Competitive Hold-Up: Monopoly Prices Too High to Maximize Profits with Perfectly Competitive Retailers or Complements

January 15, 2017

Eric Rasmusen

Abstract

If a monopolist cannot commit to a wholesale price in advance, even competitive retailers will be reluctant to enter the market, knowing that once they have entered, the monopolist has incentive to choose a higher price and reduce their quasi-rents. Even if retailers earn zero profits in long-run, this hurts the monopolist by shifting in the retailer supply curve. By “competitive hold-up” I mean this temptation of a monopolist to raise his price in reliance on the sunkness of perfectly competitive firms’ investments. A similar problem occurs if the monopolist’s product is sold directly to consumers but is complementary to a product sold in a competitive industry with U-shaped cost curves. Competitive hold-up arises upstream opportunism, not downstream market power, and so is distinct from two problems that look superficially similar, double marginalization and the two-monopoly complements externality.


This paper: http://www.rasmusen.org/papers/holdup-rasmusen.pdf.

Keywords: Monopoly, holdup-costs, opportunism, competitive hold-up.
I would like to thank Louis Kaplow, Qihong Liu, Michael Rauh and the BEPP Brown Bag Lunch for helpful comments and Bryan Stuart for research assistance.
1. Introduction

It is well-known that when two firms with market power sell related products, their prices may be too high for their own joint profit maximization, not just for the maximization of social surplus. One problem is the positive externality from complementary goods. If sales of each monopolist’s product increases sales of the other’s product, that externality will be ignored under individual profit maximization. The two firms would both have higher profits if they simultaneously reduced their prices. This idea goes back to Cournot (1838). More recent references analyses can be found at Economides & Salop (1992), Feinberg & Kamien (2001), Dari-Mattiacci & Parisi (2006), and Spulber (2016). A second problem is double marginalization, which dates back to Spengler(1950) or, in more modern form, Greenhut & Ohta (1979) or Janssen & Shelegia (2015), and is surveyed with related problems in Lafontaine & Slade (2007). If an upstream monopolist sells to a downstream monopolist who sells to consumers, they each add their own profit margin to the price they charge. Both would be better off if the upstream monopolist reduced the wholesale price in return for the downstream monopolist reducing the retail price, thus increasing the quantity ultimately sold to consumers. This can be viewed as a special case of complementary goods, with the upstream product and downstream retail services being perfect complements. Double marginalization applies, however, even if the downstream firm provides zero value-added and is just a bottleneck.

Whether the problem is double marginalization or complements, the solutions are similar. The two monopolies can be merged, to eliminate the externality. Or, depending on what actions the parties can commit to in advance, some kind of contract such as retail price maintenance can be designed to agree on mutually beneficial prices or outputs, the kind of pricing agreement that normally violates antitrust laws but in this case benefits consumers as well as producers. The problems and their solutions are prominent examples of how apparently abstruse economic theory can be useful in practice; the ideas are a standard
part of courses on managerial economics (see, for example, chapter 11 of Baye & Prince [2014]).

The same problems do not arise when only one good is monopolized, however. If an upstream monopolist sells to competitive downstream retailers who sell to consumers, the downstream retailers are not adding any second profit margin. If one complement is monopolized and the other is produced under perfect competition, the perfectly competitive product will be priced at cost, so there is no scope for mutually beneficial price reductions.

This paper points out a problem which appears similar to double marginalization and joint complements but is present even when the interaction is between a monopolist and a competitive market. This problem, which I will call “competitive hold-up”, arises because of the monopolist’s temptation to choose his price so as to take advantage of the competitive firms’ sunk costs and increase his own profits at the expense of their quasi-rents. The result of competitive holdup is that the monopolist chooses a price higher than he would if he were merged together with the competitive industry, even though in equilibrium that industry earns zero profits.

Competitive firms do not earn monopoly rents, revenues in excess of their long-run costs. In the short run, however, when the number of firms is fixed, if the market price rises unexpectedly then even when the firms take the price as given and produce up to where their marginal costs equals the price, it is quite possible for their economic profits to be positive. These profits are called “quasi-rents” because they do not persist. In the long run more firms enter, the market price falls, and profits return to zero. Even in long-run equilibrium, however, firms earn quasi-rents. Their revenue still exceeds their short-run costs and they are earning strictly more than their opportunity cost, which excludes costs that are sunk. Their quasi-rents exactly equal their sunk costs, however, which is why the industry can be in long-run equilibrium.
What we will see happening in competitive holdup is that the monopolist will try to take advantage of the short-run fixedness of the number of competitive firms. For a given size of competitive industry, it can set its price high enough that the competitive industry’s quasi-rents are too small to pay back its sunk costs. Competitive holdup is self-defeating, however, because it shrinks the competitive industry in the long run. Fewer firms will enter the competitive industry than would if the monopolist could commit not to use the tactic. The result is a competitive industry that cannot produce as cheaply on as large a scale, to the monopolist’s detriment.

This is, of course, much like the standard hold-up problem, just set in the unexpected context of a competitive industry. Hold-up costs are well understood in the context of contracting with one buyer and one seller. The buyer may refrain from undertaking value-increasing investments for fear that the seller will raise the price once the investment is a sunk cost. Williamson (1975) proposed this as a major force in explaining interactions in industrial organization, and Hart & Moore (1988) provided a formalization which showed the extent of underinvestment. Long-term contracts with various kinds of special clauses are one solution, as explored in Noldeke, Georg and Schmidt (1995) (option clauses), Edlin & Reichelstein (1996) (damage clauses) and others. The problem shows up in empirical work as well, most famously described in Joskow (1987) on the case of specific investments in coal use by electrical plants. In the present paper, the specific investment will be the fixed cost in the textbook model of U-shaped cost curves, but since marginal cost will be rising, not constant, the market will be able to exist without bilateral contracting, which is the focus of the hold-up literature.

What has been less explored is what happens when the market has many buyers—too many for contracting, or for tailoring the price to whether or not the buyer has made the investment. Hold-up can be as much a problem for impersonal markets as for bilateral contracting. The paper models this first when the monopoly sells to retailers who
must incur entry costs to sell the product and second when the monop-
oly sells directly to consumers but the good is complementary with a
good sold in a competitive market in which firms incur sunk costs.

