

July 2014 ■ RFF DP 14-21

# The Hotelling Model with Multiple Demands

---

Gérard Gaudet and Stephen W. Salant

1616 P St. NW  
Washington, DC 20036  
202-328-5000 [www.rff.org](http://www.rff.org)



# The Hotelling Model with Multiple Demands<sup>1</sup>

G rard Gaudet

Stephen W. Salant<sup>2</sup>

July, 2014

<sup>1</sup>Forthcoming in *Handbook on the Economics of Natural Resources*, eds Robert Halvorsen and Dave Layton, Cheltenham, U.K.: Edward Elgar Publ.

<sup>2</sup>G rard Gaudet is professor emeritus in the Department of Economics, University of Montreal and research fellow at CIREQ ([gerard.gaudet@umontreal.ca](mailto:gerard.gaudet@umontreal.ca)); Stephen W. Salant is a professor in the Department of Economics, University of Michigan and a nonresident fellow at Resources for the Future (RFF) ([ssalant@umich.edu](mailto:ssalant@umich.edu)).

## Abstract

*The purpose of this chapter is to provide an elementary introduction to the nonrenewable resource model with multiple demand curves. The theoretical literature following Hotelling (1931) assumed that all energy needs are satisfied by one type of resource (e.g. “oil”), extractible at different per-unit costs. This formulation implicitly assumes that all users are the same distance from each resource pool, that all users are subject to the same regulations, and that motorist users can switch as easily from liquid fossil fuels to coal as electric utilities can. These assumptions imply, as Herfindahl (1967) showed, that in competitive equilibrium all users will exhaust a lower cost resource completely before beginning to extract a higher cost resource: simultaneous extraction of different grades of oil or of oil and coal should never occur. In trying to apply the single-demand curve model during the last twenty years, several teams of authors have independently found a need to generalize it to account for users differing in their (1) location, (2) regulatory environment, or (3) resource needs. Each research team found that Herfindahl’s strong, unrealistic conclusion disappears in the generalized model; in its place, a weaker Herfindahl result emerges. Since each research team focussed on a different application, however, it has not always been clear that everyone has been describing the same generalized model. Our goal is to integrate the findings of these teams and to exposit the generalized model in a form which is easily accessible.*

## 1 Introduction

The purpose of this chapter is to provide an elementary introduction to the nonrenewable resource model with multiple demands. The theoretical literature following Hotelling (1931) assumed that each grade of a resource can be shipped to users all over the world at the same cost, that all users are subject to the same taxes and regulations; and that motorists can switch as easily from liquid fossil fuels to coal as electric utilities can.

As Herfindahl (1967) showed, these assumptions imply that in competitive equilibrium all users will exhaust a lower cost resource completely before beginning to extract a higher cost resource; moreover, users will switch to this higher cost resource at the same moment so simultaneous extraction of different grades of oil or of oil and coal will never occur. That is, Herfindahl's conclusion about the order of extraction follows from his assumption of a single demand curve.

But this assumption is highly unrealistic. A glance at the morning paper tells us that not all users pay the same cost to access a given source of energy. Consider a nonrenewable source like Canadian oil sands. At current prices, if the proposed Keystone pipeline transported tar sands oil from Alberta south to refineries in the Gulf Coast of Texas, it would be the cheapest way for properly equipped refineries to produce gasoline. In the absence of a pipeline, however, the oil sands will have to be shipped by rail and even properly equipped refineries would find using lighter oil from the Middle East to be the cheapest input to refine into gasoline.

Even with the Keystone pipeline, however, some refineries are not equipped to refine oil sands into gasoline and must incur an additional cost per unit (a "conversion cost") to refine oil sands. For this reason, properly equipped refineries may find oil sands to be the cheapest source for producing gasoline while other refineries will continue to find the lighter oil from Saudi Arabia cheapest to refine.

However, not all users of this same Saudi oil pay the same price because they may live in different jurisdictions where taxes and environmental regulations differ. Gasoline refined from Saudi oil costs Europeans more than Americans in large part because Europeans are

subject to higher gasoline taxes. As a result, it may be cheapest for someone seeking to travel within the city of Los Angeles to use a car powered by gasoline but within Paris to use a subway powered by electricity generated from nuclear power.

The cost of accessing any given source of energy differs because users incur different transportation costs, conversion costs, or taxes/regulations—or a combination of these costs. When cooking, heating and lighting their homes or traveling around town, people typically choose the cheapest alternative available to them. Because different alternatives may be cheapest to different users, however, people make different choices at any given time.

Because users accessing a given energy source pay different costs, Hotelling’s assumption of a single demand curve misses important features of world energy markets and leads to predictions that are easily refuted. In trying to apply this single-demand curve model during the last twenty years, several teams of authors have independently found a need to generalize it to account for users differing in their (1) location (Gaudet, Moreaux and Salant, 2001; Ley, Macauley and Salant, 2000, 2002), (2) regulatory environment (Fischer and Salant, 2014; Hoel, 2011), or (3) resource needs (Chakravorty and Krulce, 1994; Chakravorty, Krulce and Roumasset, 2005; Im, Chakravorty and Roumasset, 2006). Each research team found that in the generalized model, users switch from one resource to another at different times and some users skip the use of some resources entirely; and yet, at the same time, a generalized Herfindahl result emerges. Since each research team focused on a different application, however, it has not always been clear that everyone has been describing the *same* generalized model. Our goal is to integrate the findings of these teams and to exposit their generalized model in a form which is easily accessible.

We proceed as follows. In the next section we review the properties of the standard partial equilibrium model of competitive equilibrium when firms have constant per-unit costs which may differ and may change over time due to exogenous technical change. We consider both the case where the cost is the same of supplying every consumer using a given technology and where these supply costs differ. In Section 3, we introduce cumulative production constraints on some of the sources of supply. We show that the Herfindahl result that two sources of

supply will never be used at the same time disappears when we relax the standard but unrealistic assumption that a source of supply can provide the resource to everyone at the same per-unit cost. In Section 4, we decompose the per-unit cost of supplying different users from a given source and clarify its various interpretations. Section 5 concludes the paper.

## 2 Unconstrained Production

Consider a static model where the needs of  $J$  consumers are satisfied from  $I$  sources. Let  $w_{ij}$  be the per-unit cost of supplying user  $j$  with source  $i$  and  $q_{ij}$  be the amount supplied to him from that source. We consider in turn the single user case and the multiple users case. Multiple users can be either heterogeneous or homogeneous. In the latter case, all  $J$  users can be supplied from the same source  $i$  at the same per-unit cost.

