha+e^{\mu T}-e^{\mu(1-\alpha)T}\right)Q}$.

Moreover, when an early dispersed equilibrium occurs, $\Pi(\alpha^*) = x - c - \frac{b}{\rho}$ rearranges as:

$$(\alpha\mu(\rho(x-c)-b) + \rho(\rho(x-c)-d)) = \frac{\rho Q\left(e^{\mu T} - e^{(1-\alpha)\mu T} + \alpha\right)}{\alpha T\left(\rho + \mu\left(1 - e^{-\rho T}\right)\right)}\left(x - c - \frac{b}{\rho}\right).$$

We can then take the derivative $\Pi'(\alpha^*)$ and use the preceding equation to substitute for the left-hand terms where they appear in that derivative. This yields:

$$\begin{aligned}\Pi'(\alpha^*) = \frac{Te^{-\alpha\mu T}\left(\rho + \mu\left(1 - e^{-\rho T}\right)\right)}{Q\left(\alpha + e^{\mu T} - e^{\mu(1-\alpha)T}\right)^2}\Bigg(& e^{\mu T}\left(e^{\alpha\mu T} - \alpha\mu T - 1\right)(\rho(x-c)-d) + \\ & \mu\Bigg(\frac{\alpha\left(\alpha e^{\alpha\mu T} + 2e^{(\alpha+1)\mu T} - e^{\mu T}(\alpha\mu T + 2)\right)}{\rho} + \\ & \frac{\left(\alpha e^{\alpha\mu T} + e^{(\alpha+1)\mu T} - e^{\mu T}\right)^2}{\rho e^{\alpha\mu T} + \mu e^{\mu T}\left(1 - e^{-\rho T}\right)}\Bigg)(\rho(x-c)-b)\Bigg).\end{aligned}$$

Note that $\rho(x-c)-b > 0$ in this case because profits must be positive, and so $\rho(x-c)-d > 0$. Thus, for $\bar{p} > c$ (and thus enable positive profits), it must be that $Q > 0$.

In the first line, $e^{\alpha\mu T} - \alpha\mu T - 1$ equals 0 when $\alpha = 0$, and is increasing in α, so is therefore positive for all $\alpha \in [0,1]$.

The same applies to $\alpha e^{\alpha\mu T} + 2e^{(\alpha+1)\mu T} - e^{\mu T}(\alpha\mu T + 2)$ in the second line. It is 0 at $\alpha = 0$ and its derivative $\mu T e^{\mu T}\left(2e^{\alpha\mu T} - 1\right) + e^{\alpha\mu T}(\alpha\mu T + 1) > 0$.

In the third line, the numerator of the fractional term is squared, ensuring the term is always positive.

Hence, $\Pi'(\alpha) > 0$ wherever $\Pi(\alpha) = x - c - \frac{b}{\rho}$. This ensures it can have at most one zero.

Also note that if a late dispersed equilibrium occurs, then $\Pi(0) > x - c - \frac{b}{\rho}$, and it is not possible for an $\alpha > 0$ to exist where $\Pi(\alpha) = x - c - \frac{b}{\rho}$; otherwise, there would need to be some $\hat{\alpha}$ where $\Pi(\hat{\alpha}) = x - c - \frac{b}{\rho}$ and $\Pi'(\hat{\alpha}) < 0$, which contradicts. A similar argument applies when $\Pi(1) < x - c - \frac{b}{\rho}$, ruling out any equilibrium for $\alpha < 1$.

Thus, there is always a unique equilibrium α^*. $\square$

References

Akin, S. Nuray, and B. Platt. (2012) "Running Out of Time: Limited Unemployment Benefits and Reservation Wages," *Review of Economic Dynamics*, **15**, 149-170.

Akin, S. Nuray, V. Lambson, G. McQueen, B. Platt, B. Slade, and J. Wood. (2013) "Rushing to Overpay: Modeling and Measuring the REIT Premium," *Journal of Real Estate Finance and Economics*, **47**, 506-537.

Akin, S. Nuray, and B. Platt. (2014) "A Theory of Search with Deadlines and Uncertain Recall," *Economic Theory*, **55**, 101-133.

Armstrong, M, and J. Zhou. (2016). "Search Deterrence," *Review of Economics Studies* **83**, 26-57.

Burdett, Kenneth, and K. Judd. (1983) "Equilibrium Price Dispersion," *Econometrica* **51**, 955-970.

Butters, Gerard. (1977) "Equilibrium Distribution of Sales and Advertising Prices," *Review of Economics Studies* **44**, 465-491.

Choi, Michael, A. Y. Dai, K. Kim. (2018) "Consumer Search and Price Competition," *Econometrica* **86**, 1257-1281.

Choi, M. and L. Smith. (2022) "Optimal Sequential Search with Recall," working paper, University of Wisconsin.

Coey, Dominic, B. Larsen, and B. Platt. (2020) "Discounts and Deadlines in Consumer Search," *American Economic Review* **110**, 3748-85.

Curtis, E., and R. Wright. (2004) "Price setting, price dispersion, and the value of money: or, the law of two prices," *Journal of Monetary Economics*, **51**, 1599-1621.

Diamond, Peter. (1971) "A Model of Price Adjustment," *Journal of Economic Theory* **3**, 156-168.

Diamond, Peter. (1987) "Consumer Differences and Prices in a Search Model," *Quarterly Journal of Economics* **102**, 429-436.

Janssen, Maarten, and J. L. Moraga-Gonzalez. (2004). "Strategic Pricing, Consumer Search, and the Number of Firms," *Review of Economics Studies* **71**, 1089-1118.

Janssen, Maarten, and A. Parakhonyak. (2014) "Consumer Search Markets with Costly Revisits," *Economic Theory* **55**, 481-514.

Janssen, Maarten, P. Pichler, and S. Weidenholz. (2011) "Oligopolistic markets with sequential search and production cost uncertainty," *RAND Journal of Economics* **42**, 444-470.

Lemieux, James, and R. Peterson. (2011) "Purchase Deadline as a Moderator of the Effects of Price Uncertainty on Search Duration," *Journal of Economic Psychology* **32**, 33-44.

Reinganum, Jennifer. (1979) "A Simple Model of Equilibrium Price Dispersion," *Journal of Political Economy* **87**, 851-858.

Rob, Rafael. (1985) "Equilibrium Price Distributions," *Review of Economic Studies* **52**, 487-504.

Rogerson, Richard, R. Shimer, and R. Wright. (2005) "Search-Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature* **43**, 959-988.

Salop, Steven, and J. Stiglitz. (1976) "Bargains and Ripoffs: A Model of Monopolistically Competitive Prices," *Review of Economic Studies* **44**, 493-510.

Stahl, Dale O. (1989) "Monopolistic Pricing with Sequential Consumer Search," *American Economic Review* **79**, 700-712.

Stigler, George. (1961) "The Economics of Information," *Journal of Political Economy* **69**, 213-225.

van den Berg, Gerard. (1990) "Nonstationarity in Job Search Theory," *Review of Economic Studies* **57**, 255-277.

Wilde, Louis, and A. Schwartz. (1979) "Equilibrium Comparison Shopping," *Review of Economic Studies* **46**, 543-554.