

Price Discrimination between Retailers with and without Market Power

29 October 2008

Barick Chung and Eric Rasmusen

Abstract

Some retail markets are more competitive than others. A manufacturer with market power in the wholesale market who sells his product to competing retailers in cities and monopolistic ones in each of various towns must set the wholesale price difference between towns and cities to be smaller than the transportation cost to prevent “grey market” arbitrage. If he uses linear pricing, the town retail price will be even higher than under single-retailer double marginalization. Two-part tariffs do not solve the problem as they would if there were a single retailer, because the wholesale unit price must be higher than marginal cost to prevent arbitrage to the cities. If transportation costs are low, price discrimination is difficult and two-part tariffs come to resemble inefficient linear monopoly pricing. High transportation costs allow greater efficiency in contracting, and this can outweigh the negative direct effect on welfare.

Chung: Instructor, Department of Economics, The Chinese University of Hong Kong. Address: 1011 Esther Lee Building, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. Office: (852) 2609-8232. Fax: (852) 2603-5805. Barick@cuhk.edu.hk. <http://ihome.cuhk.edu.hk/~b580765/>.

Rasmusen: Dan R. and Catherine M. Dalton Professor, Department of Business Economics and Public Policy, Kelley School of Business, BU 438, 1309 E. 10th Street, Bloomington, Indiana, 47405-1701. Office: (812) 855-9219. Fax: 812-855-3354. erasmuse@indiana.edu, <http://www.rasmusen.org>.

This paper: <http://www.rasmusen.org/papers/retailers-chung-rasmusen.pdf> .

We thank Michael Baye, Rick Harbaugh, Michael Rauh and David Waterman for helpful comments.

JEL Classifications: D4, L42, M21

Keywords: Price discrimination; Double marginalization; Retail network; Transportation costs; Two-part tariffs; Vertical restraints

1 Introduction

Double marginalization and price discrimination are two of industrial organization's key ideas. The problem of double marginalization is that if the manufacturer has market power and charges higher than marginal cost to the retailer, and the retailer has market power and marks the retail price to the consumer up to above his own marginal cost, then industry profits and total welfare both will be lower than under vertical integration. The problem of price discrimination is that a manufacturer selling to two retailers with different demands would like to charge different prices to them. In this paper, we see what happens when the two problems are combined: a manufacturer wants to use two pricing schemes, one for competing retailers (in large "cities") and one for retailers with market power (in small "towns" that are natural monopolies).

For price discrimination to occur, resale must not be costless, so we will assume positive transportation costs whenever there is a resale. We will look at both linear pricing— a constant price per unit— and two-part tariffs— a constant per-unit price plus a fixed access fee. We will specify a transaction cost for using a two-part tariff, but we will also let this cost be zero, and it will not drive our most interesting results.

We establish two main results. First, in a retail network, with linear pricing the final prices can be even higher than under single-retailer double marginalization. Second, even when two-part tariffs are used, the wholesale price can be above marginal cost, and thus double marginalization cannot be totally eliminated. These two results imply that in a retail network we might expect more complex vertical restraints than when the manufacturer sells to only one retailer.

In the model, a monopolist manufacturer supplies its product to multiple retailers who sell in both small towns and large cities. We can imagine that a retailer is a grocery chain. Each retailer is a monopolist in one or more towns, which are natural monopolies, but in cities intense price competition drives prices to marginal cost. Since retailers are competitive in cities, the manufacturer earns the monopoly profit by setting wholesale prices there equal to the monopoly price. On the other hand, since retailers are monopolists in towns, there will be double marginalization if the town wholesale price is above the marginal cost. If the manufacturer chooses linear pricing contracts for both towns and cities, it suffers double marginalization in towns. Two-part tariffs can help the problem in the towns, but they are unnecessary and impractical for the cities, where retailers have no market power and their zero profits would not permit them to pay a two-part tariff's lump-sum fee.

The manufacturer may want to sign different contracts with the retailers — one for towns and one for cities, but there is a problem with this arrangement. A retailer may arbitrage its inventories from its stores in towns to its stores in cities, or vice versa. We will assume that retailers must pay a transportation cost if they arbitrage, and we will show that the manufacturer will then set the wholesale price difference between towns and cities smaller than the transportation cost, preventing arbitrage. Subject to this constraint, if transportation costs are "small", with linear pricing the retail price will be even higher

than under double marginalization. With two-part tariffs in the towns, the wholesale price will be above marginal cost. As a result, double marginalization cannot be totally eliminated by two-part tariff pricing.

Other nonlinear contracts besides two-part tariffs could solve the problem, but they must exert more direct control on the retailers, controlling either the quantities they buy or the prices they charge. We will discuss these alternatives once we have shown formally how a two-part tariff will fail to work.

The Literature

Spengler (1950) set out the classic problem of double marginalization. If an upstream manufacturer with monopoly power sells to a downstream retailer who has monopoly power in the consumer market, and they choose prices independently, the consumer price will be higher than the monopoly price. As a result, both consumer and producer surplus fall relative to the “single marginalization” one-stage monopoly. Cook (1955) and Hirshleifer (1956) show that firms can overcome the problem if the manufacturer uses a two-part tariff, charging marginal cost to the retailer plus a fixed fee.

Since marginal costs are usually not observable, putting the theory into practice can be difficult. One stream of discussion in the literature has focused on the choice of transfer prices within multi-divisional firms (for example, Menge (1961), Vaysman (1998), Baldenius, Reichelstein and Sahay (1999), and Baldenius and Reichelstein (2006)). Another stream of the literature (e.g., Mathewson and Winter (1983a,b, 1984), Rey and Tirole (1986, 2003), Tirole (1988) and Rey (2003)) studies double marginalization in the context of vertical restraints generally. If vertical externalities are present, upstream manufacturers often impose additional restraints in contracts to ensure that right actions are taken by their downstream retailers. An example is the vertical externality generated by double marginalization. By using a franchise fee or a two-part tariff, vertical firms can coordinate their pricing decisions. Not only can they maximize the industry joint profit, but consumers are better off. In this sense vertical restraints are efficiency-enhancing.

It is well known that the two-part tariff is a sufficient instrument if there is only simple double marginalization, with no uncertainty or private information. We often, however, observe more complicated and hard-to-administer contracts involving resale price maintenance, exclusive territories, exclusive dealing, quantity forcing, revenue sharing and all-units discounts. Why? Perhaps for anti-competitive purposes. For example, Salop and Scheffman (1983) show that a dominant firm can raise its rivals’ costs and earn a profit if, in doing so, the dominant firm’s residual demand increases more than its own average cost. Rasmusen, Ramseyer and Wiley (1991) show that by signing exclusion contracts with sufficiently large number of customers a less efficient firm can deter the entries of even more efficient potential competitors. Rey and Stiglitz (1988, 1994) show how monopolistic manufacturers can use vertical restraints such as exclusive territories to increase profits by reducing inter-brand competition. Moreover, a vertical structure often faces a number of decision variables, such as selling efforts, retailers’ inventory level and pre-sale service. If the upstream firm wants to control more than one dimension of the downstream firms’ behavior, more complex re-

straints are sometimes desirable. Dana and Spier (2001) argue that in the video rental industry, movie studios would like to simultaneously soften downstream retailers' competition and encourage inventory holding. Two-part tariffs are unable to accomplish both. By using revenue sharing, together with a low input price, movie studios can implement the vertical integration outcome. Kolay, Shaffer and Ordober (2004) show that all-units discount contracts can serve as a device for a monopolist manufacturer to extract greater surplus from consumers, when the downstream retailer has private information about consumers' demand, as well as eliminate the double marginalization problem.

xxx We should discuss this paper, and use it for current references: Milliou, C., E. Petrakis 2007, "Upstream Horizontal Mergers, Vertical Contracts and Bargaining," *International Journal of Industrial Organization* , 25, 963987.

Or perhaps the manufacturer's problem is not asymmetric information or how to deter entry, but how to make price discrimination successful when resale is possible at some cost. That is what we will explore here.

2 The Model

A monopolist manufacturer supplies a homogeneous product to retailers in n_a identical small towns and n_b identical large cities. Consumers must buy in their own city or town, but retailers can buy from other locations. Because there is only one retailer in each small town, town retailers are monopolists with respect to selling to consumers. City retailers, however, compete intensely using Bertrand pricing and earn zero profits. If town or city retailers sell to each other, they compete using Bertrand pricing (e.g., town retailers would choose prices simultaneously to compete in selling to city retailers).