### 2.1 Single User ( $J = 1$ )

In this subsection, we assume that there is a single user ( $J = 1$ ) and therefore dispense with the  $j$  subscript. Hence,  $w_{ij} = w_i$  and  $q_{ij} = q_i$ . Let  $Q = \sum_i q_i$  denote aggregate consumption from the  $I$  sources. Index the sources so a lower cost source has a lower index:  $w_i < w_{i'}$  for  $i < i'$ . Therefore, source 1 has the lowest per-unit cost,  $w_1$ .

Suppose a planner chooses to obtain  $Q > 0$  units of output, some from a source other than source 1. Then, the planner cannot be optimizing since marginally less can be obtained from the higher cost source and replaced by the same amount from source 1, thereby reducing the cost of obtaining the same quantity  $Q$ . If a plan is optimal, it must be obtained at least cost and that entails using only source 1. In addition, the marginal benefit of consuming more from source 1 must equal  $w_1$ , the marginal cost of obtaining it:  $U'(Q^*) = w_1$ . Assuming that  $U(\cdot)$  is a strictly concave function with  $U'(0) > w_1$ , a solution to this equation,  $Q^*$ , exists and is unique.

This planning solution can be decentralized as a competitive equilibrium. In this formulation, the owner of technology  $i$  takes the market price  $P$  as given and maximizes his profit. The market price induces aggregate supply equal to aggregate demand. If the inverse

demand curve is  $P(Q) = U'(Q)$ , then in the planning solution,  $Q^* = q_1^*$ , and in the competitive solution  $P = P(Q^*) = U'(Q^*) = w_1$ . Given this price, higher cost producers will find producing nothing to be profit-maximizing.

## 2.2 Multiple Users ( $J > 1$ )

Suppose instead that the planner has to satisfy a set of  $J > 1$  users. Let  $w_{ij}$  denote the exogenous, per-unit cost of obtaining one unit from source  $i$  and delivering it to user  $j$  and  $q_{ij} \geq 0$  the endogenous number of units user  $j$  obtains from source  $i$ . A set of users is either heterogeneous or homogeneous. Users are defined as heterogeneous if and only if there exist at least two users ( $j, j' \in J$ ) who cannot receive goods from some source ( $i \in I$ ) at the *same* per-unit cost: that is,  $w_{ij} \neq w_{ij'}$  for some  $i, j, j'$ . For our analysis, it does not matter whether the  $j$  users have the same or differing utility functions as long as each of them is strictly concave ( $U_j''(\cdot) < 0$ ), strictly increasing ( $U_j'(\cdot) > 0$ ), and passes through the origin ( $U_j(0) = 0$ ) with a finite slope larger than the cheapest source available to  $j$  ( $U_j'(0) > \min_i w_{ij}$ ).<sup>1</sup>

The planner's goal is to maximize:

$$\max_{q_{ij} \geq 0} U_1 \left( \sum_i q_{i1} \right) + \dots + U_J \left( \sum_i q_{iJ} \right) - \sum_j \sum_i w_{ij} q_{ij}, \quad (1)$$

### 2.2.1 Heterogeneous Users

Let  $Q_j$  denote the total number of units the planner sends to user  $j$  from the  $I$  sources:  $Q_j = \sum_{i=1}^I q_{ij}$ . Clearly, no plan where user  $j$  can obtain his  $Q_j$  units at lower cost by altering his sourcing can be optimal since altering his sources has no effect on the other  $J - 1$  users. Hence, in the optimal plan each user will obtain all of his output from the source which is cheapest: user  $j$  will choose source  $i$  such that  $w_{ij} < w_{i'j}$ . But since users are heterogeneous, source  $i$  is not necessarily cheapest for user  $j'$ . That is, *different* users may

---

<sup>1</sup>Sources may need to be converted in order to satisfy a particular use ( $j$ ) but, after conversion, they are assumed to be perfect substitutes. Otherwise the utility ( $U_j$ ) would depend, even after conversion, on a vector of consumptions instead of a single aggregate. Even with linear costs, users would then simultaneously access multiple sources since they would be imperfect substitutes; they would also utilize multiple sources if the *flow* of production from each source was capacity constrained.

simultaneously rely on different sources.

To illustrate, suppose two users can satisfy their needs from three sources. Hence, there will be six exogenous cost parameters. We array them in the matrix  $[w_{ij}]$ , which has  $I = 3$  rows and  $J = 2$  columns. Suppose user 1 (respectively, user 2) can obtain the same good from three sources at the per-unit costs arrayed in column 1 (respectively, column 2):

$$[w_{ij}] = \begin{pmatrix} 2 & 4 \\ 1 & 5 \\ 3 & 3 \end{pmatrix}.$$

Clearly, user 1 will rely on source 2 since, at \$1 per unit, it is cheapest for him. At the same time, user 2 will rely on source 3 since it is cheapest at \$3 per unit. Indeed, since it is optimal for user  $j$  to obtain  $Q_j^*$  such that he would gain no more from an additional unit than it would cost to obtain it, user 1 would obtain  $q_{21}^*$  units from source 2 where  $U_1'(q_{21}^*) = 1$  while user 2 would obtain  $q_{32}^*$  from source 3 where  $U_2'(q_{32}^*) = 3$ .

A problem of any dimension can be solved in the same way. To determine which source to use for a given consumer, the planner would review the entries in the column pertaining to that user and select the source with the smallest cost. He would then obtain units for that user from that source until the marginal benefit of acquiring an additional unit equalled the marginal cost for that user of acquiring it.

### 2.2.2 Homogeneous Users

If the set of  $J$  users is homogeneous, then the cost per unit of supplying any of the users is the same:  $w_{ij} = w_{ij'}$  for all  $i, j, j'$ . In that case every entry in a given row of the  $[w_{ij}]$  matrix is the same although the numbers in different rows differ. So each user  $j$  has the identical column vector listing the per-unit costs of the sources available to him, and the planner will satisfy the needs of all  $J$  users from the same source, the cheapest one. If in addition the users are assumed to have identical utility functions, then this case is the exact equivalent of the single user case of Section 2.1. From another perspective, in the single user case, the



matrix  $[w_{ij}]$  consists of a single column.

### 2.3 Technical Change

If there is exogenous technical change, then the exogenous per-unit costs will depend on time:  $w_{ij}(t)$ . At time  $t$ , the planner will still satisfy the needs of user  $j$  entirely from the source which is cheapest at time  $t$ , a consequence of the linearity of costs in (1). But some other source may be experiencing faster cost-reducing technical change and if it becomes cheapest in the future, it will be utilized instead to satisfy the needs of user  $j$ .

