SNF REPORT NO. 14/03

Competition and Collusion With Exhaustible Resources - The Case of the Crude Oil Market

by

Martin Bech Holte

SNF Project No. 8300
SIP Nærings- og konkurransepolitikk

The project is financed by the Research Council of Norway

INSTITUTE FOR RESEARCH IN ECONOMICS AND BUSINESS ADMINISTRATION
BERGEN, JUNE 2003

ISBN 82-491-0270-3
ISSN 0803-4036

Table of Contents

Summary

The market for crude oil gained prominence in the early 1970s as skyrocketing prices caused problems for the world's economies. In the academic field substantial resources were devoted to gaining a better understanding of this market; of its structure and dynamics, and how it affected importing economies. Most economists predicted that the collection of black sheep, the OPEC member countries, would soon find it impossible to sustain cooperation. However, the OPEC countries have been far more cohesive than anticipated, and the organization continues to play an important role both for the development in the oil market, but also for the world economy.

In this thesis I will study how the exhaustibility of a resource affects competition between producers, and how it affects the opportunities for collusion. The focus will primarily be theoretical, but throughout the thesis I will confront the theoretical predictions with market data.

After first having given a short summary of how production and reserve levels are distributed among producers, and how the cost levels are in different parts of the world, I start to analyze competition between producers of an exhaustible resource. After having described what is known as Hotelling's rule, models of perfect competition, monopoly and Cournot oligopoly are analyzed. With respect to the latter model I show some interesting facts about the production paths when the relative size (in terms of reserves) of producers varies.

A problem with the standard theory of exhaustible resources is that its basic prediction of steady price increases to a large extent is contradicted by the facts. In part three of the thesis I discuss why this is the case, and I also discuss the relevance of the exhaustibility assumption.

Part four is devoted to a description of what I have called 'real world' models. These models try to capture important facts about the oil market, but leaves the complicating factor of exhaustibility aside.

The parts five to seven is devoted to theory of collusion and OPEC. After having presented standard theory of collusion in repeated games, I show how exhaustibility complicates matters, and how it (negatively) affects the possibility of cooperation. In part five I also describe the dynamics of OPEC, for instance the pervasiveness of cheating on quotas and how this is punished.

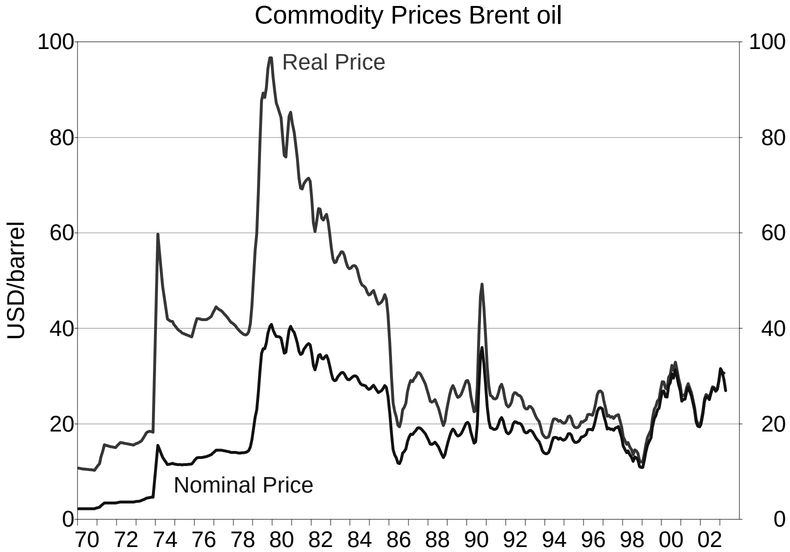
A special characteristic of the oil market is that the stream of oil revenues are of fundamental importance for the economies of many of the producers. I then argue that the producers with nationalized production should not be regarded as risk neutral producers. I then show that this

has important consequences for the workings of the OPEC cartel and the interaction between OPEC and non-OPEC. Saudi Arabia obviously plays an important part in this story, and I argue that its economic problems have caused it to follow a less aggressive strategy than in the past. I also model how risk aversion might benefit the stability of a cartel. After the collapse of oil prices in the mid 1980s, excess capacity levels among OPEC countries have dropped substantially, and I show how this might be explained.