The first model is of potential retailers who decide whether to
enter into retailing the monopolist’s product when they are worried
that after they enter he will raise his wholesale price. In this model,
the monopolist sells his product via price-taking retailers who engage
in classic perfect competition, each being infinitesimal relative to the
market with U-shaped cost curves because they have a fixed entry
cost plus rising marginal cost of providing retail services (besides the
constant marginal cost of the wholesale price of the product). In equi-
librium, the monopolist sets too high a wholesale price to maximize
his profits, much less to maximize social surplus. The reason is that
he is free to change the price after the retailers have entered, taking
away some of the quasi-rents that compensate them for their sunk en-
try costs. Foreseeing this, too few retailers enter, and the marginal cost
of the retailing services is higher than the efficient level, reducing the
monopolist’s profits (and the retail price is higher, exacerbating the
conventional triangle loss from reduced gains from trade).

This may look like double marginalization, but it is not. Dou-
ble marginalization is the problem of successive monopolies, where the
monopoly markups of the wholesaler and the retailer result in a price
higher than the price a vertically integrated monopolist would charge.
Here, the retailers have no market power and would earn zero profits
in any equilibrium. Unlike in double marginalization, two-part tariffs
would not solve the monopolist’s problem, but commitment to a one-
part wholesale price would. As with double marginalization, however,
vertical integration would solve the problem, reducing the wholesale
and retail prices and increasing the combined profits of the sellers.

In the second model, the monopolist sells directly to consumers,
but his good is complementary to a good sold in a competitive market.
This is not the usual problem of overpricing of complements. It is
well known that if two firms that sell complementary goods each have
market power, they can increase their profits by merging and reducing their prices. The problem they have is that they both mark up their prices above marginal cost, and when one of them increases his mark-up, he captures the entire gain but inflicts a negative externality on the other monopolist. If one of the markets is competitive, we will see that the same problem of high prices arises for quite a different reason. In the competitive market, firms are charging a price equal to marginal cost, and price-taking to boot, so there is no margin over cost to adjust and make too high— the profit margin is zero. Rather, the problem arises from the competitive firms’ quasi-rents. The monopoly would like to commit to a low price so as to increase the size of the competitive market and enable more firms to enter and earn back their entry cost via the quasi-rent from upward sloping marginal cost. If he cannot commit, however, potential entrants know he will charge a high price in order to weaken their demand and cause the market price to fall in the competitive industry. Not so many of them enter, which shifts in the supply curve for the competitive good to the detriment of demand for the monopolist’s complementary good.

Competitive holdup is distinct from a different opportunism problem for a monopolist facing competitive retailers, what we might call “secret quantity expansion”: after publicly agreeing to bilateral contracts with each retailer that would result in a certain total retail market quantity and price, the monopolist might secretly sell more to some retailers. This would “cheat” the other retailers who paid a wholesale price based on their belief that the retail price would be high because of the limited market quantity. Foreseeing this opportunism, the retailers will not accept a high price in the first place, dealing a serious blow to the monopolist’s profits, a problem reminiscent of the Coase Conjecture’s competition between a monopolist present and future selves. This idea of secret quantity expansion can be found in Hart & Tirole (1990) in their discussion of vertical integration with two upstream and two downstream firms with identical goods, but it was more clearly modelled as the central idea in the O’Brien & Shaffer (1992) model of a monopolist facing differentiated retailers. McAfee & Schwartz (1994)
showed that nondiscrimination clauses would not solve the problem, and Rey & Verge (2004) add close attention to the out-of-equilibrium beliefs of the retailers. Resale price maintenance does work as a solution, and Montez (2015) shows how product buybacks can serve a similar function.

Secret quantity expansion will not arise in the competitive holdup model; the opportunism is different. We will be assuming that the wholesale price is public and there is no bilateral contracting, but that is not the important difference. Rather, it is that we will assume upward sloping marginal cost for retailer services, which means that if the monopolist offers a greater quantity to a particular retailer, that retailer will have higher marginal cost and hence weaker demand than the others. Thus, the monopolist will not be tempted to offer secret discounts. If marginal costs were constant, on the other hand, and the monopolist had to use a per-unit price rather than a two-part tariff, that price would be above marginal cost, making retailers individually eager to be allowed to sell larger quantities and creating the temptation for secret quantity expansion.

Two other kinds of models of vertical relations also should be distinguished from competitive holdup. First is the opportunism problem created by unobservable actions. Telser (1960) pointed out that if retailers provide unobservable consumer services with positive externalities, the monopolist can use resale price maintenance to encourage competition in providing those services and discourage opportunistic low-service discount outlets (see, too, the discussion in the survey by Lafontaine & Slade [2007]). Here, neither monopolist nor retailer will be taking unobserved actions. Second, there is a large literature on common agency, the idea that upstream oligopolists can use contracts with retailers to soften competition between them. This literature began with Bernheim & Whinston (1985) and Bonnano & Vickers (1988); a recent example is Hunold & Stahl (2016). Relaxing competition will be irrelevant in competitive holdup, however, since the model will contain only a single monopolist and retailers will be price-takers.
2. The Retailer Model

A monopolist produces a good at a constant marginal cost $a$ in a market whose consumers demand quantity $Q_d(p)$ at retail price $p$. We will assume that $2Q'_d + (p - a)Q''_d < 0$ so monopoly profit will be concave in price.\(^1\) The monopolist sells quantity $q(t)$ at unit wholesale price $w$ to retailer $t$, who resells it to consumers at price $p$, incurring marginal costs of $c(q(t))$ to do so, where $c$ is nondecreasing. Each retailer must also incur a fixed cost of $F$. Retailers are infinitesimal, following Aumann (1964) (or, more simply, Peck [2001]). The amount of retailers is $n$, so their output is $\int_0^n q(t) dt$ and the amount of fixed cost is $nF$. We will hereafter suppress the $t$ argument, since in equilibrium all active retailers will choose the same output. Assume that consumer demand is positive if the price is at least a retailer’s minimum average cost, i.e. $Q_d(c_{min}) > 0$ where $c_{min} = \arg\min_q \frac{-F + \int_0^q c(x) dx + qw}{q}$.