Denote wholesale prices by w_a and w_b , the franchise fees by F_a and F_b , retail prices by p_a and p_b , and consumer demands by $D_a(p_a)$ and $D_b(p_b)$. We assume standard downward sloping and concave demands, and that $D_a''' \geq 0$. We assume that demand is stronger in cities than in towns—enough stronger that if arbitrage were impossible, the wholesale prices would be higher in the cities, i.e. $p_b^m > \tilde{w}_a$. This assumption will be used later in Proposition 1.

Let the manufacturer have constant marginal cost $c > 0$ and let retailers have zero distribution costs. If resale occurs between towns and cities, retailers incur a transportation cost $t > 0$.

Because we assume consumer demand is concave, there exist unique retail monopoly prices which we will denote p_a^m and p_b^m that maximize vertically integrated profits.

The manufacturer can choose from two types of contracts—linear pricing or a two-part tariff. He will not choose to use both in the same market in equilibrium. Let I_a be an indicator variable, where $I_i = 1$ if the manufacturer chooses a two-part tariff contract for

market i for $i = a, b$. Under linear pricing, the retailer pays amount w_i per unit for as many units as he wishes to buy. Under a two-part tariff, the retailer pays the lump sum $F_i > 0$ if he buys any units at all, plus w_i per unit. Linear pricing has zero transactions cost, but the two-part tariff, being more complicated, incurs an extra transaction cost of $g \geq 0$ ($g = 0$ is allowed, and is itself an interesting case). The manufacturer chooses the contracts first, and the retailers react by choosing the quantities to purchase, so in effect the manufacturer makes take-it-or-leave-it offers to the retailers.

We will denote a retailer's derived demand for the good as a function of the wholesale price by $D_a^r(w_a)$ or $D_b^r(w_b)$. Lemma 1 establishes that these are concave and strictly decreasing.

Lemma 1. $D_a^r(w_a)$ and $D_b^r(w_b)$, are strictly decreasing and concave in wholesale prices w_a and w_b .

Proof: In the Appendix.

3 The Equilibrium

We seek a subgame perfect Nash equilibrium in which the manufacturer chooses what pricing scheme to use for cities and for towns and what prices and franchise fees to charge, and the retailers react by choosing purchase amounts and wholesale and retail sales to maximize their profits.

First, consider what would happen if arbitrage could be blocked by contract— that is, if a retailer could be deterred from reselling in a forbidden market by the threat of a suit for breach of contract.

The manufacturer would choose a linear contract for cities and set the wholesale price equal to the monopoly retail price: $w_b^m = p_b^m$. A city retailer would buy the quantity $D_a(p_b^m)$ and choose the retail price p_b^m to maximize its profits under Bertrand competition with other city retailers. The manufacturer would do worse with a two-part tariff, because the retailers, knowing that they would compete their profit margin to zero, would be unwilling to pay $F_b > 0$ to buy a positive quantity.

The manufacturer would choose a two-part tariff for towns, with a wholesale price equal to marginal cost, so $w_a = c$. A town retailer would buy the quantity $D_a(p_a^m)$ and choose the retail price p_a^m to maximize its profits, and the franchise fee F_a would equal the gross profits.

If the manufacturer used linear pricing in towns, his profits would fall because of double marginalization. Denote his optimal linear price in this no-resale case by \tilde{w}_a , which would be greater than c if he is to obtain positive profits. As a result, when town retailers react with their optimal retail price, \tilde{p}_a , it will exceed the optimal monopoly price, p_a^m . More

precisely,

$$\tilde{w}_a \equiv \underset{w_a}{\operatorname{argmax}} \{(w_a - c) \cdot D_a(p_a^*(w_a))\} \quad (1)$$

and:

$$\tilde{p}_a \equiv p_a^*(\tilde{w}_a) \quad (2)$$

where $p_a^*(w_a)$ is the reaction function of the town retailers and satisfies the first order condition of the town retailers' profit maximization.

Now let us assume that contracts cannot punish intermarket arbitrage. Since each retailer sells both in towns and cities, if the wholesale price difference,

$$d \equiv (w_b - w_a),$$

is larger than the retailer transportation cost t , a retailer will arbitrage the product among its stores across locations, or one retailer will arbitrage by buying from another.

Let us define two critical values of transportation costs:

$$t_0 \equiv p_b^m - \tilde{w}_a \text{ and } t_1 \equiv p_b^m - c$$

Recall that in separate maximization problems p_b^m is the monopoly wholesale price in cities, and \tilde{w}_a and c are the wholesale prices for linear pricing and two-part tariffs in towns. Thus, t_0 is the optimal wholesale price difference between towns and cities if the retailers do not arbitrage and the manufacturer chooses linear pricing for towns. Similarly, t_1 is the optimal wholesale price difference between towns and cities if the retailers do not arbitrage and the manufacturer chooses a two-part tariff for towns. The manufacturer's profit, π , depends on what kind of arbitrage arises from its city-town pricing. The three possibilities are: (i) the retailers do not arbitrage, (ii) the retailers arbitrage goods from towns to cities, or (iii) the retailers arbitrage goods from cities to towns. Let us examine these possibilities in turn.

Case (i) (No arbitrage): If the manufacturer chooses $|(w_b - w_a)| \leq t$, then the retailers do not arbitrage the goods. The manufacturer's profit is the sum of its profits from all towns and cities:

$$\begin{aligned} \underset{w_a, w_b}{\operatorname{Maximum}} \pi|_{|(w_b - w_a)| \leq t} &= n_a \cdot [(w_a - c) \cdot D_a^r(w_a) \\ &\quad + I_a \cdot (F_a(w_a) - g)] \\ &\quad + n_b \cdot [(w_b - c) \cdot D_b^r(w_{city})] \end{aligned} \quad (3)$$

The first square bracket in (3) is the manufacturer's profit from a town, and the second square bracket is his profit from a city. The first round brackets in each square bracket are the profit margins. I_a is the indicator variable whether the manufacturer adopts two-part tariff contracts for towns, and its gain from a two-part tariff contract from a town is the franchise fee, F_a , net of the negotiation fee, g .

Case (ii) (Town retailers sell to city retailers): If the manufacturer chooses $(w_b - w_a) > t$, city retailers will buy in towns. We have assumed that town retailers compete against each

other in prices for sales to other retailers, so the resulting price paid by the city retailers will be driven down to w_a . To this the city retailers must add the transportation cost t to get their marginal cost for selling to consumers, $w_a + t$. Since the retailers compete in Bertrand equilibrium in cities, the market price in cities is also $w_a + t$, and correspondingly, the effective demand is $D_b^r(w_a + t)$. The manufacturer's maximization problem becomes:

$$\begin{aligned} \underset{w_a, w_b}{Max} \pi |_{(w_b - w_a) > t} &= n_a \cdot [(w_a - c) \cdot D_a^r(w_a) \\ &\quad + I_a \cdot (F_a(w_a) - g)] \\ &\quad + n_b \cdot [(w_a - c) \cdot D_b^r(w_a + t)] \end{aligned} \quad (4)$$

Case (iii) (City retailers sell to town retailers): If the manufacturer chooses $(w_b - w_a) < -t$, the retailers will arbitrage goods from cities to towns. The effective wholesale price for towns will be w_b , to which t must be added to get the town retailers' marginal cost, and the manufacturer's profit function will be:

$$\begin{aligned} \underset{w_a, w_b}{Max} \pi |_{(w_b - w_a) < -t} &= n_a \cdot [(w_b - c) \cdot D_a^r(w_b + t) \\ &\quad + I_a \cdot (F_a(w_b + t) - g)] \\ &\quad + n_b \cdot [(w_b - c) \cdot D_b^r(w_{city})] \end{aligned} \quad (5)$$

Proposition 1 tells us that the manufacturer prefers Case (i), no arbitrage.

Proposition 1. In equilibrium, the manufacturer will prevent arbitrage. Wholesale prices will be higher in the city, but by no more than the transportation cost: $0 < (w_b - w_a) \leq t$.

Proof: See the Appendix.