In part eight I conclude with some speculations on the future of OPEC. [1]

[1] I would like to thank my adviser Professor Lars Sørgard for valuable suggestions and comments. All errors are of course mine. I would also like to thank SNF for financial support.

Part 1 – Introduction: Description of the Oil Market



Data: Ecowin

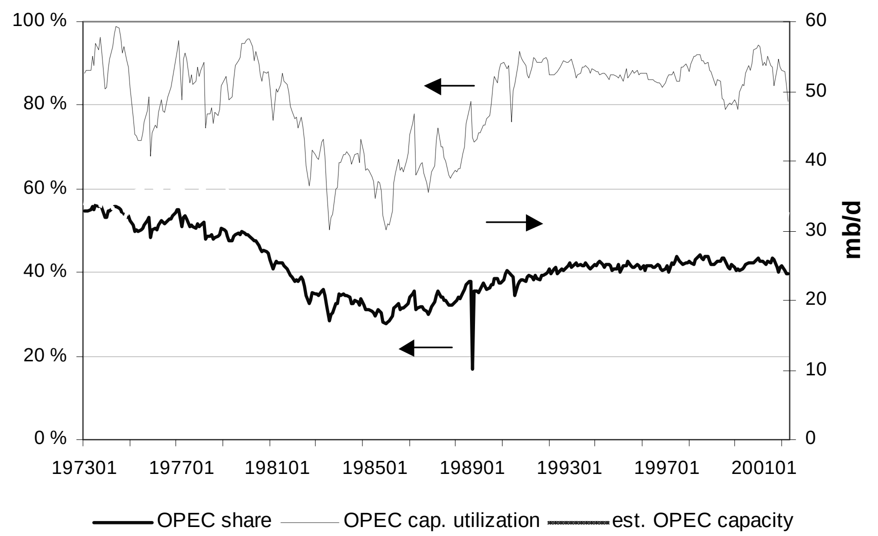
I will introduce you to the topic of this paper by presenting a figure describing the development of nominal and real crude oil prices over the past three decades. It has been a violent journey, both literally and figuratively speaking. We can see that real crude oil prices have fluctuated between $10 and almost $100 (in 2003 US Dollars), after the consequences of wars, a revolution, embargoes, sanctions, and the strategic behavior of crude oil producers have been allowed to take effect. It is impossible to point at single traits or events that have caused these oil price fluctuations, and it is hard even to describe the underlying structure and functioning of a market with such an amount of noise. Still, the objective of this paper is to describe and analyze the structure and the dynamics of the supply of crude oil. The underlying question is how the competition between producers of an exhaustible resource works.

Since 1973 the players in the oil market have composed a major force in geopolitics and in the world economy. Oil prices have surged and plummeted, and this has created instability for both producers and consumers. Quite a few commentators and analysts are daily trying to predict the development of the oil prices, but, apart from knowing that a war or a threat of war in an oil-producing region will make the prices increase and that too much production will make prices fall, it is hard to make good predictions. For instance, over the past couple of years prices have sometimes responded in a seemingly odd manner on news from OPEC, as oil prices have increased when OPEC has announced that production would increase.[2]

This paper will probably not make the analysts' short-term forecasting job much easier, but at least I will try to point out the main factors determining the long-run dynamics. For instance, what determines production behavior in exhaustible resource markets? Is exhaustibility really a relevant factor? To what extent does OPEC behave as a cartel? Is it reasonable to consider Russia and the rest of non-OPEC just as a competitive fringe? How important is Saudi Arabia, and how can one describe its strategic behavior?

Before we discuss all these questions, however, we should get a sense of how crude oil production and reserves are distributed between producers, and what the cost structure in oil exploration and production looks like in different regions.[3] Let us first take a look at production. It seems natural first to take a look at OPEC. Below is a graph that shows OPEC's share of total world oil production, and its estimated production capacity (right axis) and capacity utilization. The graph shows that OPEC's market share in January of 2002 was about 40 percent, up from its lows during the 1980s, but down from a share of about 55 percent during the 1970s. It is worth noting this market share, and already at this point we should ask ourselves whether this makes it possible to create and sustain a cartel. This issue will be thoroughly discussed below.