First, take the quantity $n$ of retailers and the wholesale price $w$ as given. The profits of an individual retailer are:

$$\pi_{retailer} = pq - \left(F + \int_0^q c(x) dx + qw\right) \quad (1)$$

Maximizing by choice of output, the retailer’s first order condition yields the conventional result of equilibrium output, $q^*$, being such that price equals marginal cost:

$$p = c(q^*) + w \quad (2)$$

This yields the short-run individual supply curve

$$q(p) = c^{-1}(p - w) \quad (3)$$

\(^1\)This is a version of the Novshek Condition of Novshek (1985) and Gaudet & Salant (1991), which is commonly used in imperfect competition models to ensure that the monopoly pricing problem is convex. Without it, the optimum choices might not change with the parameters, and we would have to qualify our results with phrases like “will fall or remain unchanged.”
and the market supply curve
\[ Q^*(p) = nc^{-1}(p - w). \]  

The market must clear, so quantity demanded must equal quantity supplied. Since all \( n \) retailers are identical, their first order conditions and equilibrium outputs are all the same, so we can write:
\[ Q^*_d(p^*) = nq^*, \]  
or, using equation (2),
\[ Q^*_d(c(q^*) + w) = nq^*. \]  

The short-run equilibrium output of a firm, \( q^*(n, w) \) (as distinguished from the firm’s short-run supply curve \( q(p) \)) is therefore
\[ q^*(w, n) = \frac{Q^*_d(c(q^*) + w)}{n}. \]  

**Lemma 1:** Individual retailers supply less if the number of retailers \( n \) or the wholesale price \( w \) is greater. The market equilibrium quantity, \( Q^*(n, w) \), increases as \( n \) increases.
\[ \frac{\partial q^*}{\partial n} < 0, \quad \frac{\partial q^*}{\partial w} < 0, \quad \frac{\partial Q^*}{\partial n} > 0. \]

**Proof.** Differentiating equation (7) yields the individual retailer comparative statics we need for \( n \) and \( w \). Differentiate with respect to \( n \) holding \( w \) constant to get:
\[ \frac{\partial q^*}{\partial n} = \frac{Q'_d c'(q^*) \frac{\partial q^*}{\partial n}}{n} - \frac{q^*}{n^2} \]  
so
\[ \frac{\partial q^*}{\partial n} \left( \frac{Q'_d c'(q^*)}{n} - 1 \right) = \frac{q^*}{n^2} \]  
and
\[ \frac{\partial q^*}{\partial n} = \frac{q^*}{n^2} \left( \frac{n}{Q'_d c'(q^*) - n} \right) \]  

Since \( Q'_d < 0 \), we can conclude that \( \frac{\partial q^*}{\partial n} < 0. \)
Second, differentiating equation (7) with respect to $w$ keeping $n$ constant yields

$$\frac{\partial q^*}{\partial w} = \frac{Q_d'c'(q^*)}{n} \frac{\partial q^*}{\partial q} + \frac{Q_d'}{n}$$

so

$$\frac{\partial q^*}{\partial w} \left( \frac{Q_d'c'(q^*)}{n} - 1 \right) = -\frac{Q_d'}{n}$$

and

$$\frac{\partial q^*}{\partial w} = -\frac{Q_d'}{Q_d'c'(q^*) - n}$$

Since $Q_d' < 0$, we can conclude that $\frac{\partial q^*}{\partial w} < 0$.

Third, the retailer market supply curve, equation (4), tells us that as $n$ increases, for a given price $p$ it follows that market quantity supplied will increase too. The increase in quantity will cause $p$ to fall since the demand function $Q_d(p)$ is downward sloping, but since both demand and supply functions are continuous, output will equilibrate at some level greater than the initial level, and $\frac{dQ^*}{dn} > 0$. ■

We can now consider the long-run equilibrium, in which the quantity $n$ of firms is determined. Since there is free entry and perfect foresight, long-run retailer profits equal zero:

$$\pi_{retailer} = pq - \left( F + \int_0^q c(x)dx + qw \right) = 0$$

(14)

We can solve for the retail price, which must equal average cost:

$$p = \text{average cost} = \frac{F}{q} + \frac{\int_0^q c(x)dx}{q} + w.$$

(15)

Since retailer profits are zero and retailers are identical, in equilibrium they will all choose sales to be the $q$ that minimizes average cost, the minimum of the U-shaped average cost curve. That value, which
we will call $q^*$, is the value that minimizes the average cost in equation \((15)\), solving the first order condition,

\[
-\frac{F}{(q^*)^2} - \int_0^{q^*} c(x) dx + \frac{c(q^*)}{q^*} = 0.
\]  \hspace{1cm} (16)

Note that the wholesale price $w$ does not appear in \((16)\); the long-run equilibrium scale $q^*$ does not depend on the wholesale price. If $w$ rises, retailers end up exactly the same size in long-run equilibrium. Equation (2) says that $p = c(q) + w$, so since output per firm is fixed at $q^*$, independently of $w$, it follows that $\frac{dp}{dw} > 0$. Since market demand slopes down ($Q'_d < 0$) and since quantity supplied is $nq^*$ in equilibrium it follows that $\frac{dn}{dw} < 0$, as stated in Lemma 2.

**Lemma 2:** The number of retailers, $n$, is decreasing in the wholesale price, $w$, in the long run:

\[
\frac{dn}{dw} < 0.
\]

When the wholesale price rises, the retail price must rise if the retailers are not to have negative profits, and since each retailer must produce at the cost-minimizing size, the number of retailers must fall.

**The Upstream Monopolist**

The upstream monopolist’s profit is

\[
\pi_{monopoly} = (w - a) \cdot n(w) \cdot q^*(w, n(w)),
\]  \hspace{1cm} (17)

where the equilibrium output function $q^*(w, n(w))$ is a function of the number of firms and the wholesale price that generate the retail price to which the retailers respond. In the long run, sales per retailer are $q^*(w, n(w))$, determined entirely by $w$. In the short run, sales per retailer are $q^*(w, n)$, with $n$ fixed by previous entry.

If the monopolist chooses $w$ after retailers enter, he takes $n$ as given and uses the short-run equilibrium sales function $q^*(w, n)$. Since the retail equilibrium price equals $p^* = c(q^*) + w$, monopoly profit \((17)\)
is concave in $w$ as a result of our assumption that $Q_d(p)$ is concave. The first order condition is

$$\frac{d\pi_{monopoly}}{dw} = nq^* + (w - a)n \frac{\partial q^*}{\partial w} = 0,$$

so

$$w = a - \frac{q^*}{\frac{\partial q^*}{\partial w}}.$$

Thus, the monopolist’s wholesale price equals his cost, $a$, plus (since $\frac{\partial q^*}{\partial w} < 0$) an amount depending on how much wholesale demand falls with the price, which in turn depends on the retail demand.