If the transportation cost is smaller than the wholesale price difference, the retailers arbitrage. To prevent this happening, the manufacturer has to set a smaller wholesale price difference. Alternatively, the manufacturer could allow arbitrage to occur. Proposition 1 says that the manufacturer's best choice is to prevent arbitrage. This is because transportation is costly, creating a social loss that also hurts the manufacturer. The city retailers are earning zero profits anyway and their higher transportation costs from arbitrage result in less demand from consumers and lower sales by the manufacturer. If the manufacturer could impose vertical control over the actions of downstream retailers, it could prevent arbitrage to realize the vertically integrated profit. Since it cannot, it must reduce the extent of its price discrimination.

If t is smaller than the critical values, t_0 and t_1 , the transportation cost is "small" and arbitrage starts to be attractive to retailers. To prevent arbitrage, the manufacturer must choose $(w_b - w_a) \leq t$. Under this constraint, the manufacturer's profit is increasing in the wholesale price difference. Thus, for t smaller than the critical transportation costs the manufacturer chooses $(w_b - w_a) = t$ and the no-arbitrage constraint is binding. For t larger than the critical values, the retailers do not arbitrage and therefore the manufacturer chooses the unconstrained optimum.

4 The Equilibrium Contracts

To determine the manufacturer's choice of contracts for towns, we will compare the manufacturer's profits under linear pricing, $\pi^*(I = 0, t)$, and under two-part tariffs, $\pi^*(I = 1, g, t)$. Recall that $g \geq 0$ is the transactions cost of the manufacturer if it adopts a two-part tariff. Let $\bar{g}(t)$, a function of transportation cost t , be the transactions cost such that the manufacturer is indifferent between choosing the two types of contracts for towns, i.e. $\bar{g}(t) : \pi^*(I = 1, \bar{g}, t) \equiv \pi^*(I = 0, t)$.

Proposition 2. (i) There exists a continuous function $\bar{g}(t) > 0$ such that if the negotiation cost is $g > \bar{g}(t)$ then in equilibrium linear pricing is used in towns but if $g < \bar{g}(t)$ a two-part tariff is used (and either might be used if $g = \bar{g}(t)$).

(ii) Let the transportation cost t be smaller than the critical value t_0 if the transaction cost is $g \geq \bar{g}(t)$ or t_1 if $g < \bar{g}(t)$ (as in Figure 1's Regions I and III). Under linear pricing (Region I), the double marginalization problem is magnified: the retail price for towns is higher than under standard double marginalization. Even under a two-part tariff (Region III), the problem in towns cannot be totally eliminated, as the wholesale price for towns will be above marginal cost and the retail price will be higher than the integrated monopoly price.

Proof: See the Appendix.

One may interpret $n_a \cdot \bar{g}(t)$ as the manufacturer's benefit from using two-part tariff contracts, because the manufacturer adopts a two-part tariff if and only if the benefit from using the two-part tariff is greater than the total cost $n_a \cdot g$. We have not and will not assume linear demands in our model, but for illustration we use linear demands in Figure 1 to generate the $\bar{g}(t)$ function (the upward sloping curve in the middle of the figure). Figure 1 shows the manufacturer's choice of contracts and pricing decisions in equilibrium (with $n_a = 2$, $n_b = 1$, $c = 2$, $Q_a = 10 - p_a$, and $Q_b = 18 - p_b$).

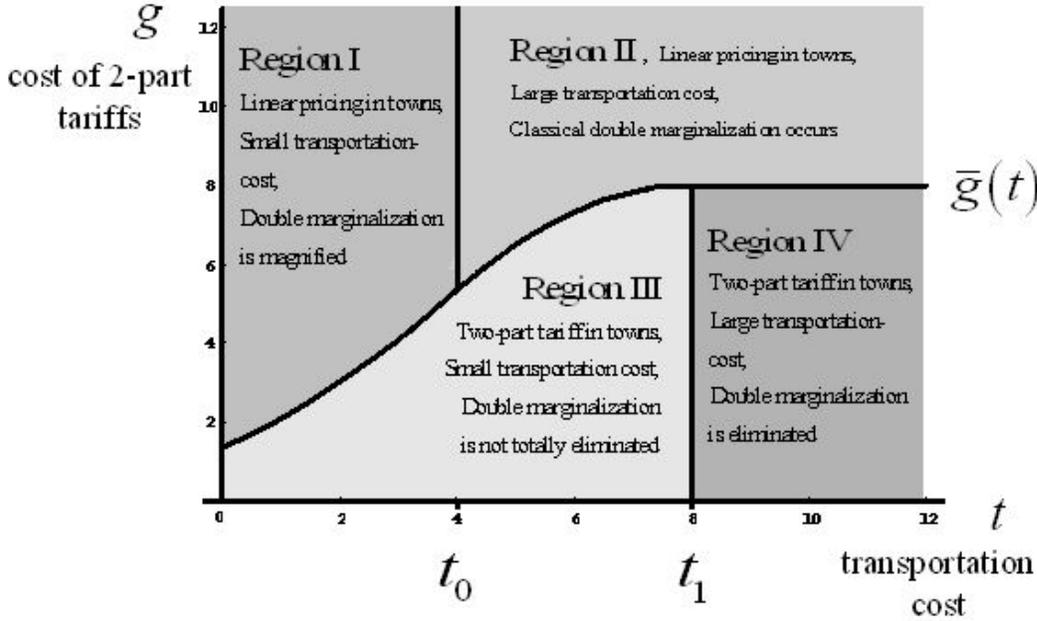


Figure 1: Double Marginalization in the Four Regions

In Figure 1, the curve $\bar{g}(t)$ and the two critical values of the transportation costs partition the space into four pieces, Regions I to IV. Above $\bar{g}(t)$ the negotiation cost is “large” and the manufacturer chooses linear contracts for towns (recall that by the manufacturer always chooses linear contracts for cities). Below $\bar{g}(t)$ the negotiation cost is “small” and the manufacturer chooses two-part tariffs for towns. Above $\bar{g}(t)$, the manufacturer uses linear pricing and thus t_0 is the critical value of transportation cost. Below $\bar{g}(t)$, the critical value is t_1 . To the left of the critical values the transportation costs are “small”. Thus in Regions I and III the manufacturer chooses $(w_b - w_a)^* = t$ to prevent arbitrage. Conversely, to the right of the critical values the transportation costs are “large”, and thus in Regions II and IV the manufacturer chooses the unconstrained optimal wholesale prices.

The four regions in Figure 1 exist for non-linear concave demands as long as t_0 and t_1 are strictly positive and finite and $\bar{g}(t)$ is continuous, positive and finite. The value $t_0 \equiv p_b^m - \tilde{w}_a$ is positive because $p_b^m > \tilde{w}_a$ and is finite because p_b^m is bounded. Similarly, t_1 is strictly positive and finite. $\bar{g}(t)$ is generally continuous because $w_a^*(t)$ and $w_b^*(t)$ are continuous at the critical values. The function is non-negative because the two-part tariff is more flexible than linear pricing. For t larger than the critical values, retailers do not arbitrage and the manufacturer’s benefit from using a two-part tariff is strictly positive. Thus, $\bar{g}(t)$ is strictly positive at the critical values, and since $\bar{g}(t)$ is continuous at these values, the four regions are not empty. Notice that t_1 is always strictly larger than t_0 . This is because under two-part tariffs the manufacturer lowers the wholesale price and thus enlarges the wholesale price difference between towns and cities. For a greater range of transportation costs the manufacturer has to choose $(w_b - w_a)^* = t$ to prevent the retailers’ arbitrage.

At the beginning of this section, we show that to the left of the critical values the transportation costs are “small”, and therefore in Regions I and III the manufacturer chooses $(w_b - w_a)^* = t$ to prevent arbitrage. To achieve this target, the manufacturer can adjust the wholesale price for only towns, only cities, or both. Proposition 2 says that the manufacturer will adjust both. While decreasing the wholesale price for cities, the manufacturer increases the wholesale price for towns. This not only achieves $(w_b - w_a)^* = t$ to prevent arbitrage, but also magnifies the double marginalization problem for towns. Even with a two-part tariff, the wholesale price will be above marginal cost and the magnified double marginalization cannot be totally eliminated.

