Rev Ind Organ
DOI 10.1007/s11151-017-9571-2

Some Simple Analytics of Vertically Linked Markets

Joseph Farrell[1]

Abstract A firm may have an incentive to encourage competition and/or efficiency in an "aftermarket" that is vertically linked to a "foremarket" in which it participates. I describe a strong form of this potential incentive, and then explore how it is weakened in plausible circumstances. Some applications to net neutrality are described.

Keywords Aftermarkets · Vertical incentives · Drip pricing · Waterbed

1 Introduction

Net neutrality and a number of other economic questions share a basic vertical linkage structure that I discuss here: A firm, F, sells a product (internet access, in the case of net neutrality), but can also choose policies that affect competitive and other conditions for the use of that product, in ways that affect: (a) F's follow-on profits; and (b) the value derived by its customers.

For instance, an internet access provider might block some traffic, or charge content providers an access fee; or it might agree to cache one provider's content but not another's, or to exempt certain content from bandwidth limits. It might choose more—or less—protective privacy policies.

In other markets, a seller of a photocopier or of a car might insist on doing certain repairs itself and charge high repair prices, or it might include repairs in a bundled

The basic ideas in this paper were presented at the Searle antitrust conference in September 2008, with a focus on the aftermarket application, including a discussion of antitrust market definition that is not key here. I thank Severin Borenstein, Carl Shapiro, and Glen Weyl for helpful comments.

✉ Joseph Farrell
joseph.farrell@berkeley.edu

1 Department of Economics, University of California, Berkeley, Berkeley, CA, USA

Published online: 17 February 2017

warranty, or allow only authorized providers to perform repairs and enforce a variety of conditions in return for authorization. An online seller might treat shipping and handling as a profit center, or might offer free shipping. A credit card issuer might charge higher or lower late fees; an airline determines pricing and service quality for its baggage handling and in-flight meals. Such policies can differ both in pricing and in non-price ways.

Some of these choices can naturally be viewed as hindering otherwise plausible competition in an "aftermarket," but others (such as shipping policies or credit card fees) are less intuitively seen that way.

If F were to choose a harmful or inefficient policy, would demand for its initial product (the photocopier, or the internet access itself) fall so much as to deter F, and would its likely price for that initial product fall so much as to compensate consumers? This article describes a highly simplified but helpful model in which to study those questions.

The basic technique below is to use a change of variable to study how F's profit-maximizing price for its access (or fore-market) product will compare, as between two policies that differ in their follow-on quasi-profits for F and in their direct impacts on the surplus that consumers will actually derive, and will expect to derive. Rather than analyze one policy at a time, I focus on the difference; indeed, it is analytically helpful below to focus on an infinitesimal change in policy that can then in principle be integrated up.

For clarity, I first describe the basic idea in a particularly simple case, with no attempt even at the more-easily-achieved increases in generality.

2 Simplest Case, Full Information

Consider a comparison, or in more lively language, a shift: from a baseline Policy 0 to Policy 1 with regard to conditions that govern firm F's product.

Under the baseline Policy 0, a consumer of type w will derive value $u_0 + w$ from the product (and consumers know their types). The distribution of w among consumers thus determines the demand curve that firm F faces, or equivalently (and more conveniently below) the inverse demand curve, which tells F that it can sell quantity Q at price $p_0(Q)$. F's costs, net of any follow-on revenues under Policy 0, are given by $C(Q)$. There is no need to make assumptions about the shape of the cost or demand functions.

Consequently, under Policy 0, F chooses Q to maximize its profits, which are given by

$$Qp_0(Q) - C(Q).$$

Policy 1 differs in two key ways:

First, Policy 1 hurts each customer by h, relative to Policy 0, so that a consumer of type w now derives value $u_1 + w = u_0 - h + w$, gross of up-front pricing. As a result, the inverse demand curve for F's product, under Policy 1, is given by $p_1(Q) = p_0(Q) - h$.