If the monopolist chooses $w$ before retailers enter, he uses the long-run equilibrium sales function $q^*(w, n(w))$. Since now $n$ is endogenous, we no longer know that monopoly profit is concave, but it is differentiable in $w$ and the optimum will not be zero or infinity under our assumptions, so the first order condition is still a necessary, if not sufficient, condition and we can compare it to the short-run optimum. The first order condition is

$$\frac{d\pi}{dw} = nq^* + (w - a)n \frac{\partial q^*}{\partial w} + (w - a)\frac{\partial}{\partial n}[nq^*(n, w)]\frac{dn}{dw} = 0,$$

so

$$w = a - \frac{q^*}{\frac{\partial q^*}{\partial w} + \frac{1}{n} \frac{\partial Q^*}{\partial n} \frac{dn}{dw}}.$$

Lemma 1 tells us that $\frac{\partial q^*}{\partial w} < 0$ and $\frac{\partial Q^*}{\partial n} > 0$. Lemma 2 tells us that $\frac{dn}{dw} < 0$. Therefore, the boxed term in the denominator is negative, which makes the negative quantity in the denominator bigger in magnitude, which means less is added to $a$ to get the wholesale price. The monopolist who can commit to $w$ therefore chooses a lower value than if he could not commit.

The monopolist who cannot precommit to a wholesale price will choose a higher price and therefore sell less to retailers. Thus, the retailer price will be higher, and consumer surplus lower. Since the value of $n(w)$ will depend on the anticipated $w$ anyway if retailers are rational, the monopolist’s neglect of that term in the short-run
profit maximization problem will reduce his profits. Thus, we have Proposition 1.

**Proposition 1:** If the upstream monopolist cannot precommit to his wholesale price before retailers enter, his price will be higher, sales lower, and profits lower. As consumer surplus will also be lower, total surplus is lower than in a vertically integrated monopoly or a monopoly that can commit to wholesale prices.

Foreseeing that without commitment the upstream monopolist will choose a high wholesale price, retailers will realize that fewer of them must enter or their individual outputs and the difference between the retail and wholesale prices would yield too low a triangle of quasi-rents to make up for the entry cost. When fewer enter, however, the retail market supply curve swivels up, reducing the monopolist’s profits.

Profits would be higher if the monopolist could commit to a lower price—though not if ex post he unexpectedly reduced his price from what retailers expected when they made their entry decisions. The monopoly ignores the zero profit constraint in setting \( w \) if the retailers have already entered, and takes \( n \) to be a constant rather than a function \( n(w) \). He will leave them with some quasi-rents, because with upward sloping marginal cost, the quasi-rents are always positive. Foreseeing the monopolist’s choice of \( w \) based on \( n \), a number \( n(w) \) of firms will enter that will make for zero profits.

**Discussion**

One solution to the problem of competitive holdup is vertical integration. The scale of operation would still be \( q^* \) per retailer, since that is the minimum efficient scale. The monopolist would use the resulting minimum average cost as his marginal cost of providing retailing services, since he could choose the amount \( n \) of retailers. He would choose it to be the same as if he could commit to a wholesale price, since in that case, too, he would leave the retailer with zero profit.
Note that the hold-up problem in the absence of commitment to wholesale prices is not double marginalization, even though both problems result from a monopolist selling downstream through retailers and can be solved by vertical integration. Double marginalization is the inefficiency that results when a monopoly manufacturer sells about marginal cost to a retailer who then sells above his own marginal cost to consumers. The result is higher prices than a vertically integrated monopoly would charge.

Here, however, retailers have no market power, so any mark-up they add is justified by marginal costs. Two-part tariffs, which are a solution to double marginalization, merely exacerbate competitive holdup. If the monopolist can make a take-it-or-leave-it offer to sell amount $q^*$ to a retailer at fixed price $w_1$ plus marginal price $w_2$ per unit, he can seize all the quasi-rents. Foreseeing this, no retailer will enter. The possibility of using a two-part tariff is purely bad for the monopolist. One way to view the problem in double marginalization is that the retailers are making monopoly profits at the expense of the manufacturer, so a solution is to transfer all the bargaining power to the manufacturer. An alternative is to give all the bargaining power to the retailers— the ability to make a take-it-or-leave it offer to buy the good at manufacturer marginal cost would solve the double marginalization problem. In competitive holdup, the monopolist’s bargaining leverage is the source of the problem, so increasing it just makes things worse.

3. A Numerical Example of the Retailer Model

Numerical examples are helpful in understanding ideas and are often more persuasive than proofs, so let us lay one out here. An upstream monopolist produces a good at constant marginal cost $a = 1$ which he will sell at wholesale price $w$ per unit. A continuum of length $n$ of competing retailers with identical cost curves enter at cost $F = 1$. Each retailer chooses to sell $q(p)$ of the good at marginal cost $c(q) + w$, with $c = q$ in this example, so a retailer’s total cost is

$$ .5 + .5q^2 + wq. $$

(22)
Since retailers are infinitesimal, if we index retailers by \( t \) then the retail industry generates a fixed entry cost of \( F_n \), output of \( \int_0^n q(p,t) \, dt \), and variable costs of 
\[
\int_0^n \left[ \int_0^{q(p,t)} (c(q(p,t)) + w) \, dq \right] \, dt.
\] (23)

Thus an individual retailer will have U-shaped cost curves, as in Figure 1.

**Figure 1: A Retailer’s Cost Curves**

Let consumer market demand be linear:
\[
Q(p) = 1000 - 200p
\] (24)

Since retailers compete, and since free entry will make their profits equal to zero, all retailers must produce at exactly the minimum average cost. This is found by minimizing \( \frac{5}{q} + .5q + w \), which yields output of \( q = 1 \) per retailer.