If we introduce such dynamics into the homogeneous users case, where entries in a row of the  $[w_{ij}(t)]$  matrix are identical, then *every* user will rely on the same source at date  $t$  and if one user subsequently switches to another source, every user will switch to that same resource simultaneously. Clearly, this behavior is unrealistic. We do not see everyone in the world getting the same good or service from the same technology and switching at the identical moment to another technology if it becomes cheapest. But this prediction follows from the unrealistic assumption that at any date any of the  $J$  users can acquire a unit of the good from a given source  $i$  at the same cost. Rather than discarding the useful partial equilibrium framework with constant per-unit costs because it gives unrealistic predictions when this unrealistic assumption is made, it seems more sensible to replace that assumption with something more realistic.

If we introduce exogenous technological change into the heterogeneous users case, where entries in a row of  $[w_{ij}(t)]$  *differ*, it is again optimal for the planner to rely entirely on the lowest cost source at  $t$  when satisfying the needs of user  $j$ . But as we have seen in the example above, what is lowest cost for one user need not be lowest cost for another. And if a different source becomes cheapest for the first user at some future date, he will switch to that source. Other users will not switch at the same moment since that source may not at that moment become cheapest for them.

We conclude this section by noting that the planning problem(s) we have been discussing can all be decentralized as competitive equilibria. Consider the most general case, where

users are heterogeneous and the exogenous per-unit costs differ and change due to technical progress. The planning problem in (1) can be modified as:

$$\max_{q_{ij}(t) \geq 0} U_1 \left( \sum_i q_{i1}(t) \right) + \dots + U_J \left( \sum_i q_{iJ}(t) \right) - \sum_j \sum_i w_{ij}(t) q_{ij}(t), \quad (2)$$

At time  $t$  if the planner sends  $q_{ij}(t)$  from source  $i$  to user  $j$ , then in the decentralized solution the price in market  $j$  is  $P^j(t) = w_{ij}(t)$ .

### 3 Constraints on Cumulative Production: the Hotelling Model

In this section, we constrain cumulative supply from some or all of the  $I$  sources, a modification first introduced by Hotelling (1931). To begin we assume that every source is constrained:  $S_i \geq \sum_j \int_{t=0}^{\infty} q_{ij}(t) dt$  for all  $i = 1, \dots, I$ . This would apply to the case where every source is a nonrenewable like oil. In Section 3.3, we will assume that a subset of the sources have no such constraint. Typical examples are solar or geothermal energy. These are referred to in the Hotelling literature as “backstops.” To simplify we revert to our assumption that the matrix  $[w_{ij}]$  is stationary, taking up the nonstationary case in Section 3.4.

#### 3.1 Single User ( $J = 1$ )

In this subsection, we return to the case of a single user ( $J = 1$ ) and therefore dispense again with the  $j$  subscript. Hence,  $w_{ij} = w_i$  and  $q_{ij}(t) = q_i(t)$ . Let  $Q(t) = \sum_i q_i(t)$  denote aggregate consumption from the  $I$  sources. Index the sources so a lower cost source has a lower index:  $w_i < w_{i'}$  for  $i < i'$ . Therefore, source 1 has the lowest per-unit cost,  $w_1$ . In terms of the matrix  $[w_{ij}]$ , it consists of a single column with  $I$  components which increase as one moves from the first to the last component.

It was Hotelling’s insight that  $w_i$  is *no longer the full cost* at time  $t$  of using another unit of resource  $i$  provided the reserve constraint is binding. For if  $S_i = \int_{t=0}^{\infty} q_i(t) dt$ , then providing more of resource  $i$  at time  $t$  requires cutting back at some other time(s). Suppose a program is optimal. Then whenever source  $i$  is utilized ( $q_i(t) > 0$ ), the net present value of

using more of it must be the same for every  $t$ . For if the net present values at  $t$  and  $t' \neq t$  of marginally increasing use of source  $i$  were not the same, the planner could marginally expand usage on a date with a strictly higher net present value and could contract by an equal amount usage on the date with the strictly lower net present value. This perturbation would strictly increase the planner's payoff and cannot occur if his program is optimal. Denote the common net present value of marginally increasing source  $i$  as  $\lambda_i$ . For any  $t$  such that  $q_i(t) > 0$ ,  $U'(q_i(t)) = w_i + \lambda_i e^{rt}$  for all  $i, t$ . Moreover, whenever  $q_i(t) = 0$ ,  $U'(0) \leq w_i + \lambda_i e^{rt}$ . For otherwise, a gain in net present value could be achieved by marginally expanding usage on the date when nothing is used and contracting it equally on date(s) where  $q_i(t) > 0$ .

Hence, at  $t$  the analog of the exogenous per-unit cost  $w_i$  in Section 2.1 is the augmented per-unit cost:  $w_i + \lambda_i e^{rt}$ . But since the augmented per-unit cost changes over time, the closest analog is the model with exogenous technical change in Section 2.3. Once we assign the  $I$  multipliers, we can construct the augmented matrix  $[w_i + \lambda_i e^{rt}]$ . In the case of a single user it will consist of a single column. The planner will use only the source with the lowest augmented per-unit cost. But when a different source becomes cheapest, the planner will immediately switch to it.

It was left to Herfindahl (1967)) to clarify what intertemporal cost-minimization implies in the single user case. If  $w_i < U'(0)$  for all  $i$  then the planner will want to exhaust each source eventually and so  $\lambda_i > 0$  for all  $i$ . Hence, the multipliers must be assigned so that each source  $i$  has the least augmented per-unit cost for some interval of time.

This limits what multipliers can be assigned. For example, if  $w_i < w_{i'}$  one cannot assign  $\lambda_i \leq \lambda_{i'}$  or source  $i'$  will never have the smallest augmented per-unit cost and will never be used. Hence, one must set  $\lambda_i > \lambda_{i'}$ . Moreover, one cannot assign the multipliers so that  $\lambda_i + w_i > \lambda_{i'} + w_{i'}$  or source  $i$  will never be used. For every source to be least cost over some time interval, therefore,  $\lambda_1 > \dots > \lambda_I$  and  $\lambda_1 + w_1 < \dots < \lambda_I + w_I$ . Given our indexing convention, the planner will exhaust source  $i$  before beginning to utilize source  $i + 1$ .