5 Welfare

If the two-part tariff contracting cost g is low, two-part tariffs are used in towns regardless of t . When t increases from a low level, the city price rises and the town price falls. Manufacturer profits from both sources rise, and both town and city retailer profits do not change. Town consumers have higher consumer surplus, and city consumers have lower consumer surplus. After t rises to t_1 , however, arbitrage is no longer a constraint and further increases have no effect.

If the two-part tariff contracting cost g is high enough, it always exceeds $\bar{g}(t)$. Linear pricing is used, and if t is smaller than t_0 , the rising transportation cost increases prices in the cities (the larger markets) and reduces them in the towns. The profits of city retailers do not change and the consumer surplus of city consumers fall in t , and those of town retailers and consumers rise. After t rises to t_0 , arbitrage is no longer a binding constraint and further increases have no effect.

In Figure 2 we use the same demands used in Figure 1 (fixing the two-part tariff transaction cost at $g = 3$) to show the changes of social welfare with transportation costs for a specific functional form. $\bar{t}(g)$ is the transportation cost, as a function of g , so that the manufacturer is indifferent between choosing the two types of contracts for towns.

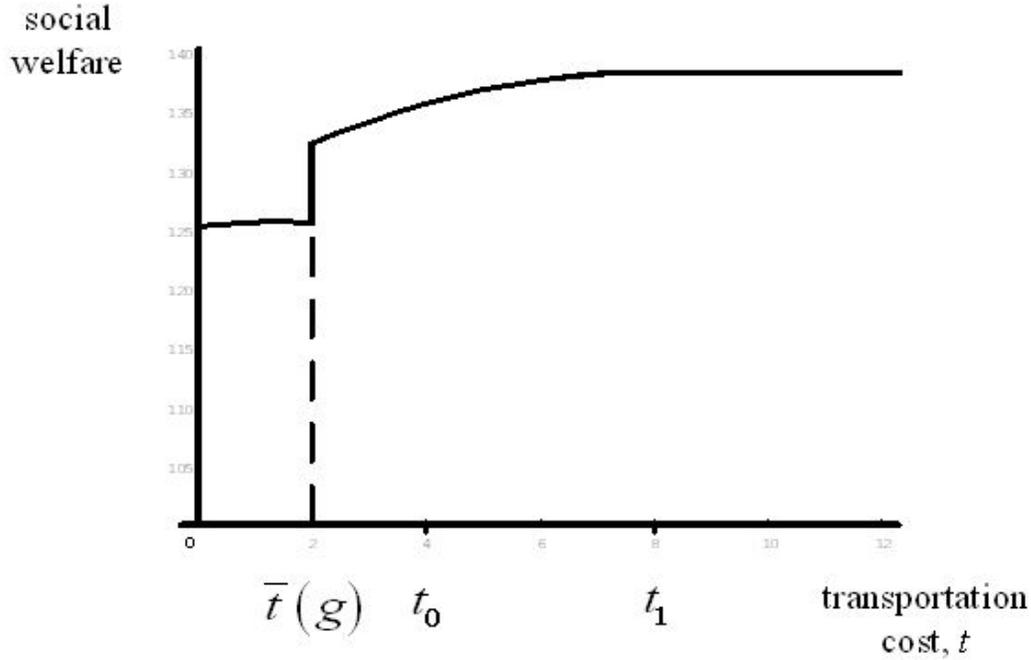


Figure 2. Social Welfare Changes with Transportation Costs

The manufacturer chooses linear pricing contracts if $t \leq \bar{t}(g)$ and two-part tariff contracts otherwise. This is because if the constraint is binding by the Envelope theorem the manufacturer's profit is increasing in t . Since the negotiation cost g is fixed, if under unconstrained optimization the manufacturer's benefit from using a two-part tariff is larger than $n_a \cdot g$, we can find $\bar{t}(g)$ such that the manufacturer is indifferent between the two types of contracts. As t increases, the manufacturer switches from using linear pricing to a two-part tariff at $\bar{t}(g)$. Since in a two-part tariff the manufacturer chooses lower wholesale prices for both cities and towns, after the switch the retail prices decrease in both markets and social welfare increases.

Thus, although it is hard to make general statements about welfare changes with t because increasing the amount of price discrimination can in general make welfare either rise or fall, we can make a more limited statement:

Proposition 3. (a) An increase in the retailer transportation cost t increases social welfare in towns and reduces it in cities if t is small enough to have any effect at all. (b) If t rises to exceed $\bar{t}(g)$, prices fall in both towns and cities and consumer surpluses rise.

Proof. See the appendix.

If the retailer transportation cost is small, the manufacturer will set a small wholesale price difference to prevent arbitrage. This increases the wholesale price for towns and decreases the wholesale price for cities. An increase in transportation cost relaxes the constraint of the manufacturer and thus increases his profit, but it allows him to increase

the disparity in prices, so in general we cannot say whether the increased welfare from town demand exceeds the decreased welfare from city demand. As t increases, the manufacturer switches to use a two-part tariff at $\bar{t}(g)$. The wholesale prices for both towns and cities will fall once a two-part tariff is used in the towns, and the retail prices are strictly increasing in the wholesale prices, so social welfare increases at $\bar{t}(g)$. Two-part tariffs do incur the transaction cost g , however. If g is small, overall social welfare increases and there is a jump at $\bar{t}(g)$.

How social welfare changes with the negotiation cost, g , is simpler. Consider a particular value of the transportation cost, t . Given t , we can compute $\bar{g}(t)$. If $g \geq \bar{g}(t)$, the manufacturer chooses linear pricing contracts. In this situation, the negotiation cost g does not go into the manufacturer's profit function, and thus is irrelevant to social welfare. Conversely, if $g < \bar{g}(t)$, the manufacturer chooses a two-part tariff. In this situation, negotiation cost becomes part of the social cost, and social welfare falls linearly with g .

6 Conclusion

We have explored one reason why a two-part tariff might not be sufficient to solve the problem of double marginalization in a retail network: the danger of resale while attempting price discrimination. This insufficiency could help explain why vertical firms sometimes use more complicated and hard-to-administer contracts than the simple two-part tariff.

If retailers sell to markets with different market powers, wholesale price difference between markets may cause retailer arbitrage. If a monopolist manufacturer can impose a vertical control over the actions of the downstream retailers, it can prevent arbitrage directly, but otherwise it must do so by setting the wholesale price difference to be smaller than the transportation cost between retailers. Taking as given this strategy of the manufacturer, we examined the classical double marginalization problem. If the transportation cost is small, the manufacturer will raise the wholesale price for towns and reduce the wholesale price for cities. If the manufacturer adopts linear pricing, this distortion of optimal wholesale pricing magnifies the double marginalization problem in towns. Even with a two-part tariff, the wholesale price will be higher than the marginal cost, and the double marginalization problem cannot be totally eliminated.

There are occasions where a two-part tariff fails to be a sufficient vertical restraint; for example, when the retailers have private information or the retailers are risk averse and there is uncertainty about demands. In this paper, we did not assume private information or uncertain demands. Neither did we assume inter-brand competition or retailer services that introduce horizontal externalities. Even so, in a retail network, if transportation cost is small, a two-part tariff cannot achieve the profits of vertical integration. Manufacturers usually sell to multiple markets through retailers, and retailers usually have different market power in different locations. Since the two-part tariff is not a sufficient instrument in a retail network, we may observe firms using more sophisticated combinations of vertical restraints instead, or, if two-part tariffs are administratively costly, going to the other extreme of giving

up on two-part tariffs and using linear prices.

Preventing resale directly may be prohibitively costly, but one contractual feature that would be helpful, if legal, would be a prohibition on retailers selling the manufacturer's good without having paid the franchise fee. This would eliminate the possibility of city-to-town arbitrage, because the manufacturer could shut down a town retailer who was selling the good to consumers (having bought it from the city) without having paid the franchise fee. As a result, the manufacturer could raise the town franchise fee higher than in our model's equilibrium, setting it to equal the town retailer's entire gross profit. The manufacturer would also set a city franchise fee, but it would equal zero. Since the town two-part tariff would be more profitable, the manufacturer would choose it for values of the transaction cost g higher than in our model. The problem described in this paper would by no means disappear, however, because the more serious form of arbitrage is from towns to cities. Although town retailers would pay the franchise fee and end up with zero profit, it would not follow that their wholesale price could equal marginal cost, because then city retailers (having paid their franchise fees of zero) would buy from town retailers instead of directly from the manufacturer.