The social marginal cost of serving consumers equals the minimum average cost when the wholesale price is set to the wholesale marginal cost, so \( w = a = 1 \), in which case the average cost is \( \frac{5}{T} + .5(1) + 1 = 2 \). If we set \( p = 2 \), then \( Q = 1000 - 200p = 600 \). Dividing this by \( q = 1 \) yields the quantity of retailers at the social optimum, \( n = 600 \). This will yield zero profits to the monopolist since \( w = a \) and zero profits to the retailers since retailer revenue of \( pnq = (2)(6)(1) = 12 \) equals the
retailer cost of \( n(0.5 + 0.5q^2 + wq) = (6)(0.5 + 0.5(1)^2 + 1(1)) = 12 \). It will yield consumer surplus of \( 0.5(5 - 2)(600) = 900 \).

**Retailer behavior in equilibrium**

A retailer’s profit is

\[
\pi_{\text{retailer}} = pq - \int_0^q c(x)dx - wq - F
\]  

(25)

Maximizing this profit equation by choice of output with respect to quantity yields the familiar rule of price equals marginal cost:

\[
p = c(q) + w
\]  

(26)

Because of free entry, retailer profits equal zero. Setting \( p = c(q) + w \) and substituting our example’s marginal cost function, \( c(q) = q \), revenue equals \( (q + w)q \). Equation (22) gives us the total cost. Thus, the profit function in (25) can be rewritten and set to zero to yield the long-run equilibrium condition

\[
\pi_{\text{retailer}} = (q + w)q - 0.5q^2 - wq - F = 0.
\]  

(27)

The \( wq \) terms cancel, and we have assumed an entry cost of \( F = 0.5 \), so this solves to \( q = 1 \). Substituting that \( q \) into the market equilibrium condition (30) gives us the amount of retailers \( n^* \) as a function of the wholesale price \( w \):

\[
n^* = 800 - 200w.
\]  

(28)

**Figure 2: A Retailer’s Cost Curve Shift when the Wholesale Price Rises**
In the short run, \( n \) is fixed and profits are not necessarily zero. Following the rule of marginal cost equals price, \( c(q) = q = p + w \) so the individual retailer short-run supply curve is

\[
q(p, w) = p - w. \tag{29}
\]

In equilibrium, market supply equals market demand, so

\[
Q_s = nq = Q_d = 1000 - 200p \tag{30}
\]

Solving equation (30) for \( p \) and substituting into (29) yields \( q = 5 - n/200 - w \), which, solving for \( q \), leaves us with an expression for the individual retailer’s output as a function of \( n \) and \( w \) when \( p \) takes its resulting short-run equilibrium value:

\[
q(n, w) = \frac{5 - w}{1 + .005n}. \tag{31}
\]

As Figure 3 shows, the market supply curve swivels down as the number of retailers \( n \) increases, since to increase market supply requires moving less far up the individual retailer’s marginal cost curves.

**Figure 3: How Market Supply Changes with More Retailers**
Monopolist behavior with commitment to a wholesale price

The monopolist’s profits equal the wholesale price minus the wholesale production cost times output. Since the monopolist can commit to the price before retailers choose whether to enter, he takes into account his effect on $n$ and uses the retailer’s long-run equilibrium condition (28) in choosing $w$:

$$
\pi_{\text{monopolist}} = (w - a)qn = (w - 1)(1)(800 - 200w) = 1000w - 800 - 200w^2. \tag{32}
$$

Setting the derivative with respect to $w$ equal to 0 and solving for $w$ yields

$$
w = 2.5. \tag{33}
$$

Using this $w$, in equilibrium $p = 3.5, n = 300, q = 1$ and $Q = 300$. The monopolist earns $\pi_{\text{monopolist}} = (w - a)Q = 450$, compared to 0 at the social optimum. Consumer surplus falls from 900 to $.5(5-3.5)(300) = 225.$

Monopolist behavior without commitment to a wholesale price

Now suppose the monopolist does not choose $w$ until after the retailers have made their entry decisions. The monopolist takes $n$ as
given when he chooses \( w \), and uses the retailers’ short-run supply function (31) for his expectation as to the level of \( q \) they will choose. Note that the monopolist’s choice of \( w \) will result in the retailer’s zero-profit entry condition being satisfied if retailers are not myopic, but not otherwise. Thus,

\[
\pi_{\text{monopolist}} = (w - a) n q(n, w) \\
= (w - 1) n \left( \frac{5 - w}{1 + .005n} \right) \\
= \left( \frac{n}{1 + .005n} \right) (5w - 5 - w^2 + w)
\]  

(34)

The first order condition for choice of \( w \) is

\[
\frac{d\pi_{\text{monopolist}}}{dw} = \left( \frac{n}{1 + .005n} \right) (6 - 2w) = 0
\]  

(35)

Thus, the monopolist will choose a wholesale price of \( w = 3 \), regardless of the level of \( n \).

The actual level of \( n \) will anticipate the monopolist’s decision. We know that to achieve zero profit, a retailer must be operating at minimum efficient scale of \( q = 1 \) with marginal cost of 1, so \( p = w + 1 = 4 \). The demand curve tells us that \( Q_d = 1000 - 200p = 200 \), which when \( q = 1 \) means that \( n = 200 \) also. The monopoly profit is \( \pi_{\text{monopolist}} = (w - a)Q = 400 \), and consumer surplus equals \( .5(5 - 4)(200) = 100 \).

**The case when the monopolist cannot commit to \( w \) but retailers are deceived.**

It is helpful to see what happens if the monopolist is able to get away with lying to retailers, promising them the commitment wholesale price of \( w = 2.5 \) but actually being free to change to \( w = 3 \). As we have seen, if they believe that \( w = 2.5 \) then amount \( n = 300 \) of retailers will enter, expecting zero profits. Now, however, the monopolist will rely on their short-run supply curve and use first-order condition (35),

\[2\]It is a general property of linear demand and constant marginal cost that the monopoly price is the same for all demand curves with the same \( P(0) \) intercept. That property is not important to the conclusion that without commitment the monopolist will choose too high a price.
which yields \( w = 3 \). The retailers will use their short-run supply curve (31) and choose

\[
q(n, w) = \frac{5 - w}{1 + .5n} = \frac{5 - 3}{1 + 1.5} = .8,
\]

so output will be \( Q = (300)(.8) = 240 \) and \( p = 5 - 2.4/2 = 3.8 \). The monopoly’s profit will be \( (w - a)Q = 480 \), and consumer surplus will be \(.5(5 - 3.8)(240) = 144 \). Retailer profits will now be negative, equalling

\[
pQ - n(.5q^2 + wq + F) = 912 - 300(.5 \times .8^2 + 3 \times .8 + .5) = -54 \quad (37)
\]

Note that industry profit is higher if the retailers are deceived than if they are not—a total of 426 instead of 400.