This makes perfect sense. For the planner cannot be optimizing if he could obtain the same quantity path in a cheaper manner. Suppose at some date  $t$  he was (mistakenly)

taking some reserves from pool  $i'$  despite its strictly higher cost *before* exhausting pool  $i$ , with  $w_i < w_{i'}$ . Suppose he planned to subsequently utilize source  $i$  after some interval of time  $\Delta > 0$ . Then he could lower his costs by marginally reducing his current usage of pool  $i'$  and replacing it with an equal quantity from pool  $i$ ; at date  $t + \Delta$  he would then provide the marginal amount of  $i'$  that he had saved during the interval of time  $\Delta$  so that the same quantity was restored at that date, although the replacement would come from pool  $i'$  instead of pool  $i$ . This perturbation is feasible and would lower costs since, although the current cost reduction is equal to the cost increase that occurs after the interval  $\Delta$ , the former dominates because of discounting: the discounted value of the cost change is:  $-w_{i'} + w_i + e^{-r\Delta}(w_{i'} - w_i) = (w_i - w_{i'})(1 - e^{-r\Delta}) < 0$  where the inequality follows since  $w_{i'} > w_i$ .

## 3.2 Multiple Users

### 3.2.1 Heterogeneous Users

Suppose the planner uses the  $I$  sources to satisfy the needs of a heterogeneous set of  $J$  users with preferences  $U_j$ . That is, his goal is to maximize:

$$\max_{q_{ij}(t) \geq 0} \int_{t=0}^{\infty} e^{-rt} \left[ U_1 \left( \sum_i q_{i1}(t) \right) + \dots + U_J \left( \sum_i q_{iJ}(t) \right) - \sum_j \sum_i w_{ij} q_{ij}(t) \right] dt \quad (3)$$

subject to the constraint that  $S_i \geq \sum_{j=1}^J \int_{t=0}^{\infty} q_{ij}(t) dt$  for  $i = 1, \dots, I$ .

As before, if source  $i$  is exhausted in the optimal plan, then the full cost of using more of it at  $t$  must reflect the opportunity cost of using less of it at other times. As before, if a program is optimal, every user of source  $i$  must derive from it the same present value of marginal net benefit. Denoting that common value as  $\lambda_i$ , then in the optimal plan,  $U'_j(q_{ij}) = w_{ij} + \lambda_i e^{rt}$  for every  $i, j, t$  provided  $q_{ij}(t) > 0$ ; moreover, when  $q_{ij}(t) = 0$ ,  $U'_j(0) \leq w_{ij} + \lambda_i e^{rt}$ .

A unique set of multipliers is associated with the solution to the planning problem provided the objective function is smooth and strictly concave. Every resource  $i$  such that  $w_{ij} < U'_j(0)$  for some  $j$  will have cumulative usage equal to the exogenous upper bound

( $S_i > 0$ ) and a strictly positive multiplier ( $\lambda_i > 0$ ). Every resource  $i$  such that  $w_{ij} > U'_j(0)$  for every  $j$  will be unused and will have a multiplier of zero ( $\lambda_i = 0$ ).

In the optimal plan, user  $j$  relies at time  $t$  entirely on the resource with the lowest augmented per-unit cost. One can determine the source which user  $j$  will rely on at time  $t$  by constructing the lower envelope of his augmented per-unit cost paths:  $\min\{w_{1j} + \lambda_1 e^{rt}, \dots, w_{Ij} + \lambda_I e^{rt}\}$ . As long as every source with a strictly positive multiplier is used by at least one of the  $J$  users over some time interval, we can construct a planning problem that these paths solve. To construct it, one simply computes the cumulative usage over time of each source by the  $J$  users and assumes the initial reserve of each source ( $S_i$  for  $i = 1, \dots, I$ ) is equal to the usage computed for that source.

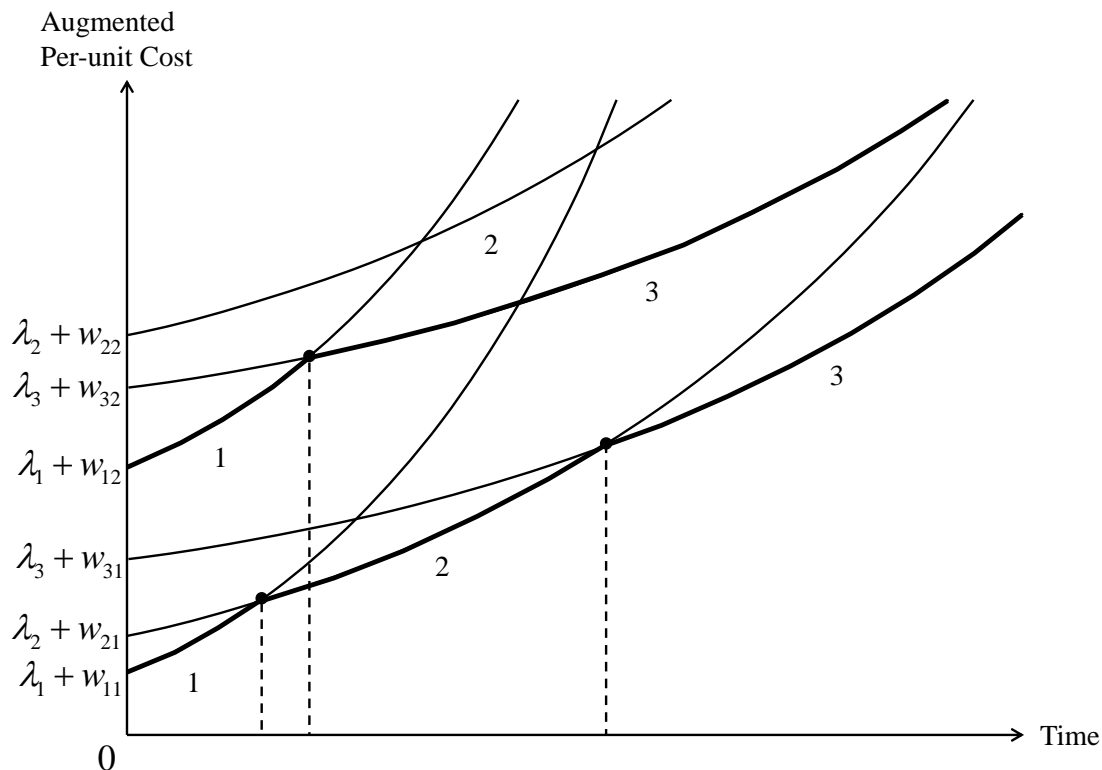
Figures 1 and 2, therefore, depict the solutions to two planning problems with  $J = 2$  and  $I = 3$ . We use them to illustrate behaviors that occur in the multiple demand model with heterogeneous users. In Figure 1, user 1 is assigned source 1, followed by source 2 and then source 3. User 2 skips from source 1 to source 3 without ever being assigned source 2. So in the multiple demand case, users need not utilize the same sources. Moreover, although source 1 has the lowest exogenous per-unit cost for user 1 ( $w_{11} = \min\{w_{11}, w_{21}, w_{31}\}$ ), the planner switches user 1 to source 2 *before* source 1 is exhausted. This is efficient because allocating additional units of source 1 to user 2 has a higher net present value than allocating any more of them to user 1. The “invisible hand” induces user 1 to abandon source 1 when reserves with lower exogenous per-unit costs remain by making the full price of source 2 (the exogenous per-unit cost *plus* the endogenous multiplier, capitalized to  $t$ ) smaller than the full price of source 1. Note that in this example different sources are simultaneously in use: at the time that user 1 is relying on source 2, user 2 is relying on source 3. In Figure 2, source 2 is utilized by user 1, then utilized by nobody, and then utilized by user 2.<sup>2</sup>