Second, Policy 1 gives firm F an additional follow-on revenue of m per customer, relative to Policy 0. As a result, F's profit function under Policy 1 has an added term mQ.

Incorporating both of these differences, we see that under Policy 1, F's profit function becomes:

$$Q[p_0(Q) - h] - C(Q) + mQ,$$

which we can equivalently write as

$$Qp_0(Q) - [C(Q) + (h - m)Q].$$

Comparing this to the profit function under Policy 0, we see that it is the same except that in moving to Policy 1, firm F has in effect raised its cost function by a uniform shift of $(h - m)$, which may be positive or negative. It is a uniform shift because I assumed that both m and h are the same for all customers, as I discuss below.

If $h - m > 0$, the higher costs would hurt both firm F and (through the pass-through of the higher effective costs) consumers; thus both would be worse off under Policy 1 than under Policy 0. If $h - m < 0$, both would be better off under Policy 1: the firm because of the lower (effective) cost function; and the consumers because that lower cost function leads to higher output: The marginal type w is lower, and each consumer who buys the product gets surplus equal to the difference between his type and the marginal type.

As the mention of pass-through suggests, this arises because of the effect on firm F's price in the foremarket; but keeping track of it in terms of literal prices involves the profit-maximizing cost pass-through rate, k, for a firm that faces inverse demand conditions $p_0(Q)$. We have $p_0(Q_1) = p_0(Q_0) + k(h - m)$, so for consumer types who buy under both policies, the change in surplus is equal to $-k(h - m)$. But recall that the foremarket price under Policy 1 is not $p_0(Q_1)$ but $p_0(Q_1) - h$. Thus the change in foremarket price is $p_1(Q_1) - p_0(Q_0) = p_0(Q_1) - p_0(Q_0) - h = k(h - m) - h$.

In effect, the change in the foremarket price both compensates fully for the upcoming "harm" h and further gives the consumer the pass-through (via k) of firm F's reduction in effective cost equal to the quasi-efficiency term $m - h$.

Thus inframarginal consumers gain; as noted above, a profitable policy shift also expands output, and since consumer choice is optimal, the new consumers also gain.

In summary: The feed-through effects of a policy change on foremarket pricing cause an optimistic decentralization or alignment result [essentially the one monopoly rent theorem, or "internalization of complementary efficiencies" (Farrell and Weiser 2003)]: firm F will find a policy shift profitable if and only if its direct effects are efficient in terms of per-customer total surplus ($m > h$), in which case consumers also gain.

3 Muted Consumer Response

The optimistic result just derived is extreme in a number of dimensions. A familiar one was identified by Spence (1975): in general h and/or m will vary among consumers. In simple models, firm F's incentive to avoid policy shifts with unduly

large values of h comes from the downward movement of the foremarket demand curve, and specifically that movement at or around the equilibrium point(s) Q_i. That is, it depends on the value of h that applies to "the marginal" consumer. As I discuss below, I will retain here the simplifying assumption that marginal and averages coincide (both as to m and as to h, both of which matter—although differently).

The issue I explore instead is that even if consumers correctly foresee the value they will derive from the product under Policy 0 (perhaps because Policy 0 is conventional or well-established), they may well not fully foresee the impact of Policy 1: They may be unaware (when buying in the foremarket) that it will be applied, or they may not see its full consequences, etc.

In principle consumer foresight might be limited in this way for either or both policies (or neither): It seems most intuitive, however, to think about Policy 0 as well understood but not implemented, and Policy 1 as implemented but less well seen or understood at the time of initial purchase.

Specifically, I assume that although a consumer of type w in fact derives value (as above) of $u_0 - h + w$ under Policy 1, he believes (at foremarket time) that he will derive value $u_0 - th + w$, where t is a parameter that measures consumer information, attention, or sophistication, and we will think of $0 < t < 1$. As a result, we have $p_1(Q) = p_0(Q) - th$. As far as F is concerned (though not for consumer welfare), the inverse demand curve under Policy 1 is shifted down (assuming for concreteness that $h > 0$) not by h but only by th.