Table 1 summarizes the various outcomes. When the monopolist cannot commit to a wholesale price, he sets the wholesale price higher, fewer retailers enter, retail prices are higher, and monopoly profits fall from 450 to 400. If he can promise a price of 2.5 and fool the retailers into trusting him, then he would raise the price to 3 and earn profits of 480.

**TABLE 1:**
**Equilibrium Values in the Numerical Example**

<table>
<thead>
<tr>
<th></th>
<th>Monopoly with commitment</th>
<th>Monopoly without commitment</th>
<th>Monopoly with deception</th>
<th>Social Optimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wholesale price, ( w )</td>
<td>2.5</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Retail price, ( p )</td>
<td>3.5</td>
<td>4</td>
<td>3.8</td>
<td>2</td>
</tr>
<tr>
<td>Amount of retailers, ( n )</td>
<td>300</td>
<td>200</td>
<td>300</td>
<td>600</td>
</tr>
<tr>
<td>Output per retailer, ( q )</td>
<td>1</td>
<td>1</td>
<td>.8</td>
<td>1</td>
</tr>
<tr>
<td>Monopolist profit</td>
<td>450</td>
<td>400</td>
<td>480</td>
<td>0</td>
</tr>
<tr>
<td>Retailer profit</td>
<td>0</td>
<td>0</td>
<td>-54</td>
<td>0</td>
</tr>
<tr>
<td>Consumer surplus</td>
<td>225</td>
<td>100</td>
<td>144</td>
<td>900</td>
</tr>
<tr>
<td>Total surplus</td>
<td>675</td>
<td>500</td>
<td>570</td>
<td>900</td>
</tr>
</tbody>
</table>
Figure 4: Monopoly Pricing and Demand by Retailers and Consumers

Note that for a given level of total output $Q$ there is no inefficiency in production as the result of hold-up: retailers operate at the efficient scale, $q^*$ and the only problem is that there are too few of them. If the retailers were, instead, surprised by the high wholesale price, then too many would enter and they would end up each producing at too small a scale. They foresee the high wholesale price, though, so just enough enter to earn zero profit at the minimum efficient scale. On the other hand, the level of total output is the result of an industry which is inefficiently small given the level of demand. If the industry tried to produce the first-best output using the equilibrium short-run supply curve, the cost would be inefficiently high, despite minimizing the cost of producing the second-best output.

4. A Model of a Competitive Complement to a Monopolized Good

The hold-up problem of a monopolist who sells through competitive retailers is very similar to the hold-up problem of a monopolist who sells a good directly to consumers that is a complement for a second good sold by competitive retailers. After all, the retail services the retailer provides can be seen as a separate good that is a complement
to the monopolist’s good. The analysis is essentially the same, but the intuition is less clear and the setting is different enough for an explicit model to be useful.

There are two goods, the monopolized good and the competitive good. A monopolist produces the monopolized good at constant marginal cost \( a \). Infinitesimal price-taking firms indexed by \( t \) produce quantity \( y(t) \) of the competitive good by incurring a sunk fixed cost \( F \) and a nondecreasing marginal cost of \( c(q(t)) \). Firms are infinitesimal, so if the amount of retailers is \( n \), their output is \( \int_0^n y(t) \, dt \). As in the retailer model, we will suppress the \( t \) argument, since in equilibrium all active firms will choose the same output.

The monopolized good with demand \( Q_d(w, r) \) at price \( w \) and a competitive good with demand \( Y_d(r, w) \) at price \( r \). The two goods are complements, so \( \partial Q_d / \partial r < 0 \) and \( \partial Y_d / \partial w < 0 \), but quantity demanded is affected more by a good’s own price, so \( \partial Q_d / \partial w > \partial Q_d / \partial r < 0 \) and \( \partial Y_d / \partial w < \partial Y_d / \partial r < 0 \). We will make the conventional assumption that \( 2 \partial Q_d / \partial w + (w - a) \partial^2 Q_d / \partial w^2 < 0. \)

**The Competitive Firm’s Problem**

First, let us take the quantity \( n \) of firms and the monopoly price \( w \) as given. The profits of an individual firm are

\[
\pi_{\text{firm}} = ry - \left( F + \int_0^y c(x) \, dx \right)
\]

Maximizing by choice of output, the retailer’s first order condition yields the conventional result that price equals marginal cost at the optimal output:

\[
r = c(y),
\]

which gives us the short-run individual supply curve,

\[
y(r) = c^{-1}(r)
\]

\(^3\)As in the retailers model, this is a version of the Novshek Condition of Novshek (1985) and Gaudet & Salant (1991) which is commonly used in imperfect competition models to ensure that the monopoly pricing problem is convex.
and a short-run market supply curve
\[ Y_s(r, n) = nc^{-1}(r). \] (41)

The market must clear, so quantity demanded must equal quantity supplied:
\[ Y_d(r, w) = ny(r). \] (42)

We can rewrite this as
\[ Y_d(c(y), w) = ny(c(y)). \] (43)

The short-run equilibrium output of a firm \( y^*(n, w) \), as distinguished from the firm’s short-run supply curve \( y(r) \), is the \( y \) that solves the preceding equation:
\[ y^*(w, n) = \frac{Y_d(w, c(y^*))}{n} \] (44)

**Lemma 3:** In the short run, if \( w \) rises then \( r \) falls, but at rate less than unity:
\[-1 < \frac{dr}{dw} < 0\]

**Proof.** Totally differentiate the short run market supply equation (42) with respect to \( w \), keeping \( n \) fixed and recognizing that the equilibrium value of \( r \) is a function \( r(w) \):
\[ \frac{\partial Y_d}{\partial w} + \frac{\partial Y_d}{\partial r} \frac{dr}{dw} = n \frac{dy}{dr} \frac{dr}{dw} \] (45)

Rearranging, we get
\[ \frac{dr}{dw} = \frac{\frac{\partial Y_d}{\partial w}}{n \frac{dy}{dr} - \frac{\partial Y_d}{\partial r}} \] (46)