The solution of the planning problem and its competitive analog displays what we have elsewhere called the “Generalized Herfindahl Principle” (Gaudet et al., 2001, p. 1153). Any user  $j$  who is assigned source  $i$  early in the program and  $i'$  later in the program must have

---

<sup>2</sup>The resumption of usage of a source by one user after it has been abandoned by another user and has remained idle is illustrated in Figure 1 of Im et al. (2006).

Figure 1: Simultaneous Use of Different Resources and User 2 Never Uses Source 2

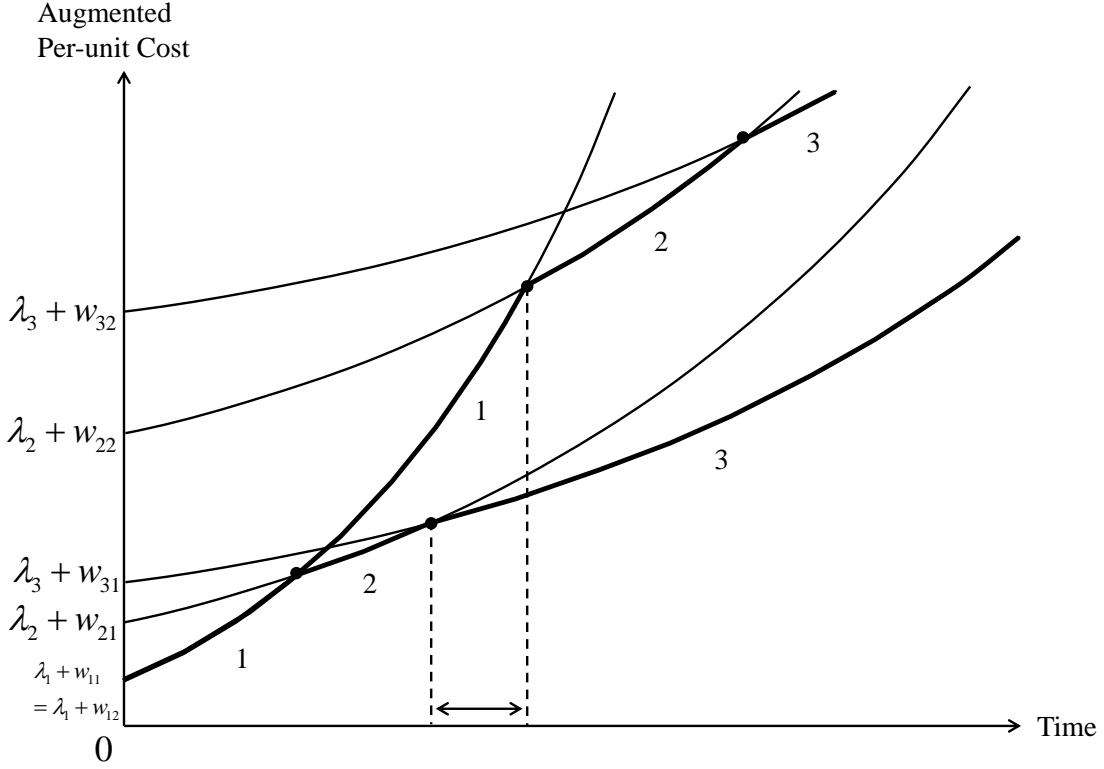


$w_{ij} < w_{i'j}$ . For using sources in that order can only occur if  $\lambda_i > \lambda_{i'}$ . But then for source  $i$  ever to be used by  $j$  we must have  $w_{ij} < w_{i'j}$ . This result in turn implies that if user  $j$  and user  $j'$  use source  $i$  and then some time later source  $i'$ , then  $i$  is the cheaper source for both users.

If  $j$  regards source  $i$  as cheaper but, to the contrary,  $j'$  regards source  $i'$  as cheaper, then at most one of the users will use both sources. For suppose, user  $j$  was assigned both resources. Then he will use source  $i$  first since  $w_{ij} < w_{i'j}$  and will use source  $i'$  later only if  $\lambda_i > \lambda_{i'}$ . But then user  $j'$  will *never* use source  $i$  since, for him,  $w_{i'j'} > w_{ij'}$  and so  $w_{i'j'} + \lambda_{i'}e^{rt} > w_{ij'} + \lambda_i e^{rt}$  for all  $t \geq 0$ . Thus, it will never be optimal for user  $j$  to use the two resources in one order and user  $j'$  to use the same two sources in the opposite order.

Now suppose source  $i$  has the lower per-unit cost for both user  $j$  and  $j'$  and in the optimal plan each user switches to  $i'$  immediately after  $i$ . Then user  $j$  would switch at the instant  $t^j$

Figure 2: No One Uses Source 2 for an Interval of Time



defined by

$$\lambda_i - \lambda_{i'} = (w_{i'j} - w_{ij})e^{-rt^j} > 0, \quad (4)$$

while user  $j'$  would switch at instant  $t_{i'}^{j'}$  defined by

$$\lambda_i - \lambda_{i'} = (w_{i'j'} - w_{ij'})e^{-rt^{j'}} > 0. \quad (5)$$

Notice that both equation (4) and equation (5) have the same left-hand sides. If  $w_{i'j} - w_{ij} = w_{i'j'} - w_{ij'}$ , the two users must switch at exactly the same time. Hence if the two users switch from source  $i$  to  $i'$  and the difference between the component  $i$  and component  $i'$  of their respective column vectors coincide, then they will switch simultaneously. This would occur if the entire column vector of the two users in the augmented per-unit cost matrix were identical, as will be discussed in Section 3.2.2; it would also occur if every component of one

column vector differed from the corresponding components of the other column vector by a constant.