F's profit function under Policy 1 is therefore

$$Qp_0(Q) - [C(Q) + (th - m)Q].$$

By the same logic as above, firm F finds that the policy shift is profitable if and only if $m > th$, and alters its profit-maximizing choice of Q in such a way that $p_0(Q_1) - p_0(Q_0) = k(th - m)$. As in the previous section, that difference is the change in the marginal type w who buys, so it is also each inframarginal consumer's change (reduction) in his subjectively *expected* surplus. Such a consumer's change in *realized* surplus, however, is his change in expected surplus minus an additional $(1 - t)h$ (one might call this the "surprise term"): Each inframarginal consumer loses $k(th - m) + (1 - t)h = (1 - t + kt)h - km$; this is an expression that can be either positive or negative even conditioning on the "revealed preference" or "true profitability" condition $m > th$.

Thus while the short-run, ex post, or apparent profit and consumer harm levels are by definition m and h respectively, the true profitability of the policy shift is (by the envelope theorem) to a first-order approximation equal to $m^* = m - th$ per customer, and the true harm per inframarginal customer is $h^* = (1 - t)h - km^* = (1 - t + kt)h - km$. It follows that for truly profitable policy shifts ($m^* \geq 0$), unsurprisingly $h^* \leq (1 - t)h$, and less obviously, $h^* \leq (1 - t)m/t$. These bounds tend to confirm that better information (higher values of t) will help protect consumers; so does the fact that a rational F will only adopt policy shifts for which $m \geq th$. But in contrast to the case with $t = 1$, these assurances and bounds do not rule out problems, but only limit them.

Marginal customers who buy under one policy but not under the other must be considered separately. Because true profitability implies that $m \geq th$, F will (in this specific model) always accompany a profitable policy shift with an increase in Q. New customers (that is, those who buy under Policy 1 but not under Policy 0) are those with $u_0 + w < p_0(Q_0)$ but $u_0 + w - th \geq p_1(Q_1) = p_0(Q_1) - th = p_0(Q_0) - k(m - th) - th$. Such a customer thus gets anticipated consumer surplus under Policy 1 of between zero and $k(m - th)$, and gets actual consumer surplus under Policy 1 equal to that anticipated surplus minus the "surprise term" $(1 - t)h$. That difference has an ambiguous sign, even with the simplifying assumption that consumers accurately perceive (and optimally respond to) Policy 0.

However, with that assumption, if m and h are small, then the consumer surplus of these marginal consumers is second-order, because both the number of such consumers and their per-consumer surplus are first-order in m and h. Thus the first-order effects (m^*, h^*) will dominate.

4 Muted Consumer Response to Infinitesimal Shift

It is helpful to consider the case where the shifts that we called m and h are infinitesimal. Formally, we can think of creating a continuous path of policies that are indexed by a parameter z that varies continuously from 0 to 1, and compare Policy z with Policy $(z + \mathrm{d}z)$. The infinitesimal change dz can be viewed as artificially "causing" changes dm and dh, and consequently dQ. Costs and consumer preferences, and the consumer sophistication parameter t, are assumed constant as z varies.

From the envelope theorem, we can calculate the effect on firm F's true profits as if Q were held constant:

$$\mathrm{d}\pi = (\mathrm{d}m - t\mathrm{d}h)Q.$$

We thus derive a more precise version of the intuitive "per consumer" true profit equation $m^* = m - th$ above: The true effect on profits is the integral of this expression for dπ, and this would be equal to $(m - th)$ times demand, as above, except that demand in general changes as z grows (of course, the way that we used the envelope theorem is irrelevant to that point). Since different increments $(\mathrm{d}m - t\mathrm{d}h)$ get multiplied by different values of Q before being integrated, the extent to which one can easily express the integral in terms of the total changes in m and in th depends on Q not varying "too much," as the intermediate value theorem points out. Thus the m^* equation from above is most useful to the extent that demand does not change "too much" along the path from $z = 0$ to $z = 1$. That will normally be the case for changes that are small relative to the overall customer cost, as well as for relatively inelastic demand.