The numerator is \( \frac{\partial Y_d}{\partial w} \), which is negative by assumption. The denominator is composed of two positive terms, since \( \frac{dy}{dr} \) is positive and \( \frac{\partial Y_d}{\partial r} \) is negative by assumption. Since, in addition, we assumed direct demand effects are bigger than cross-effects, we have \( |\frac{\partial Y_d}{\partial w}| > |\frac{\partial Y_d}{\partial r}| \) and we can conclude that \(-1 < \frac{dr}{dw} < 0\).
We can now consider the long-run equilibrium, in which the quantity $n$ of firms is determined. Since there is free entry and perfect foresight, long-run competitive profits equal zero:

$$\pi_{firm} = ry - \left( F + \int_{0}^{y} c(x) dx \right) = 0. \quad (47)$$

We can solve this equation for the long-run price of the competitive good, which must equal the minimum average cost. The average cost is

$$\text{Average cost} = \frac{F}{y} + \frac{\int_{0}^{y} c(x) dx}{y}. \quad (48)$$

Since retailer profits are zero and retailers are identical, in equilibrium they will all choose sales $y$ to be the value that minimizes average cost, the minimum of the U-shaped average cost curve. That value of $y$ that minimizes the average cost in equation (15) solves the first order condition,

$$-F \frac{c''}{(y)^2} - \frac{\int_{0}^{y} c(x) dx}{(y)^2} + \frac{c(y)}{y} = 0. \quad (49)$$

Note that the price of the complementary monopolized good does not appear in (49); the long-run equilibrium scale does not depend on anything on the demand side. If the monopolist chooses a higher price $w$, retailers end up exactly the same size in long-run equilibrium. Moreover, the long-run equilibrium price $r^*$ is independent of $w$ and of $n$. It is determined entirely on the supply side, by the minimum average cost, at the efficient scale. The effect of increasing the price of the monopolized good is just to reduce the number of competitive firms and industry output of the competitive good.

**The Monopolist’s Problem**

The monopolist’s profit is

$$\pi_{monopolist} = (w - a)Q_d(w, r(w)) \quad (50)$$
In the long run, \( r \) equals \( r^* \), the average cost at the efficient scale, and so does not vary with \( w \). In the short run, \( r \) falls as \( w \) increases, by Lemma 3.

It will be convenient here, unlike in the retailers model, to start with the case of the monopolist who can commit to \( w \). When the monopolist chooses \( w \) before retailers enter, he knows he will not affect the competitive good’s price, which will equal \( r^* \). His first order condition is

\[
\frac{\partial \pi_{\text{monopoly}}}{\partial w} = Q_d(w, r) + (w - a) \frac{\partial Q_d(w, r)}{\partial w} = 0, \tag{51}
\]

so

\[
w = a - \frac{Q_d(w, r)}{\frac{\partial Q_d(w, r)}{\partial w}} \tag{52}
\]

Since \( \frac{\partial Q_d(w, r(w))}{\partial w} < 0 \) from the assumption that the demand curve sloped down, this means that the monopolist’s wholesale price equals his cost, \( a \), plus an amount depending on the price sensitivity of demand.

Without commitment, in the short-run, when the monopolist chooses \( w \) after retailers enter, his choice will affect \( r \). His first order condition is

\[
\frac{d\pi_{\text{monopoly}}}{dw} = Q_d(w, r(w)) + (w - a) \frac{\partial Q_d(w, r(w))}{\partial w} + (w - a) \frac{\partial Q_d(w, r(w))}{\partial r} \frac{dr}{dw} = 0, \tag{53}
\]

so

\[
w = a - \frac{Q_d(w, r(w))}{\frac{\partial Q_d(w, r(w))}{\partial w} + \frac{\partial Q_d(w, r(w))}{\partial r} \frac{dr}{dw}} \tag{54}
\]

This differs from the wholesale price with commitment via the boxed term. That term is positive because the fact that the two goods are complements tells us that \( \frac{\partial Q_d(w, r(w))}{\partial r} < 0 \) and Lemma 3 tells us that \( \frac{dr}{dw} < 0 \). The first term in the denominator is negative because the demand curve slopes down: \( \frac{\partial Q_d(w, r(w))}{\partial w} < 0 \). The first term is bigger in magnitude than the second because we assumed that cross-price effects are smaller than own-price effects (\( |\frac{\partial Q_d(w, r(w))}{\partial r}| < |\frac{\partial Q_d(w, r(w))}{\partial w}| \)) and Lemma 3 tells us that \( |\frac{dr}{dw}| < 1 \). As a result, the effect of the
boxed term is to make the denominator smaller in magnitude, but still negative. The equilibrium price without commitment will therefore be higher than with commitment.

If the competitive firms are rational, they will foresee that \( w \) will be higher in the no-commitment case, so \( Y^d \) will be weaker and \( n \) must be smaller for each firm to sell \( y^* \) at \( r^* \) and earn zero profits. Therefore, in the equilibrium without commitment, \( y \) and \( r \) end up the same as when there is commitment. But we have seen that when \( r \) equals \( r^* \), the price that solves the monopolist’s optimization problem is lower than our no-commitment monopoly price, so the no-commitment monopoly must be earning lower profit.

In the end, the monopolist cannot affect the equilibrium competitive price. If the monopolist can commit to a price for his own good, then he will choose a value low enough to encourage the right amount of entry into the competitive-good market. If he cannot commit, then fewer competitive firms will enter. He will find it optimal to react to that with a higher monopoly price, but though that will discourage sales of the competitive good, that discouragement will have been factored into the entry decision and so \( n \) will be just right to allow for zero profits in the competitive-good industry. Thus, we have Proposition 2.

**Proposition 2:** If the monopolist cannot precommit to his output before competitive firms enter the market for a complementary good, his output and profit will be lower. As consumer surplus will also be lower, total surplus is lower than in a vertically integrated monopoly or a monopoly that could commit to future output.

One way to understand Proposition 2 is to think of the monopolist as an innovator who starts production knowing that he will stimulate a secondary market for a different, complementary good next period. Firms thinking of entering the secondary market will have to think about how much the monopolist will produce then. They know that he will produce more once the second market opens up. Once the second period arrives, though, the monopolist will be conferring a positive
externality on the firms in the secondary market. Thus, he will not produce enough to maximize social surplus. Foreseeing this, fewer firms will enter the secondary market. But since the monopolist gets all the surplus in the end anyway, he is worse off because he neglects the externality.