OPEC's share of world crude oil production

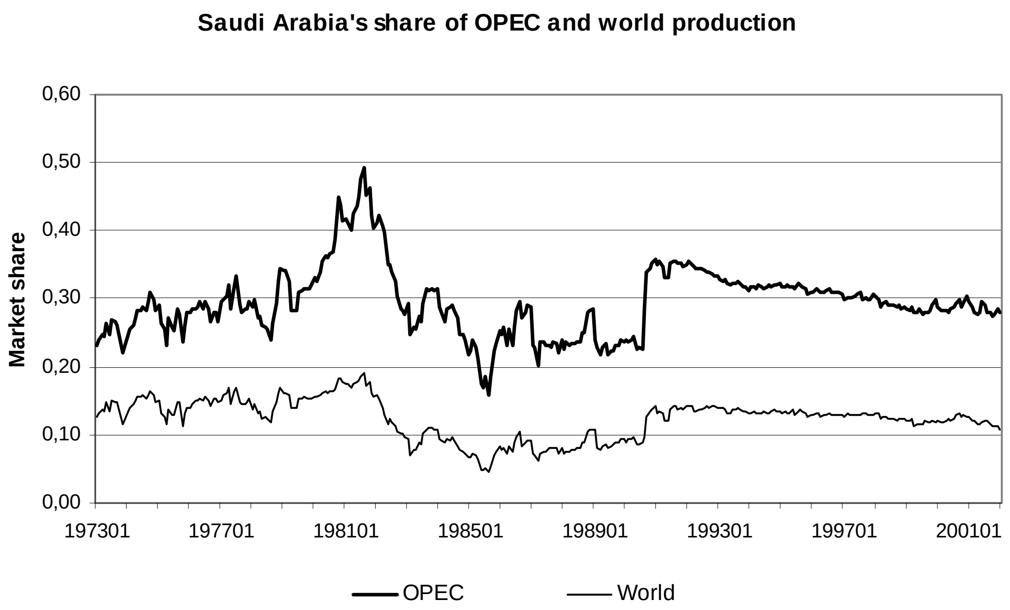


Data: IEA, CERA

[2] Of course, this could be because the production increases were lower than expected, but this does not seem to be the case. Rather, the price increases seem to be due to a realization that the production capacity is lower than previously expected.

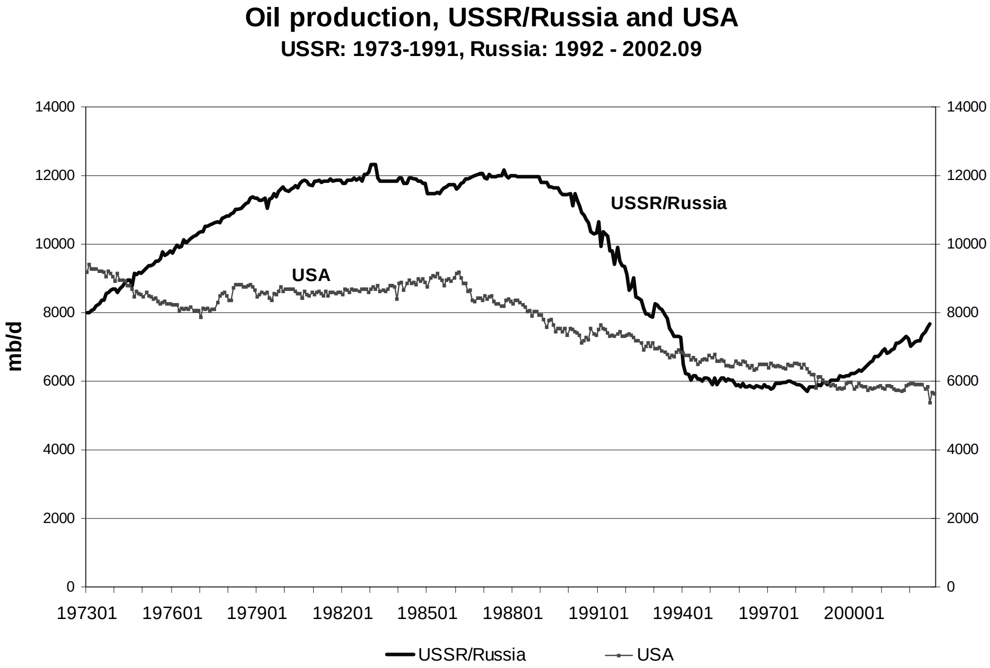
[3] Natural gas liquids are included in the production figures that follow.