The effect on consumer surplus has three parts: The first, which applies (even) with fully informed and fully rational consumers, combines the direct effect on each consumer's (perceived) willingness to pay and the price effect, which is the pass-

through of the change in m^*. This gives us the change in inframarginal perceived consumer surplus:

$$\mathrm{d}S_I = k(\mathrm{d}m - t\mathrm{d}h)Q.$$

The second term is the "surprise" effect on inframarginal consumers:

$$dS_s = -Q(1-t)dh.$$

Finally, there is a term for the surplus of marginal consumers, which is more problematic. As noted above, this is second-order in some circumstances, but we cannot assume that those conditions hold all along the z-path.

More generally, the term that reflects marginal consumers' *anticipated* consumer surplus will be second-order for the usual reasons, so to a first-order approximation, their *realized* surplus will be equal to the level of (not the change in) the surprise term—$-(1-t)h$—times the number of such consumers. The number of consumers will in turn be equal to the change in perceived consumer surplus, $k(\mathrm{d}m - t\mathrm{d}h)$, times the responsiveness of demand (that is, change in demand for a unit change in price). That responsiveness or derivative of demand is equal to the absolute value of price-elasticity of demand, times demand Q, divided by the appropriate measure of price.[1] One can also invoke the Lerner equation, but I will leave this implicit because even after these calculations, there remain several factors in this expression that will vary as z goes from 0 to 1.

I have not been able to find a reliable and tractable approximation that simplifies integration along that path. Perhaps at this stage it is most useful simply to observe that when the surprise is everywhere adverse, along a path on which the number of consumers is everywhere growing (recall that both of these held in the discrete comparison in the previous section), this component of change in consumer surplus will be negative.

In contrast, the expressions for the change in profits and in the other components of consumer surplus lend themselves to informative use of the intermediate value theorem provided that (we can choose a path from Policy 0 to Policy 1 such that) Q doesn't change too much along that path. What can we say about this?

Sometimes we may have direct information that a change in policy did not (or is not expected to) result in proportionally large changes in (foremarket) quantity. In other cases, however, we may need to draw inferences. One helpful observation may be that foremarket quantity is driven by price, modified by the anticipated portion th of h; and, intuitively, if that modified price does not change drastically along the path then quantity will not either, unless demand elasticity is extreme.

Thus for "relatively small" discrete effects m and h, the quantity will not (proportionately) vary greatly in the differential equations for profit and the first two components of consumer surplus. In turn, this would justify an approach such as that taken in the previous section; and the third component of consumer surplus, while

[1] I believe that one could legitimately use more than one concept of price here, provided that one is consistent; but the perceived system price most directly drives demand.

not mathematically "second order," will by the same token be limited, because it hinges on and is bounded by the change in output.

5 Implications and Applications

If F chooses t (perhaps by disclosing its policy more prominently than is required or by advertising it, or by obfuscating it into the fine print), this model suggests that it will maximize t if $h<0$ and minimize t if $h > 0$. Conditional on the adoption of Policy 1, the calculations above indicate that if, for instance, $h > 0$, true consumer harm h^* is decreasing in t if and only if $k<1$, which is not guaranteed although it is often viewed as the plausible case. Higher values of t may be more reliably beneficial through their screening effect: Among potential policies with strictly positive m and h (the usual locus of public-policy concern), higher t brings true profit and true consumer effects into better alignment, with the extreme case $t = 1$ being a useful mnemonic.

I view these results as being broadly supportive of initiatives to protect consumers by improving consumer information. Of course, these initiatives are apt to be most necessary where F is not keen to disclose the information ($h > 0$).