Welfare of the manufacturer, the retailers, and the consumers change in different ways depending on the difficulty of arbitrage. In our model, a monopolist sells to competing retailers in cities and monopolistic ones in each small town. With high transportation costs, he uses two-part tariffs in the towns (provided that the transactions cost g in our model is low enough) and linear pricing in the cities. If transportation costs are lower, the manufacturer can still use a two-part tariff in the towns, but he cannot make the unit price equal to marginal cost or city retailers will shift their demand to the towns. We have shown that the manufacturer will set the unit price in the town two-part tariff high enough that city retailers will not buy from the towns. He will also reduce the city price below the monopoly level, as a second way to reduce the temptation to arbitrage. Higher transportation costs could therefore increase welfare by causing manufacturers to shift to lower prices in the towns. Higher transportation costs can help because they cause the manufacturer to shift from linear pricing to the more efficient two-part tariff, and in fact prices can fall in both towns and cities as a result. The welfare effect of higher transportation costs in general is ambiguous, however, because higher welfare in the towns, where prices fall, must be balanced against lower welfare in the cities, where they fall.

7 Appendix

Lemma 1. The retailers' derived demands for the good, $D_a^r(w_a)$ and $D_b^r(w_b)$, are strictly decreasing and concave in wholesale prices w_a and w_b .

Proof: I First, we show that $D_b^r(w_b)$ is strictly decreasing and concave in w_b . Differentiating the derived demand function $D_b^r(w_b) \equiv D_b(p_b^*(w_b))$ two times with respect to w_b we obtain $\frac{dD_b^r(w_b)}{dw_b} = D_b' \cdot \frac{dp_b^*}{dw_b}$ and $\frac{d^2D_b^r(w_b)}{dw_b^2} = D_b'' \cdot \left(\frac{dp_b^*}{dw_b}\right)^2 + D_b' \cdot \frac{d^2p_b^*}{dw_b^2}$. Since in cities the retailers compete in prices, $p_b^*(w_b) = w_b$, and therefore $\frac{dp_b^*(w_b)}{dw_b} = 1$ and $\frac{d^2p_b^*(w_b)}{dw_b^2} = 0$. Thus, we obtain $\frac{dD_b^r(w_b)}{dw_b} < 0$ and $\frac{d^2D_b^r(w_b)}{dw_b^2} \leq 0$.

Second, differentiating two times $D_a^r(w_a) \equiv D_a(p_a^*(w_a))$ with respect to w_a we obtain $\frac{dD_a^r(w_a)}{dw_a} = D_a' \cdot \frac{dp_a^*}{dw_a}$ and $\frac{d^2D_a^r(w_a)}{dw_a^2} = D_a'' \cdot \left(\frac{dp_a^*}{dw_a}\right)^2 + D_a' \cdot \frac{d^2p_a^*}{dw_a^2}$. Then we differentiate the retailer's first order condition with respect to w_a and obtain $\frac{dp_a^*(w_a)}{dw_a} = \frac{D_a'}{2D_a' + (p_a^* - w_a) \cdot D_a''} > 0$ and therefore $\frac{dD_a^r(w_a)}{dw_a} < 0$. Differentiating the retailer's first order condition again with respect to w_a , we obtain $\frac{d^2p_a^*(w_a)}{dw_a^2} = \frac{p_a^* \cdot D_a'' \cdot (2 - 3p_a^*) - (p_a^*)^2 \cdot (p_a^* - w_a) \cdot D_a'''}{2D_a' + (p_a^* - w_a) \cdot D_a''}$. Since $\frac{dp_a^*(w_a)}{dw_a}$ is smaller than $\frac{1}{2}$, non-negative D_a''' is a sufficient condition for $\frac{d^2p_a^*(w_a)}{dw_a^2} \geq 0$ and hence $\frac{d^2D_a^r(w_a)}{dw_a^2} \leq 0$. \square

Proposition 1. In equilibrium, the manufacturer will prevent arbitrage. Wholesale prices will be higher in the city, but by no more than the transportation cost: $0 < (w_b - w_a) \leq t$.

Proof: To show that there is no arbitrage we need to show that the manufacturer's maximum profit in (3) is strictly greater than in (4) or (5). Let w'_a solve the maximization problem in (4). Choose $w_a = w'_a$ and $w_b = w'_a$ in (3) and compare the resulting profit functions in (3) and (4). By Lemma 1, $D_b^r(w_b)$ is strictly decreasing in w_b , so the manufacturer's maximum profit in (3) is strictly greater than in (4), regardless of the choice of I_a .

Let w'_b solve the maximization problem in (5). Choose $w_a = w'_b$ and $w_b = w'_b$ in (3) and compare the resulting profit functions in (3) and (5). Since $D_a^r(w_a)$ is strictly decreasing in w_a by Lemma 1 and $F_a(w_a)$ is decreasing in w_a , the manufacturer's maximum profit in (3) is strictly greater than in (5), regardless of the choice of I_a .

Thus, if arbitrage occurred in any equilibrium, the manufacturer would decrease the absolute value of the wholesale price difference until it did not, which rules out equilibria with a wholesale price difference of $|(w_b - w_a)| > t$ with arbitrage occurring.

Next let us show that $(w_b - w_a) > 0$. First remove the constraint $|(w_b - w_a)| \leq t$ from maximization problem (3) and let (w'_a, w'_b) be the solution. Without the constraint $|(w_b - w_a)| \leq t$, the manufacturer's problem is separable into maximization problems for towns and for cities. If the manufacturer chooses linear pricing, it chooses wholesale prices \tilde{w}_a for towns and p_b^m for cities. Since we have $p_b^m > \tilde{w}_a$, the unconstrained opti-

mal $(w_b - w_a)' = p_b^m - \tilde{w}_a$ is positive. If the manufacturer chooses a two-part tariff, the wholesale price is even lower for towns and the sign of $(w_b - w_a)'$ does not change.

Now, put back the constraint $|w_b - w_a| \leq t$ into maximization problem (3). If $(w_b - w_a)'$ satisfies the constraint $|w_b - w_a| \leq t$, the manufacturer will choose (w'_a, w'_b) and thus $(w_b - w_a)^* > 0$, though possibly that difference is strictly less than t .

If $(w_b - w_a)'$ does not satisfy $|w_b - w_a| \leq t$, positive $(w_b - w_a)'$ implies $(w_b - w_a)' > t$. Since the manufacturer's profit in (3) is twice differentiable and $\frac{\partial^2 \pi}{\partial (w_b - w_a)^2} = n_b \cdot [2D'_b + (w_b - c) \cdot D''_b] < 0$, the profit function is concave in $(w_b - w_a)$. Because of the concavity, $(w_b - w_a)' > t$ implies π is increasing over $-t \leq (w_b - w_a) \leq t$ and the manufacturer will choose $(w_b - w_a)^* = t$. Thus, the price difference is positive and might be as great as t . \square

Proposition 2. (i) There exists a continuous function $\bar{g}(t) > 0$ such that if the negotiation cost is $g > \bar{g}(t)$ then in equilibrium linear pricing is used in towns but if $g < \bar{g}(t)$ a two-part tariff is used (and either might be used if $g = \bar{g}(t)$).

(ii) Let the transportation cost t be smaller than the critical value t_0 if the transaction cost is $g \geq \bar{g}(t)$ or t_1 if $g < \bar{g}(t)$ (as in Figure 1's Regions I and III). Under linear pricing (Region I), the double marginalization problem is magnified: the retail price for towns is higher than under standard double marginalization. Even under a two-part tariff (Region III), the problem in towns cannot be totally eliminated, as the wholesale price for towns will be above marginal cost and the retail price will be higher than the integrated monopoly price.