Another way is to think of the benefits to the monopolist from successful deception. Suppose the monopolist promises to produce $Q_1$ once the secondary market opens up. After the market does open up, the monopolist will want to break his promise. He has already gotten the entry he wanted to stimulate so his own good would sell more, and the entrants will not leave (in the short run) even if their revenues do not cover their entry costs. So the monopolist will produce more than if the secondary market did not exist, but not enough to make profits positive there. Foreseeing this, fewer firms would enter the competitive market.

5. A Numerical Example of the Competitive Complements Model

Let consumer demand be linear in the markets for the monopolized and competitive good:

\[ Q_d(w, r) = 2100 - 200w - 180r \]  
\[ Y_d(r, w) = 1400 - 200r - 150w \]

Assume that the monopolist’s marginal cost is $a = 1$ and the competitive firms have fixed cost of $F = .5$, and marginal cost $c(y) = y$, so a retailer’s total cost is

\[ .5 + .5y^2 \]

In the competitive industry, free entry makes profits equal to zero so all firms produce at exactly the minimum average cost in long-run equilibrium. This is found by minimizing $\frac{5}{q} + .5q$, and yields marginal cost of $r^* = 1$ and output of $q = 1$ for each firm.
The monopolist’s profit is

\[ \text{Profit(\text{monopolist})} = Q(w, r(w))(w-a) = (2100-200w-180r(w))(w-1) \]  
(58)

where in the long run, the competitive good’s price is \( r^* = 1 \), but in the short run it could be different. When the monopolist can commit, he will maximize

\[ \text{Profit(\text{monopolist, commitment})} = Q(w, 1)(w-a) = (2100-200w-180)(w-1) \]  
(59)

This is maximized at \( w = 5.3 \). In that case, \( n = Y_d = 1400 - 200r - 150w = 405 \). Profit is 3,698. The price in the competitive market is \( r = 1 \).

Now consider the monopolist without commitment. Short run supply of \( Y \) is \( Y_s = nr \). Thus, we have \( Y_d = 1,400 - 200r - 150w \), \( Y_s = nr \), so \( r(w) = (1,400 - 150w)/(200 + n) \). In that case, the monopolist’s profit function is

\[ \text{Profit(\text{monopolist, no-commitment})} = Q(w, r(w))(w-a) \]
\[ = (1,400 - 200w - 180r(w))(w-1) \]
\[ = (1,400 - 200w - 180\frac{1,400-150w}{200+n})(w-1), \]  
(60)

which is maximized at

\[ w(n) = (9,050 + 115n)/(20(65 + n)) \]  
(61)

We found that \( n = 405 \) in the commitment case. If \( n \) happens to equal 405 in the no-commitment case, then \( w \approx 5.92 \). The price in the competitive market is \( r(405, 5.6) \approx .85 \). Profit is 3,757.

A price of \( r = .85 \) is not a long-run equilibrium, however, because the competitive firms cannot earn negative profits. The equilibrium is where

\[ r(n) = \frac{1,400 - 150w(n)}{200 + n} = 1, \]  
(62)

which solves to \( n \approx 305 \) and \( w \approx 5.96 \). This is a profit of 3,610.
Thus, if competitive firms think the monopolist will not keep his promise to set $w = 5.3$ and fear he would raise it to 5.92 to get them to reduce their price from 1 to .85, fewer of them enter. Since fewer enter, the monopolist ends up charging 5.96, even more than 5.92.

6. Concluding Remarks

The extent of the hold-up problem described in this paper is hard to judge. The phenomenon would be difficult to measure because except in the less realistic case of perfect two-part tariffs the conclusion is not that retailing is blocked entirely, but just that too few retailers enter. The one possible example is Pepsi’s experience in China in trying to buy the particular quality of potatoes needed to make their Lay’s brand of potato chips. Pepsi has had difficulty persuading local farmers to grow the new potato varieties needed, and has had to grow most of them itself, despite its preference for not vertically integrating into farming.\footnote{“To Bag China’s Snack Market, Pepsi Takes Up Potato Farming; Locals Get Taste for Lay’s Chips But Raising Enough Spuds Proves to Be a Challenge; Ms. Huang’s 2 a.m. Scooter Ride,” Chad Terhune, The Wall Street Journal, December 19, 2005; Page A1.} That is an example not of small retailers but of small suppliers, but the principle is the same. In many cases, however, we could hope that the desire to maintain a reputation would help a large monopolist facing small retailers or suppliers to credibly claim that it would maintain its prices at a promised level.

Both models reveal a cost of monopoly not previously acknowledged. This is a supply problem rather than a demand problem such as arises from successive monopolies or monopolies of related products. It is not a problem of inefficient production within the monopoly itself (see, e.g. Farrell (2001) on the problem of incentivizing management), or even of inefficient production within related firms. Instead, it is a problem of an inefficiently costly downstream industry supply curve because there is an efficient scale for a firm to produce, and to attain
even the optimal monopoly level of output, more firms would have to enter.

The retailers in this paper have been perfectly competitive, but competitive holdup would also occur if they were monopolistically competitive, with market power but free entry. In that case, the situation would be complicated by the presence of double marginalization. Further modelling would be necessary to discover whether the two problems have interesting interactions. It might also be interesting to analyze the effect of monopolies that do not compete with each other but which use the same retailers, or of retailers which sell independent goods as well as those supplied by monopolies. This is not the usual sort of common agency situation, where the common agent helps oligopolists to soften competition. Instead, it is a situation of supply-side externalities, where each supplier benefits from the quasi-rents the retailers earn from selling the goods of other suppliers. This would make the transactions costs of a fully inclusive contract even higher, and I conjecture that free-riding would exacerbate the high prices arising from competitive holdup.
References


Cournot, Augustin (1838) Researches into the Mathematical Principles of the Theory of Wealth, original published in French as Recherches sur les principes mathmatiques de la thorie des richesses.


To be added and discussed on tying, bundling, innovation, bottleneck, reducing the threat of entry. : THE ACTIVITIES OF A MONOPOLY FIRM IN ADJACENT COMPETITIVE MARKETS: Economic Consequences And Implications For Competition Policy Patrick Rey - Paul Seabright - Jean Tirole Institut dEconomie Industrielle, Université de Toulouse-1 September 21, 2001