Within OPEC, Saudi Arabia is the largest and dominant producer. Saudi Arabia currently represents about 30 percent of OPEC production, but as can be seen from the graph below, this share has been fluctuating quite dramatically over the past three decades. Interesting stories lie behind the fluctuations in this market share, and to find an explanation of its development is of great importance for our understanding of the market dynamics. Again, I will have to refer to the discussion below.



Data: IEA

The OPEC countries have never been the world's sole oil producers. Norway and Mexico as well as other countries are established as important exporters, and we have already seen that non-OPEC's share of world oil production is about 60 percent. The two most important players within this group appears to be the US and Russia. The importance of the US stems from its huge demand for petroleum products and from its dwindling production. Currently the US is producing 3,5 million barrels a day (mbd) less than it did during the mid- 1980s, while its demand is steadily increasing (see graph below). Currently the US consumes close to 20 mbd, which means that it is an enormous importer of petroleum products. The US consumption of petroleum products was close to 26 percent of the world total in 2001 (this share is down from 31 percent in 1973). A back of the envelope estimate then says that the US import demand equals the aggregate production of Mexico, Norway, and Russia. With this in mind, it is understandable that secure access to oil supply is one of the most important issues in US foreign policy.



Data: US Energy Information Administration

The increases in Russian oil production during the latter part of the 1990s and early 2000s have been given a lot of attention in the media. Russia is now the largest oil producer in the world together with Saudi Arabia. While this is true, one should also note the (estimated) production level of the former USSR during the 1980s. The USSR was in fact the world's largest oil producer continuously from 1973 to the early 1990s.

The production of exhaustible resources is different from other production in that the producers will have to decide upon the intertemporal allocation of production, that is, when and how the petroleum wealth should become financial wealth. The amount of petroleum reserves will be of first-order importance for developing this strategy, and it is therefore important to get a sense of the distribution of these reserves. The table below presents the data on 'proven' reserves. Even though only fields that have been discovered are included, precise estimates are hard to give, and this is illustrated by the differences in estimates given by the journals *Oil and Gas Journal* and *World Oil*. If we disregard the uncertainties for a moment, we clearly get a picture of where most of the reserves are located. Saudi Arabia alone has 25 percent of total reserves, while the countries located in the Middle East possess around 70 percent of the reserves. The horizon of several of these countries is very long, with reserve ratios[4] of plus minus 100 years. Outside the Middle East, most oil reserves are found in Venezuela and Russia.

Location of World Crude Oil Reserves*				
Ranked by *Oil and Gas Journal* 2001 Reserve Estimates	*Oil and Gas Journal*	*World Oil*	Total Production 2001	Reserve Ratio
Saudi Arabia	261,7	265,3	2 920,0	91
Iraq	112,5	115,0	902,0	127
United Arab Emirates	97,8	62,8	796,0	79
Kuwait	96,5	98,8	759,0	130
Iran	89,7	96,4	1 343,0	72
Venezuela	76,9	47,6	1 055,0	45
Russia	48,6	54,3	2 449,0	22
Libya	29,5	30,0	489,0	61
Mexico	28,3	26,9	1 110,0	24
China	24,0	30,6	1 179,0	26
Nigeria	22,5	24,1	763,0	32
United States	22,0	22,0	2 121,0	10
Qatar	13,2	5,6	277,0	20
Norway	9,4	10,1	1 142,5	9
Algeria	9,2	12,7	296,0	43