Given $t < 1$, and again focusing on potential policies with strictly positive m and h, we have $m^* < m$ and, when $0 \leq k \leq 1$, also $h^* < h$. Intuitively, F internalizes part of the impact on consumers via the downward shift in demand, and consumers internalize or benefit from part of the impact on F via the effect on F's profit-maximizing foremarket price. Thus the vertical linkage helps bring their interests into partial alignment; this is an effect that becomes full alignment in this simple model in the extreme case $t = 1$.

A category of examples of such policies (or policy shifts) with positive values of m and h, arises if F has an ownership or contractual interest in one of several rival providers of a follow-on product or service. For example, a "cable" ISP will normally have an ownership interest in a cable TV (or MVPD) video entertainment business, and a "telco" ISP will normally have an ownership interest in telecom operations. Such a service can also be provided over the ISP's connection by another provider. If the ISP "steers" customers to its TV or phone services, how should we analyze this?

The same issue potentially arises without historical ownership interests: For instance, an ISP might offer travel websites the opportunity to bid for favorable treatment, either in absolute terms or relative to their direct rivals.

If competition among (say) travel websites is well-functioning, then one might expect such steering to reduce total value, and this would suggest $h > m > 0$ (perhaps especially if there are no substantial rents at the website level to be extracted). To some extent this depends on the details of how users are steered, however. For example, if travel websites are not functionally differentiated but each has some degree of market power based on customer inertia or search costs, and if an ISP can steer customers to a website that offers the best deal (without imposing real switching costs or search costs on its customers, of course), it is essentially

providing a comparison-shopping service to its customers. That may be more a case of negative h, and perhaps also negative m.

In a standalone market, if both h and m are negative (F can provide a valuable service but only at a cost to itself), some medium needs to be found for consumers to pay for the service, which can sometimes be awkward with individually low-value online services (indeed, this is often a role for online advertising, which can raise privacy issues: see below).

With vertically linked markets as here, there is no need for F to charge separately: Its reward comes in the foremarket (or perhaps it is better to say in the system market). If F incurs or imposes substantial costs in the process of charging customers, in fact, that might suggest that t is well below 1 (because that evidence suggests that F can't capture customer gains via the demand shift), making the revealed preference test (the fact that F chooses to do it) less informative in diagnosing whether the policy or service is desirable, and the pass-through mechanism less helpful in protecting consumers, as above.

6 Related Literature and Further Comments

As the 2008 version of this paper stressed, it is most closely related to the antitrust economics literature on aftermarkets. In that literature, the early contribution of Shapiro (1995) suggested that under foremarket (perfect) "competition," any (increase in) per-consumer follow-on profits (m in my notation) would be fully passed through to consumers, so that the net effect on consumers of a policy shift would be $m - h$.

My original goal was to explore the extent to which profit-maximizing pass-through would preserve or extend this reassuring lesson in more realistically "competitive" foremarket/aftermarket environments. My model is a single-firm treatment, assuming a stable residual demand curve (based on perceived value), which does not yet capture all the effects if multiple competing firms all adopt, or all are prohibited from adopting, policy shifts such as those considered. Nor does it capture issues arising when a single buyer may purchase a durable good repeatedly in the foremarket but can treat aftermarket "repair" services as a partial substitute: Carlton and Waldman (2010) describe ways such as this in which a competitive aftermarket is not necessarily an efficient solution.

In part to allow for such possibilities, I treat m and h as primitives rather than derive them from standard pricing models of competition and monopoly, except for the brief discussion in Sect. 5. Other contributions include Borenstein et al. (2000), Cabral (2014), and Coppi (2007).

In the analysis of privacy policies, m and h may flow from F's re-use or sale of a customer's information that has been disclosed to F (in the course of what the customer may have thought was the transaction but that as far as F was concerned was only the foremarket). A simplified example was described in Farrell (2012), and a much more general recent treatment is in O'Brien and Smith (2014). Among other things, O'Brien and Smith allow for non-uniform valuations of privacy, which is to say (in my notation) that h differs among customers. The observation that such

variation in willingness to pay for product characteristics (which one might broadly call "quality") affects the welfare economics is of course classically due to Spence (1975). Veiga and Weyl (2016) give a sophisticated extension and generalization of those issues.