If instead the per-unit cost increase in switching from source  $i$  to  $i'$  is *larger* for user  $j$  than for  $j'$  ( $w_{i'j} - w_{ij} > w_{i'j'} - w_{ij'}$ ), then equations (4) and (5) imply that  $j$  will linger longer on source  $i$  before switching to  $i'$  ( $t^j > t^{j'}$ ). Intuitively, user  $j$  would *delay* switching to the higher cost source because the cost increase is larger than it is for user  $j'$ . Elsewhere, we have referred to this property as “The Principle of Comparative Advantage” (Gaudet et al., 2001, p. 1153).<sup>3</sup>

In the multiple demand model, it is impossible for a source to be used by  $j$ , abandoned by everybody, and then utilized again by the *same* user  $j$ . For, after source  $i$  is abandoned for any alternative  $i'$ ,  $\lambda_i e^{rt} + w_{ij} > \lambda_{i'} e^{rt} + w_{i'j}$  and source  $i$  will never again have the lowest augmented per-unit cost. If setup costs are added to the multiple demand model, however, vacillation of this kind can occur. Gaudet et al. (2001) provide an example with  $I = 3$  and  $J = 2$ . One source has the lowest per-unit cost but, unlike the other two sources, is not utilized initially because it has not yet been set up. Instead, each of the two users relies on a different source. When one of those resources runs out, the planner sets up for that user the low cost resource and supplies his needs from it. Since the per-unit cost of supplying the needs of the other user as well from that setup resource is strictly smaller than continuing to supply them from the source he has been using, the planner will switch him as well to the resource he has just set up. At this point, all the setup costs have been paid and we are in familiar territory. The remaining reserves in the abandoned resource must eventually be used by at least one of the two users. If the user who abandoned the resource before it was exhausted has the comparative advantage in using that source (a strictly smaller per-unit cost difference than the other user), then he must eventually return to it.<sup>4</sup>

---

<sup>3</sup>Chakravorty et al. (2005, Section 2.4) introduce various notions of comparative advantage, including this one, which they refer to as “pairwise comparative advantage.” Prior to that the concept of comparative advantage was also mentioned in Chakravorty and Krulce (1994, pp. 1445 and 1448) and Chakravorty, Roumasset and Tse (1997, p. 1222).

<sup>4</sup>Unlike the case in Im et al. (2006), discussed in footnote 2, the *same* user returns to the source he had abandoned.



### 3.2.2 Homogeneous Users

In the case of homogeneous users, every per-unit cost in a given row of the matrix  $[w_{ij}]$  is the same but the number in a different row may differ. We can again index the  $w_{ij}$  so that row 1 has the lowest per-unit cost...and row  $I$  has the highest per unit marginal cost. If we add  $\lambda_i e^{rt}$  to every term in row  $i$  for  $i = 1, \dots, I$ . we obtain the augmented per-unit cost matrix:  $[w_{ij} + \lambda_i e^{rt}]$ . Every column of that matrix will be the same. Given Herfindahl's insight, if every row is to be lowest over some time interval then the first row must be assigned the largest multiplier...the last row must be assigned the smallest multiplier. Nonetheless the multipliers must be assigned so that at  $t = 0$  the first row in the augmented matrix is smallest ... and row  $I$  is largest. Over time, the second row then becomes the smallest, then the third row,... until the  $I^{\text{th}}$  row becomes the smallest. As in the single user case, nothing is used of a higher cost resource until the lower cost resource is exhausted. Moreover, in the case of homogeneous users (and  $J > 1$ ), every user switches from one resource to the next at the same instant. This is a consequence of assuming homogeneous users; it does not occur in the Hotelling model when users are heterogeneous (except in some circumstances when two columns in the matrix differ by a constant).

### 3.3 Backstops

Assume now that  $B < I$  backstop technologies are available, that can serve the users indefinitely without facing the constraint on the cumulative amount supplied that limits each nonrenewable source. Their cumulative production being unconstrained,  $\lambda_i = 0$  for each of these  $B$  sources of supply and their augmented per-unit cost of serving user  $j$  reduces to the exogenously given  $w_{ij}$ . We assume, without loss of generality, that the first  $I - B$  sources are nonrenewables and the remaining  $B$  sources are backstops.

In the unlikely case that for each  $j = 1, \dots, J$  there exists some backstop technology  $b = I - B + 1, \dots, I$  (which may differ across the  $j$ 's) such that  $w_{bj} < w_{ij}$  for  $i = 1, \dots, I - B$ , then, as long as  $w_{ij}$  is stationary for all  $i, j$ , all users would be served forever by some backstop technology; no nonrenewable source would ever be used. We would essentially find ourselves

in the situation of unconstrained production assumed in Section 2.2 and the analysis carried out therein would apply. Let us assume that this is not the case and that at least some of the users are, at the outset, supplied from a nonrenewable source for which  $\lambda_i > 0$ , as in Section 3.2.

In fact the entire analysis of Section 3.2 continues to hold. As before, the source which user  $j$  will be using at any time  $t$  is determined by constructing the lower envelope of his augmented per-unit cost paths; this lower envelope is given by  $\min\{w_{ij} + \lambda_1 e^{rt}, \dots, w_{I-B,j} + \lambda_{I-B} e^{rt}, w_{I-B+1,j}, \dots, w_{I,j}\}$ . The only unique situation which may arise when backstop technologies are present is that if user  $j$  switches to some backstop technology  $b$ ,  $b = I - B + 1, \dots, I$ , at some date  $\hat{t}$ , then he must have been using a sequence of one or more nonrenewable sources for  $t \in [0, \hat{t})$  and this backstop must have been the last source in his sequence of sources: he will use it for all  $t \geq \hat{t}$  and will never return to a nonrenewable source. Indeed, had the cost-minimizing choice of user  $j$  been some backstop  $b'$  in the interval of time preceding  $\hat{t}$ , switching to backstop  $b$  could not reduce his cost. He must therefore have been using a nonrenewable source of supply. If nonrenewable source  $i$  is being used in the interval of time preceding the switch, it has a lower augmented per-unit cost than any other nonrenewable source during that interval of time,  $w_{ij} + \lambda_i e^{rt} < w_{bj}$  for  $t < \hat{t}$  and  $w_{ij} + \lambda_i e^{r\hat{t}} = w_{bj}$ . Since  $w_{ij} + \lambda_i e^{rt}$  is growing over time for all  $i$  and all  $t$ , relying on the unconstrained backstop  $b$  will therefore remain cost-minimizing forever for user  $j$ .