Proof:

(i) The manufacturer's piecewise maximization problem in (3) to (5) can be reduced to the following, with Proposition 1 justifying the constraint:

$$\begin{aligned} \underset{w_a, w_b}{Max} \pi &= n_a \cdot [(w_a - c) \cdot D'_a(w_a) + I_a \cdot (F_a(w_a) - g)] \\ &+ n_b \cdot [(w_b - c) \cdot D'_b(w_b)] \text{ such that } 0 < (w_b - w_a) \leq t \end{aligned} \quad (6)$$

The Lagrangian function is:

$$\begin{aligned} L &= n_a \cdot [(w_a - c) \cdot D'_a(w_a) + I_a \cdot (F_a(w_a) - g)] \\ &+ n_b \cdot [(w_b - c) \cdot D'_b(w_b)] + \lambda \cdot [t - (w_b - w_a)] \end{aligned} \quad (7)$$

We may regard the wholesale price difference, $(w_b - w_a)$, as a choice variable of the manufacturer. Instead of choosing both wholesale prices, the manufacturer chooses the wholesale price for towns, w_a , and the wholesale price difference, $(w_b - w_a)$. Since the wholesale price difference $(w_b - w_a) > 0$ (Proposition 1) and $w_a \geq c > 0$, the Kuhn-Tucker

conditions are:

$$\begin{aligned} \frac{\partial L}{\partial w_a} &= n_a \cdot [D_a^r(w_a^*) + (w_a^* - c) \cdot D_a^{r'}(w_a^*) \\ &\quad + I_a \cdot F_a'(w_a^*)] \\ &\quad + n_b \cdot [D_b^r(w_b^*) + (w_b^* - c) \cdot D_b^{r'}(w_b^*)] = 0 \end{aligned} \quad (8)$$

$$\frac{\partial L}{\partial (w_b - w_a)} = n_b \cdot [D_b^r(w_b^*) + (w_b^* - c) \cdot D_b^{r'}(w_b^*)] - \lambda^* = 0 \quad (9)$$

and

$$t \geq (w_b^* - w_a^*) \quad (\text{with } t = (w_b^* - w_a^*) \text{ if } \lambda^* > 0) \quad (10)$$

Notice that variables I_a for whether a two-part tariff is used and λ for whether the arbitrage constraint is binding each have values either positive or zero. There are in total four different combinations of I_a and λ . Let us relate the four combinations of policies in Figure 1 to the two critical values of transportation costs, t_0 and t_1 . We will need to show three things about this:

- (a) If the manufacturer chooses linear pricing for towns, then $\lambda^* > 0$ and the arbitrage constraint is binding if and only if $t < t_0$.
- (b) If the manufacturer chooses a two-part tariff for towns, then $\lambda^* > 0$ and the arbitrage constraint is binding if and only if $t < t_1$.
- (c) The manufacturer's optimal choices of wholesale prices, $w_a^*(t)$ and $w_b^*(t)$, are continuous at the two critical values of transportation costs.

(a) Consider linear pricing contracts ($I_a = 0$). If $\lambda^* = 0$, the arbitrage constraint is not binding and the Kuhn-Tucker conditions (8) and (9) give the same first order conditions as unconstrained maximization. Thus, the manufacturer chooses \tilde{w}_a for towns and p_b^m for cities. This implies $(w_b - w_a)^* = p_b^m - \tilde{w}_a$ and according to the definition of t_0 we obtain $(w_b - w_a)^* = t_0$. On the other hand, since $\lambda^* = 0$, (10) implies $t \geq (w_b - w_a)^*$. Combining these two conditions we obtain that $t \geq t_0$. Conversely, if $\lambda^* > 0$, then (9) implies that $n_b \cdot [D_b^r(w_b^*) + (w_b^* - c) \cdot D_b^{r'}(w_b^*)] > 0$. The left hand side of this inequality is $\frac{\partial \pi}{\partial (w_b - w_a)}$. Since π is concave in $(w_b - w_a)$, the inequality $\frac{\partial \pi}{\partial (w_b - w_a)} > 0$ implies $(w_b - w_a)^* < p_b^m - \tilde{w}_a \equiv t_0$. On the other hand, since $\lambda^* > 0$, (10) implies that the constraint is binding and $(w_b - w_a)^* = t$. Combining the two conditions we obtain that $t < t_0$.

(b) The proof is parallel to that for (a), with t_1 being the critical value instead.

(c) By parts (a) and (b), for t greater than the critical values, $\lambda^* = 0$ and the constraint is not necessarily binding. If $I_a = 0$, wholesale prices \tilde{w}_a and p_b^m solve the Kuhn-Tucker conditions (8) and (9), and if $I_a = 1$, then marginal cost c and p_b^m solve both (8) and (9). For t smaller than the critical values, $\lambda^* > 0$ and by (10) $(w_b - w_a)^* = t$. Kuhn-Tucker conditions (8) and (9) become:

$$\begin{aligned} n_a \cdot [D_a^r(w_a^*) + (w_a^* - c) \cdot D_a^{r'}(w_a^*) \\ + I_a \cdot F_a'(w_a^*)] \\ + n_b \cdot [D_b^r(w_a^* + t) + (w_a^* + t - c) \cdot D_b^{r'}(w_a^* + t)] = 0 \end{aligned} \quad (8')$$

and

$$n_b \cdot [D_b^r(w_a^* + t) + (w_a^* + t - c) \cdot D_b^{r'}(w_a^* + t)] - \lambda^* = 0 \quad (9')$$

As the transportation cost t approaches the critical values, $\lambda^* \rightarrow 0$. We can substitute $t = t_0 \equiv p_b^m - \tilde{w}_a$ into (8') and (9') and find that \tilde{w}_a and p_b^m solve these equations when $I_a = 0$. Similarly, substitute $t = t_1 \equiv p_b^m - c$ into (8') and (9') to find that c and p_b^m solve these equations when $I_a = 1$. Thus, the manufacturer's optimal wholesale prices are continuous at the critical values of transportation costs.

The manufacturer's profit $\pi^*(I = 1, g = 0, t)$ cannot be strictly less than $\pi^*(I = 0, t)$. Moreover, $\pi^*(I = 1, g, t)$ is strictly decreasing in g , and therefore we can always find $\bar{g}(t) \geq 0$ satisfying the identity. In other words, $\bar{g}(t)$ exists for all $t > 0$. Equating the manufacturer's optimal profits in (6) for $I = 0$ and $I = 1$, we can solve explicitly $\bar{g}(t)$ in terms of $w^*(I = 0, t)$ and $w^*(I = 1, t)$. By result (c) above, $w_a^*(t)$ and $w_b^*(t)$ are continuous at the critical values for both $I = 0$ and $I = 1$. Thus, $\bar{g}(t)$ is also continuous at the critical values and is in general continuous for $t > 0$.

(ii) In (a), (b) and (c) we have established that if transportation costs are less than the critical values (Figure 1's Regions I and III), the Lagrange multiplier is positive: $\lambda^* > 0$. From Kuhn-Tucker condition (9), $\lambda^* > 0$ implies $n_b \cdot [D_b^r(w_b^*) + (w_b^* - c) \cdot D_b^{r'}(w_b^*)] > 0$. The terms in the square bracket of this inequality is the first derivative of the manufacturer's profit from a city, π_b , with respect to w_b . Since π_b is concave in w_b and p_b^m solves $\frac{\partial \pi_b}{\partial w_b} = 0$, $\frac{\partial \pi_b}{\partial w_b} > 0$ implies $w_b^* < p_b^m$. Substituting $n_b \cdot [D_b^r(w_b^*) + (w_b^* - c) \cdot D_b^{r'}(w_b^*)] > 0$ into (8) we obtain $n_a \cdot [D_a^r(w_a^*) + (w_a^* - c) \cdot D_a^{r'}(w_a^*) + I_a \cdot F_a'(w_a^*)] < 0$. The terms in the square bracket of this last inequality is the first derivative of the manufacturer's profit from a town, π_a , with respect to w_a . If $I_a = 0$ (in Region I), π_a is concave in w_a and \tilde{w}_a solves $\frac{\partial \pi_a}{\partial w_a} = 0$. Thus, $\frac{\partial \pi_a}{\partial w_a} < 0$ implies that $w_a^* > \tilde{w}_a$. Since with linear pricing the retail price is increasing in wholesale price, $w_a^* > \tilde{w}_a$ implies $p_a^* > \tilde{p}_a$. If $I_a = 1$ (as in Region III), then $w_a = c$ solves $\frac{\partial \pi_a}{\partial w_a} = 0$ and the stationary point locally maximizes the manufacturers profit. Then $\frac{\partial \pi_a}{\partial w_a} < 0$ implies $w_a^* > c$ and $p_a^* > p_a^m$. \square

Proposition 3. (a) An increase in the retailer transportation cost t increases social welfare in towns and reduces it in cities if t is small enough to have any effect at all. (b) If t rises to exceed $\bar{t}(g)$, prices fall in both towns and cities and consumer surpluses rise.