In contrast, I have chosen to downplay the "Spence distortion." This is both for simplicity, and because when price discrimination by F is important, more than one customer type will be "the" marginal customer,[2] and it becomes analytically difficult (especially if one tries to take the analysis to real-world policy shifts) to predict which types that will be. As a result, it seems to me (informally) that a prediction of the effect of a policy on the willingness to pay of "the marginal customer" will involve some weighted average of gauging the effect on the willingness to pay of many potentially-marginal customers, in a way that reflects more uncertainty (and perhaps less multidimensional variety) than the "average marginal customer" calculations of Veiga and Weyl. A limiting case would view all customer types as being potentially marginal, which then brings together the analysis of their willingness to pay (which is central to F's incentives) with the welfare economics analysis that adds up all consuming customers' willingness to pay.

I don't mean to suggest that the analysis of Spence-type effects is unimportant; but those are my reasons for not having given it priority here.

Follow-on revenue effects are important in a wide range of contexts, well beyond "just" privacy policy, net neutrality, and the aftermarkets that are of concern in antitrust. Indeed, one might identify such effects as key to multiproduct pricing and even to the scope of the firm. They are key to understanding penetration pricing, for example.

Empirically, Agarwal et al. (2015) consider whether the "waterbed effect" (essentially an effect of m on foremarket pricing) can be detected in a large credit card dataset, although their results are largely negative. Even the US 2010 Horizontal Merger Guidelines—traditionally focused entirely on horizontal concerns—mention the topic (albeit in a footnote).

Better understanding of follow-on effects and their back-echo, the consequences for foremarket pricing and competition, as well as the concerns that arise when customers do not adequately foresee the follow-on consequences, seem to me a key topic for further economics research.

References

Agarwal, S., Chomsisengphet, S., Mahoney, N., & Stroebel, J. (2015). Regulating consumer financial products: Evidence from credit cards. *The Quarterly Journal of Economics, 130*(1), 111–164.

[2] Severin Borenstein has suggested a possible illustration: In airlines, the introduction of "premium economy" seating, or more generally the shifting around in relative amenities and terms that are offered by airlines to different groups of passengers, repeatedly changes the identity of "the marginal customer" (for any one airline, or, if it matters, for the industry).

Borenstein, S., MacKie-Mason, J., & Netz, J. (2000). Exercising market power in proprietary aftermarkets. *Journal of Economics and Management Strategy, 9,* 157–188.

Cabral, L. (2014). Aftermarket power and foremarket competition. *International Journal of Industrial Organization, 35,* 60–69.

Carlton, D., & Waldman, M. (2010). Competition, monopoly, and aftermarkets. *Journal of Law Economics and Organization, 26,* 54–91.

Coppi, L. (2007). Aftermarket monopolization: The emerging consensus in economics. *Antitrust Bulletin, 52,* 53–71.

Farrell, J. (2012). Can privacy be just another good? *The Journal on Telecommunications and High-Technology Law, 10,* 251–264.

Farrell, J., & Weiser, P. (2003). Modularity, vertical integration, and open access policies. *Harvard Journal of Law and Technology, 17,* 85–134.

O'Brien, D., & Smith, D. (2014). *Privacy in online markets: A welfare analysis of demand rotations*. Federal Trade Commission Bureau of Economics working paper no. 323.

Shapiro, C. (1995). Aftermarkets and consumer welfare: Making sense of Kodak. *Antitrust Law Journal, 63,* 483–511.

Spence, M. (1975). Monopoly, quality and regulation. *Bell Journal of Economics, 6,* 417–429.

U.S. Department of Justice and Federal Trade Commission. (2010). *Horizontal merger guidelines*.

Veiga, A., & Weyl, G. (2016). Product design in selection markets. *Quarterly Journal of Economics, 131,* 1007–1056.