Note that if it is optimal for two users, say  $j$  and  $j'$ , to switch from the same nonrenewable source  $i$  to the same backstop  $b$ , they will not switch at the same time unless  $w_{bj} - w_{ij} = w_{bj'} - w_{ij'}$ , by the Principle of Comparative Advantage discussed in Section 3.2.1. To see this, simply set  $i' = b$  and  $\lambda_{i'} = 0$  in equations (4) and (5). More generally, not only can users switch from a nonrenewable to a backstop technology at different times, but when they do switch it may well be to different backstop sources (and from different nonrenewables).

### 3.4 Technical Change

We have so far assumed in this section that  $w_{ij}$  is time-invariant for all  $i, j$ . But if source  $i$  is subject to technical change, then  $w_{ij}$  will be decreasing over time towards some lower bound and we should write it as time-dependent:  $w_{ij}(t)$ . If the source is a backstop and hence  $\lambda_b = 0$ , then its augmented per-unit cost reduces to  $w_{bj}(t)$ , which will be decreasing over time. If the source is a nonrenewable, then its augmented per-unit cost of supplying user  $j$  is  $w_{ij}(t) + \lambda_i e^{rt}$ . Since this is the sum of a decreasing and an increasing function, it can have regions where it increases and regions where it decreases; it will be decreasing if and only if the reduction over time in the exogenous  $w_{ij}(t)$  due to technical progress is sufficient to more than offset the growth at the rate of interest of the present value of marginally increasing the finite stock of source  $i$ .

The introduction of technical change does not alter the fact that at any date  $t$  each user will rely on the source which has the lowest augmented per-unit cost to him. But now less can be said in general about the lower envelope of these time paths since each time path may have both increasing and decreasing regions. As a result, many scenarios for the sequence of use of the different sources are theoretically possible. To take but one example, if it is optimal for user  $j$  to switch to some backstop technology  $b$  at date  $\hat{t}$ , then, unlike in Section 3.3 where the  $w_{ij}$ 's are stationary, we cannot rule out that he may have been using a different backstop technology prior to  $\hat{t}$ . Nor can we rule out that he may return to a nonrenewable source in the future if technical progress has sufficiently reduced the exogenous component of the augmented per-unit cost of this alternative source. A user may also abandon a nonrenewable source  $i$  at some date  $t_1$  in favor of another nonrenewable source  $i'$  and then come back to source  $i$  at a future date  $t_2$ , depending on the evolution of  $w_{ij}(t)$  and  $w_{i'j}(t)$  over the interval of time  $[t_1, t_2]$ . In the absence of technical change, this could only happen if there were significant set-up costs, as shown in Gaudet et al. (2001) and briefly discussed in Section 3.2.1.

#### 4 Decomposition of $w_{ij}$

Source of supply  $i$ ,  $i = 1, \dots, I$  can be thought of as being obtained from a resource of type  $n$ ,  $n = 1, \dots, N$ , of grade  $g$ ,  $g = 1, \dots, G$ , located geographically at site  $\ell$ ,  $\ell = 1, \dots, L$ . As for demand  $j$ , it can be thought of as serving end use  $k$ ,  $k = 1, \dots, K$ , at location  $\ell$ ,  $\ell = 1, \dots, L$ . Hence  $I = NxGxL$  and  $J = KxL$ . It then becomes clear that the cost  $w_{ij}$  of providing for demand  $j$  from source  $i$  will not depend only on the cost of extraction (or production, in the case of a reproducible substitute). It will also depend on the cost of transporting to location  $\ell$  the resource of type  $n$  of grade  $g$  converted for end use  $k$ , which can be specific to demand  $j$ . Hence, if we denote by  $c_i$  the unit cost of extraction (or production), by  $d_{ij}$  the transport cost from source  $i$  to demand  $j$ , by  $z_{ij}$  the conversion cost of source  $i$  to satisfy demand  $j$ , and by  $\tau_{ij}$  any taxes (or subsidies) specific to source  $i$  serving demand  $j$ , we have that

$$w_{ij} = c_i + d_{ij} + z_{ij} + \tau_{ij}. \quad (6)$$

The full marginal cost, which also includes the imputed cost of depleting supply source  $i$  by an extra unit, is  $w_{ij} + \lambda_i e^{rt}$ , with  $\lambda_i$  positive whenever the cumulative supply faces a binding constraint and zero otherwise. Decomposing the exogenous cost  $w_{ij}$  in this way sets the stage for useful applications of the multi-demand framework described in the previous sections.

In some cases, one may want to stress the fact that the many users of the multiple sources of supply for a single end use (i.e.  $K=1$ ) are spatially distributed, as in Gaudet et al. (2001) and Ley et al. (2000, 2002). The cost of transporting from each source to each user then becomes the dominant factor in explaining the equilibrium allocation of sources to users. If regulatory constraints can be neglected (i.e.  $\tau_{ij}$  set to zero), this amounts to writing  $w_{ij} = c_i + d_{ij}$ . A particularly interesting example of this can be found in Ley et al. (2002), who apply the multi-demand model developed in Gaudet et al. (2001) to study the efficient intertemporal allocation of the solid waste of cities in the United States to

spatially distributed landfills and incinerators (the backstop technology).<sup>5</sup> They append to the planning problem time-varying constraints on the maximal activity at each landfill and the maximal flow between particular city-landfill pairs. The multipliers associated with these constraints can be added to obtain  $\tau_{ij}(t)$ . They then use this model to evaluate the positive and normative consequences of proposed regulatory policies.

In other instances it may be appropriate to neglect the spatial dimension in order to emphasize the cost of converting different sources of supply (e.g. various grades of oil, coal, natural gas, etc.) to satisfy demand for different end uses (e.g. heating, transportation, electricity, etc.). This then assumes users and supply sources to be identically located (i.e.  $L = 1$ ) and, in the absence of regulatory interventions, we may interpret the different costs as  $w_{ij} = c_i + z_{ij}$ . This is the interpretation retained in Chakravorty and Krulce (1994), Chakravorty et al. (1997) and Chakravorty et al. (2005). Chakravorty et al. (1997) provide a nice application of this approach to the multi-demand framework. As in Nordhaus (1979), they retain four demand sectors (electricity, industry (mainly process and space heating), residential/commercial (mainly heating) and transportation) and four energy sources (oil, coal, natural gas and solar (as a backstop technology)). Using actual data for the conversions costs  $z_{ij}$ , they simulate the model for the world economy and draw conclusions for the projected long term sequences of use of the resources, as well as the resulting carbon emissions and their effects on global warming.