(a) Denote social welfare in towns and cities as SW_a and SW_b . The social welfare for towns is

$$SW_a^*(g, t) = n_a \cdot \int_0^{D_a^*(g, t)} [p_a(Q) - c] dQ \quad (11)$$

where $p_a(Q)$ is the inverse demand function and $D_a^*(g, t) \equiv D_a(p_a^*(w_a^*(g, t)))$. Using Leibnitz's rule, we obtain

$$\frac{\partial SW_a^*(g, t)}{\partial t} = n_a \cdot \left\{ \begin{array}{l} \int_0^{D_a^*(g, t)} \frac{\partial}{\partial t} [p_a(Q) - c] dQ \\ + [p_a(D_a^*(g, t)) - c] \cdot \frac{\partial D_a^*(g, t)}{\partial t} \\ - [p_a(0) - c] \cdot \frac{\partial 0}{\partial t} \end{array} \right\} \quad (12)$$

Since $\frac{\partial}{\partial t} [p_a(Q) - c] = 0$, $\frac{\partial 0}{\partial t} = 0$, $p_a(D_a^*(g, t)) = p_a^*$ and $D_a^*(g, t) = D_a^{r*}(g, t) \equiv D_a^r(w_a^*(g, t))$, by using the chain rule we obtain $\frac{\partial D_a^*(g, t)}{\partial t} = \frac{\partial D_a^{r*}(g, t)}{\partial t} = \frac{\partial D_a^r(w_a^*(g, t))}{\partial w_a} \cdot \frac{\partial w_a^*(g, t)}{\partial t}$

and thus

$$\frac{\partial SW_a^*(g, t)}{\partial t} = n_a \cdot (p_a^* - c) \cdot D_a^{r'}(w_a^*(g, t)) \cdot \frac{\partial w_a^*(g, t)}{\partial t} > 0 \quad (13)$$

The effect of the retailer transportation cost on social welfare in towns is positive because $(p_a^* - c) > 0$, $D_a^{r'} < 0$ and $\frac{\partial w_a^*(g, t)}{\partial t} < 0$. Similarly, the effect on social welfare in cities is

$$\frac{\partial SW_b^*(g, t)}{\partial t} = n_b \cdot (p_b^* - c) \cdot D_b^{r'}(w_b^*(g, t)) \cdot \frac{\partial w_b^*(g, t)}{\partial t} < 0 \quad (14)$$

To show that $\frac{\partial w_a^*(g, t)}{\partial t} < 0$ and $\frac{\partial w_b^*(g, t)}{\partial t} > 0$ we differentiate the Kuhn-Tucker conditions in (8), (9) and (10) with respect to t and solve that if the constraint is binding

$$\frac{\partial w_a^*(g, t)}{\partial t} = \frac{-n_b \cdot [2D_b^{r'}(w_b^*) + (w_b^* - c) \cdot D_b^{r''}(w_b^*)]}{\Phi} < 0 \quad (15)$$

$$\frac{\partial w_b^*(g, t)}{\partial t} = \frac{n_a \cdot [2D_a^{r'}(w_a^*) + (w_a^* - c) \cdot D_a^{r''}(w_a^*) + I_a^* \cdot F_a''(w_a^*)]}{\Phi} > 0 \quad (16)$$

where $\Phi = \sum_{j=\text{town, city}} [2n_j \cdot D_j^{r'}(w_j^*) + n_j \cdot D_j^{r''}(w_j^*) \cdot (w_j^* - c)] + n_a \cdot I_a^* \cdot F_a''(w_a^*) < 0$

and the value of I_a^* is 1 if and only if $g < \bar{g}(t)$ and is 0 otherwise; and that if the constraint is not binding then $\frac{\partial w_a^*(g, t)}{\partial t} = \frac{\partial w_b^*(g, t)}{\partial t} = 0$.

(b) As the transportation cost t increases to $\bar{t}(g)$, the manufacturer switches to use a two-part tariff. We want to show that the wholesale prices in both towns and cities are lower when $I_a = 1$ than when $I_a = 0$. There are three separate cases depending on the size of t .

(Case i) $t \geq t_1$. The constraint is not binding. $\lambda^* = 0$ and p_b^m and \tilde{w}_a solve the Kuhn-Tucker conditions in (8) and (9) when $I_a = 0$ and p_b^m and c solve the Kuhn-Tucker conditions when $I_1 = 1$. Thus, the wholesale prices are higher when $I_a = 0$ than when $I_a = 1$.

(Case ii) $t < t_0$. The constraint is binding. $\lambda^* > 0$ and (9) implies $\frac{\partial \pi_b}{\partial w_b}(w_b^*) > 0$, and (8) implies

$$\frac{\partial \pi_b}{\partial w_b}(w_b^*) = -\frac{\partial \pi_a}{\partial w_a}(w_a^*) \quad (17)$$

where $\frac{\partial \pi_b}{\partial w_b}$ is the slope of the manufacturer's profit function for cities and $\frac{\partial \pi_a}{\partial w_a}$ is the slope for towns. We let $w_a^*(I_a = 0)$ and $w_b^*(I_a = 0)$ solve the Kuhn-Tucker conditions when $I_a = 0$, and let $w_a^*(I_a = 1)$ and $w_b^*(I_a = 1)$ solve the conditions when $I_a = 1$. If $I_a = 0$ it follows that $\frac{\partial \pi_a}{\partial w_a}(w_a^*(I_a = 0)) = n_a \cdot [D_a^r(w_a^*(I_a = 0)) + (w_a^*(I_a = 0) - c) \cdot D_a^{r'}(w_a^*(I_a = 0))]$. If $I_a = 1$, $\frac{\partial \pi_a}{\partial w_a}(w_a^*(I_a = 1)) = n_a \cdot [D_a^r(w_a^*(I_a = 1)) + (w_a^*(I_a = 1) - c) \cdot D_a^{r'}(w_a^*(I_a = 1))] +$

$n_a \cdot 1 \cdot F'_a(w_a^*(I_a = 1))$ and thus $\frac{\partial \pi_a}{\partial w_a}|_{I_a=1}(w_a^*(I_a = 1)) = \frac{\partial \pi_a}{\partial w_a}|_{I_a=0}(w_a^*(I_a = 1)) + n_a \cdot F'_a(w_a^*(I_a = 1))$. Since $n_a \cdot F'_a(w_a^*(I_a = 1)) < 0$, we obtain

$$\frac{\partial \pi_a}{\partial w_a}|_{I_a=1}(w_a^*(I_a = 1)) < \frac{\partial \pi_a}{\partial w_a}|_{I_a=0}(w_a^*(I_a = 1)) \quad (18)$$

If $I_a = 0$, (17) implies

$$\frac{\partial \pi_b}{\partial w_b}(w_b^*(I_a = 0)) = -\frac{\partial \pi_a}{\partial w_a}|_{I_a=0}(w_a^*(I_a = 0)) \quad (19)$$

and if $I_a = 1$, (17) implies $\frac{\partial \pi_b}{\partial w_b}(w_b^*(I_a = 1)) = -\frac{\partial \pi_a}{\partial w_a}|_{I_a=1}(w_a^*(I_a = 1))$. By (18), we obtain

$$\frac{\partial \pi_b}{\partial w_b}(w_b^*(I_a = 1)) > -\frac{\partial \pi_a}{\partial w_a}|_{I_a=0}(w_a^*(I_a = 1)) \quad (20)$$