* All numbers are billion barrels

Sources: Oil and Gas Journal, World Oil, EIA

As have been pointed out, 'proven' reserves are not the same as true reserves. However, the same is true when one compares desired and true reserves. During the 1990s, euphoric descriptions of the 'Gulf of the 21st Century', the Caspian Sea, were widespread. The problem is that the current production levels are still depressed after the collapse of the Soviet Union, and the enormously lucrative oil fields everyone was hoping for are to a large extent yet to be found. The countries surrounding the Caspian Sea (excluding Iran and Russia) currently account for just one percent of international petroleum trade flows. The 2002 combined production of Azerbaijan, Kazakhstan, and Turkmenistan is just one third of the Norwegian production. The table below illustrates the discrepancy between current reality and the high hopes that still exist.

[4] These are given by dividing the reserves with the current production level

Caspian Sea Region: Reserves and Production			
Country	Proven* Oil Reserves	Possible** Oil Reserves	Production (mbd)****
Azerbaijan	1,2 bbl	32 bbl	0,26
Iran***	0,1 bbl	15 bbl	0
Kazakstan	5,4 bbl	92 bbl	0,6
Russia***	2,7 bbl	14 bbl	0,14
Turkmenistan	0,6 bbl	80 bbl	0,13
Total	10 bbl	233 bbl	1,13

Sources: Oil and Gas Journal, EIA

* Proven reserves are defined as oil deposits that are considered 90% probable

** Possible reserves are defined as oil deposits that are considered 50% probable

*** Only the regions near the Caspian are included

**** As of July 2002

Even though the optimists have been proven wrong so far, it is easy to justify plowing even more resources for exploration into the Caspian Sea Region. In the Caspian Sea the US sees its chance in reducing its dependence on Middle East oil, and after the terror attacks against the US, the military and US oil companies have joined forces with the local governments in order to discover the true potential of the region. According to Business Week, ChevronTexaco Corp., Exxon Mobil Corp., BP PLC, and Halliburton are among the companies that are about to undertake some heavy investments in the region, and BP alone plans to invest as much as $12 billion in the region over the next eight years. During the early 1990s some US companies formed joint ventures with the national companies of the region, but due to low-quality oil and political difficulties interest in the region faded after a while. Now however, the companies have got the political and military backing they need in order to do create a background that supports business operations. An important factor in this game is the construction of new oil pipelines, which the US does not want to go across Iran. As the US is now negotiating with countries that badly need foreign investment, chances are that US will get what it wants. Currently it seems to be premature to dismiss the potential of the region. Still, if the exploration activities over the next years are unsuccessful, the optimistic reserve estimates will have to be reconsidered.

We will finish this introductory part by looking at the structure of worldwide costs in finding, developing and producing crude oil. The cost structure is of course essential in all markets, but perhaps especially so in markets for exhaustible resources. As we will see below, under perfect competition the theory implies that optimal extraction involves exhausting low-cost reserves prior to the less accessible reserves. The table below shows estimated upstream costs in various regions in the world. We can see that both finding and development costs and production costs are lowest in the Middle East, while costs for instance in South American countries and in the North Sea are far higher[5].

Worldwide Upstream Costs

Country	Exploration (Finding) Only	Development	Finding & Development Cost	Production Cost	Royalties included?	Total
Iraq	0,25	1,75	$2,00	$1,50	no	$3,50
Saudi Arabia	0,25	1,75	2,00	1,50	no	$3,50
Kuwait	0,25	1,75	2,00	1,80	no	$3,80
Iran	0,25	1,75	2,00	2,50	no	$4,50
UAE	1,00	2,00	3,00	2,00	no	$5,00
Algeria	0,50	1,50	2,00	3,00	yes	$5,00
Alaska	1,50	2,00	3,50	2,50	no	$6,00
Oman	1,50	2,25	3,75	2,50	no	$6,25
Nigeria	0,50	3,00	3,50	2,75	no	$6,25
West Siberia	1,00	3,00	4,00	3,00	no	$7,00
Kazakhstan	1,00	2,00	3,00	4,00	no	$7,00
Venezuela	0,25	1,50	1,75	5,50	yes	$7,25
Brazil	1,00	3,00	4,00	3,50	yes	$7,50
Mexico	1,00	4,00	5,00	3,00	no	$8,00
US Lower-48	2,00	3,25	5,25	3,60	yes	$8,85
Indonesia	1,50	1,00	2,50	8,00	no	$10,50
North Sea	2,50	5,00	7,50	3,75	yes	$11,25

Source: Cambridge Energy Research Associates.