In some situations it is the effect of regulatory environments that differ across users that one may wish to emphasize (i.e. varying  $\tau_{ij}$ ). Hoel (2011) and Fischer and Salant (2014) are examples of this. Both recognize that climate policies differ considerably across countries and adopt a version of the multi-demand model to analyze some of the consequences of this fact for the intertemporal and spatial effects of various changes in climate policies. Hoel considers a case where a single nonrenewable carbon emitting resource and a backstop are two perfectly substitutable sources of energy ( $I = 2$ ), the latter having a higher (constant)

---

<sup>5</sup>The volume of space contained in a landfill at any given time can be viewed as a stock of nonrenewable resource that is being depleted: dumping solid waste in the landfill is equivalent to extracting the displaced volume of space from it and shipping it to satisfy a demand for solid waste disposal.

unit cost. This energy is demanded by two countries ( $J = 2$ ) who have different climate policies (carbon taxes or subsidies), which is the only thing that distinguishes them. Hence one may set  $d_{ij} = z_{ij} = 0$  and  $w_{ij} = c_i + \tau_{ij}$ . He studies various scenarios whereby the taxes or the subsidies to the backstop are increased and shows that such increases can have very different effects on climate costs if the initial taxes or subsidies are different to begin with than if they are the same. Fischer and Salant consider a world of identical countries, a fraction of which are regulated and the rest unregulated ( $J = 2$ ). Therefore, as in Hoel, the only thing that distinguishes the two sets of countries is whether they are regulated or not. There are  $I$  sources of energy:  $I - 1$  different grades of the polluting nonrenewable resource characterized by different time-invariant per-unit costs, different emission factors and different underground reserves, and a clean backstop that experiences exogenous cost-reducing technical change. The cost of energy from the backstop exceeds initially that of the highest cost nonrenewable sources. There are no transport costs ( $d_{ij} = 0$ ) and no conversion costs ( $z_{ij} = 0$ ). Denote the exogenous emissions factor (emissions per barrel) at source  $i$  by  $\mu_i$  and the initial emissions tax by  $\tau_{ij}$ . Over time, the emissions tax (or, price of bankable emissions permits) is assumed to rise at the rate of interest. Hence we may write  $w_{ij}(t) = c_i(t) + e^{rt}\tau_{ij}(\mu_i)$ , where  $c_i(t) = c_i$  for  $i = 1, \dots, I - 1$ . They consider the effects on intertemporal and spatial leakage of an increase in the emission tax, in the speed of technical change in the backstop, and in the fraction of regulated countries.<sup>6</sup> They identify circumstances where even a “negative leakage” can occur, in the sense that a policy change that induces reductions in emissions in the regulated countries also induces reductions in emissions in the unregulated countries. This suggests that promoting technical change in clean substitutes, for instance, could be an alternative to attempting to secure international cooperation on setting target emission reductions and implementing them.

---

<sup>6</sup>The term leakage is used here to describe a situation where, for instance, an increase in a carbon tax at a particular date (or in a particular region) would result in an increase in emissions at another date (or in another region). The latter can be referred to as “spatial leakage” and the former as “intertemporal leakage”.

## 5 Conclusion

The Hotelling model is often criticized for its unrealistic prediction that every user of a non-renewable resource will exhaust it and will then simultaneously move on to the cheapest of the remaining nonrenewables. We trace this unrealistic prediction to the unrealistic assumption that each resource is available to every user at the same exogenous per-unit cost. This assumption fails to hold when users are (1) spatially separated, (2) subject to different taxes, or (3) requiring different processes for the source to be useful. Each of these cases has been studied in isolation in the literature, but the unity of these disparate cases has not always been clear to readers. In this chapter, we have examined the general case. We discussed its main properties: the Generalized Herfindahl Principle and the Principle of Comparative Advantage. Finally, we have shown how some contributions to the literature are special cases of this general framework.

## References

- Chakravorty, Ujjayant, and Darrell L. Krulce (1994) ‘Heterogeneous demand and order of resource extraction.’ *Econometrica* 62, 1445–1452
- Chakravorty, Ujjayant, Darrell Krulce, and James Roumasset (2005) ‘Specialization and non-renewable resources: Ricardo meets Ricardo.’ *Journal of Economics Dynamics and Control* 29, 1517–1545
- Chakravorty, Ujjayant, James Roumasset, and Kiping Tse (1997) ‘Endogenous substitution among energy resources and global warming.’ *Journal of Political Economy* 105, 1201–1234
- Fischer, Carolyn, and Stephen W. Salant (2014) ‘Limits to limiting greenhouse gases: Intertemporal leakage, spatial leakage, and negative leakage.’ Resources for the Future, DP 14–09
- Gaudet, Gérard, Michel Moreaux, and Stephen W. Salant (2001) ‘Intertemporal depletion of resource sites by spatially distributed users.’ *American Economic Review* 91, 1149–1159
- Herfindahl, Orris C. (1967) ‘Depletion and economic theory.’ In *Extractive Resources and Taxation*, ed. Mason Gaffney (University of Wisconsin Press) pp. 63–90
- Hoel, Michael (2011) ‘The supply side of CO2 with country heterogeneity.’ *Scandinavian Journal of Economics* 113, 846–865
- Hotelling, Harold (1931) ‘The economics of exhaustible resources.’ *Journal of Political Economy* 39, 137–175
- Im, Eric Iksoon, Ujjayant Chakravorty, and James Roumasset (2006) ‘Discontinuous extraction of a nonrenewable resource.’ *Economics Letters* 90, 6–11
- Ley, Eduardo, Molly K. Macauley, and Stephen W. Salant (2002) ‘Spatially and intertemporally efficient waste management: The costs of interstate trade restrictions.’ *Journal of Environmental Economics and Management* 43, 188–218
- Ley, Eduardo, Molly Macauley, and Stephen W. Salant (2000) ‘Restricting the trash trade.’ *American Economic Review* 90, 243–246
- Nordhaus, William D. (1979) *The Efficient Use of Energy Resources* (New Haven, Conn.: Yale University Press)