Since $\lambda^* > 0$, (10) implies $w_b^* - w_a^* = t$ regardless of $I_a = 1$ or $I_a = 0$. Thus, $w_b^*(I_a = 1) - w_a^*(I_a = 1) = t$ and $w_b^*(I_a = 0) - w_a^*(I_a = 0) = t$. Combining the two conditions we obtain

$$w_b^*(I_a = 1) - w_b^*(I_a = 0) = w_a^*(I_a = 1) - w_a^*(I_a = 0) \quad (21)$$

There are two possible subcases for the wholesale price in cities — $w_b^*(I_a = 1) \geq w_b^*(I_a = 0)$ and $w_b^*(I_a = 1) < w_b^*(I_a = 0)$. To show that in fact the latter is correct, suppose that the opposite is true and thus $w_b^*(I_a = 1) \geq w_b^*(I_a = 0)$. Since the constraint is binding, by Proposition 1 $w_a^* > \tilde{w}_a$ and $w_b^* < p_b^m$. Since π_b is strictly concave and $\frac{\partial \pi_b}{\partial w_b}(w_b^*) > 0$, $w_b^*(I_a = 1) \geq w_b^*(I_a = 0)$ implies

$$\frac{\partial \pi_b}{\partial w_b}(w_b^*(I_a = 0)) \geq \frac{\partial \pi_b}{\partial w_{city}}(w_b^*(I_a = 1)) \quad (22)$$

Substituting (19) and (20) into (22) and rearranging the terms we obtain

$$\frac{\partial \pi_a}{\partial w_a}|_{I_a=0}(w_a^*(I_a = 1)) > \frac{\partial \pi_{town}}{\partial w_a}|_{I_a=0}(w_a^*(I_a = 0)) \quad (23)$$

Since $\pi_a|_{I_a=0}$ is also strictly concave and $\frac{\partial \pi_a}{\partial w_a}(w_a^*) < 0$, (23) implies $w_a^*(I_a = 1) < w_a^*(I_a = 0)$. This implies that the right hand side of (21) is strictly negative. However, we have assumed that $w_b^*(I_a = 1) \geq w_b^*(I_a = 0)$, which implies the left hand side of (21) is non-negative. This is a contradiction and we can thus conclude that $w_b^*(I_a = 0) > w_b^*(I_a = 1)$. Since $w_b^*(I_a = 1) = w_a^*(I_a = 1) + t$ and $w_b^*(I_a = 0) = w_a^*(I_a = 0) + t$, we obtain also $w_a^*(I_a = 0) > w_a^*(I_a = 1)$.

(Case iii) $t_0 \leq t < t_1$. The constraint is binding for $I_a = 1$ but not for $I_a = 0$. If $I_a = 0$, the manufacturer chooses wholesale prices \tilde{w}_a for towns and p_b^m for cities. On the other hand, if $I_a = 1$, by Proposition 1 $w_b^* < p_b^m$ and therefore $w_b^*(I_a = 0) > w_b^*(I_a = 1)$. From the proof in part (ii) we know that $w_a^*(I_a = 1) < w_a^*(I_a = 0)$ for all $t < t_0$. By Lemma

4 the manufacturer's optimal wholesale prices are continuous at the critical values of the transportation cost and therefore $w_a^*(I_a = 1) < w_a^*(I_a = 0)$ at $t = t_0$. From the proof in part (a) we know that $\frac{\partial w_a^*(g,t)}{\partial t} < 0$ if the constraint is binding and $\frac{\partial w_a^*(g,t)}{\partial t} = 0$ if the constraint is not binding. These conditions together imply $w_a^*(I_a = 1) < w_a^*(I_a = 0)$ for $t_0 \leq t < t_1$.

We have shown that the wholesale prices are lower when $I_a = 1$ than when $I_a = 0$. Retail prices are strictly increasing in the wholesale prices, so all prices will fall if t rises enough to induce a switch to two-part tariffs in the towns. \square

8 References

Baldenius, Tim, and Stefan Reichelstein, 2006, "External and Internal Pricing in Multidivisional Firms," *Journal of Accounting Research*, Vol. 44, No. 1, pp. 1-28.

Baldenius, Tim, Stefan Reichelstein and Savita A. Sahay, 1999, "Negotiated versus Cost-Based Transfer Pricing," *Review of Accounting Studies*, 4, 67-91.

Chung, Barick, 2007, *Essays on Supermarket Pricing and Coupon Strategies*. Ph.D. dissertation, Indiana University.

Cook, Paul W., 1955, "Decentralization and the Transfer-Price Problem," *The Journal of Business*, Vol. 28, No. 2, pp. 87-94.

Dana, James, D. Jr. and Kathryn E. Spier, 2001, "Revenue Sharing and Vertical Control in the Video Renting Industry," *Journal of Industrial Economics*, Vol. 49, No. 3, pp.223-245.

Hirshleifer, Jack, 1956, "On the Economics of Transfer Pricing," *The Journal of Business*, Vol. 29, No. 3, pp. 172-184.

Kolay, Sreya, Greg Shaffer and Janusz A. Ordover, 2004, "All-Units Discounts in Retail Contracts," *Journal of Economics and Management Strategy*, Volume 13, No. 3, pp. 429-459.

Mathewson, G.F., and R. A. Winter, 1983a, "Vertical Integration by Contractual Restraints in Spatial Markets," *The Journal of Business*, Vol. 56, No. 4, pp. 497-517.

Mathewson, G.F., and R. A. Winter, 1983b, "The Incentives for Resale Price Maintenance under Imperfect Information," *Economic Inquiry*, Vol. 21, No. 3, pp. 337- 348.

Mathewson, G.F., and R. A. Winter, 1984, "An Economic Theory of Vertical Restraints," *The RAND Journal of Economics*, Vol. 15, pp. 27-38.

Menge, John A., 1961, "The Backward Art of Interdivisional Transfer Pricing," *The Journal of Industrial Economics*, Vol. 9, No. 3, pp. 215-232.

Milliou, C., and E. Petrakis 2007, "Upstream Horizontal Mergers, Vertical Contracts and Bargaining," *International Journal of Industrial Organization*, Vol. 25, pp. 963-987.

Rasmusen, Eric B., J. Mark Ramseyer and John S. Wiley, Jr., 1991, "Naked Exclusion," *The American Economic Review*, Vol. 81, No. 5, pp. 1137-1145.

Rey, Patrick, and Jean Tirole, 1986, "The Logic of Vertical Restraints," *The American Economic Review*, Vol. 76, No. 5, pp. 921-939.

Rey, Patrick, and Joseph Stiglitz, 1988, "Vertical Restraints and Producers' Competition," *NBER Working Paper Series*, Working Paper No. 2601.

Rey, Patrick, and Joseph Stiglitz, 1994, "The Role of Exclusive Territories in Producers' Competition," *NBER Working Paper Series*, Working Paper No. 4618.

Rey, Patrick, 2003, "The Economics of Vertical Restraints." http://esnie.u-paris10.fr/pdf/textes_2004/rey_vertical-res.pdf. Accessed on May 10, 2007.

Rey, Patrick, and Jean Tirole, 2003, "A Primer on Foreclosure," forthcoming, in *Handbook of Industrial Organization*, Volume 3, Mark Armstrong and Robert S. Porter (eds), North Holland. http://esnie.u-paris10.fr/pdf/textes_2004/rey_primer.pdf. Accessed on May 10, 2007.

Salop, Steven C. and David T. Scheffman, 1983, "Raising Rivals' Costs," *American Economic Review*, Vol. 73, pp. 267-71.

Spengler, Joseph, 1950, "Vertical Integration and Antitrust Policy," *Journal of Political Economy*, Vol. 58, pp. 347-352.

Tirole, Jean, 1988, *The Theory of Industrial Organization*, MIT Press.

Wilson, Robert B., 1993, *Nonlinear Pricing*, Oxford University Press, New York.

Vaysman, Weger, 1998, "A Model of Negotiated Transfer Pricing," *Journal of Accounting and Economics*, Vol. 25, pp. 349-384.

THINGS TO DO

1. Read Milliou, C., and E. Petrakis 2007, "Upstream Horizontal Mergers, Vertical Contracts and Bargaining," *International Journal of Industrial Organization*, Vol. 25, pp. 963-987. and add more current references. Take, for instance, the recent contributions of Marx/Shaffer (in Rand) as well as of Rey/Thal/Verge on "three-part-tariffs". What about the price discrimination on inputs literature, the referee asks.

2. Discuss what happens if other pricing schemes besides linear pricing or a two part tariff could be used. Quantity requirements and retail price ceiling.

3. Discuss stylized facts. Where does this model apply?

Maybe we could send this to the *The B.E. Journal of Economic Analysis and Policy*. They have quick turnaround and publication at several levels, which would be useful for you.