[5] Ideally, I would have liked to present production costs excluding royalties in order to illustrate the true social costs of production. Unfortunately I have not been able to find this information. The overall pattern of costs is not affected by this shortcoming, however.

Part 2 - How to Model the Oil Market: Exhaustible Resource Theory

The oil market is a complex market to model. The problems stem from two main attributes of the market, one with respect to the product itself, and the other with respect to the producers. The former problem is due to the fact that oil is an exhaustible resource. Ever since Hotelling (1931), economists have understood how to think about the intertemporal considerations with respect to the production of fixed exhaustible resources. However, even though oil in principle is an exhaustible resource, it is not clear that Hotelling's approach is the appropriate one for the oil market. So, as economists have problems even classifying the nature of the product of the industry, it is no wonder why it is hard to develop models that describe the market adequately. Second, even though exhaustibility implies a markup in prices relative to marginal costs, the price of oil is by most estimates far higher than that one would expect to see under perfect competition. This means that some kind of market imperfection must exist. If the producers were regular profit maximizing firms, one might expect that it would be possible to explain the nature of the imperfections with standard economic theory. However, the producers, or the owners of the resource that is extracted, are sovereign nations[6]. We would expect that this would influence the market behavior of the producers, and that we might have to depart from the standard assumption of purely profit maximizing firms.

Traditional Theory: Exhaustibility and the Hotelling Rule

The basic theory underlying the models describing the oil market is the principle of resource exhaustibility. It is necessary to discuss this theory in some detail if we want to understand the methodological challenges one faces when we want to model the market for exhaustible resources. Dasgupta and Heal (1979, p. 153) define an exhaustible resource as a resource that 'is used up when used as an input in production and at the same time its undisturbed rate of growth is nil. In short, the intertemporal sum of the services provided by a given stock of an exhaustible resource is finite'.

One of the consequences of exhaustibility is that the price will exceed marginal cost. Hotelling (1931) was the first to study this formally, and he reasoned in the following manner: First of all, an owner of an exhaustible resource should be indifferent as to when he sells his product. Hence, the value of the unextracted resource must appreciate at the same pace as the

[6] There is one exception: In the US oil reserves are mainly private property.

current profits from extraction. Then, if we allow for constant marginal cost of production, the net spot price of the unextracted resource (that is, price less marginal cost) will appreciate at the rate of interest. Formally we have that $p_t=p_0e^{rt}$, where p is the net price, r is the interest rate, and the subscripts represent the periods. Hence, even under perfect competition, the nature of an exhaustible good will make the difference between price and marginal costs strictly positive. The strictly positive markup for oil producers is known as the scarcity rent, while the sum of marginal costs of production and the scarcity rent is known as user cost, which is the definition of the opportunity cost of selling oil in any given period (Crémer and Salehi Isfahani (1991)). One of the problems in modeling exhaustible natural resource industries is that the markup in prices above marginal cost stems not only from the market structure and the degree of market power, but also from the value of the scarcity rent. All models should try to disentangle these two effects.

In this section we will study the theory of exhaustible resource markets. We will start out with the case of perfect competition, and then look at the cases of monopoly and Cournot oligopoly.

Perfect Competition

We can show the principles underlying Hotelling's rule by solving the allocation problem for an individual firm (a price taker).[7] The firm will maximize the discounted profits over an endogenous horizon of T years, $t= (1,\dots,T)$, subject to a constraint that total production cannot exceed the amount of initial reserve. The profit function (in discrete time) takes the following form:

(2.1) $$\Pi = \max_{q,T} \sum_{t=0}^{T} \left[p(t)q(t) - C(q(t),t) \right] e^{-rt} \ ,$$

where r is the interest rate, $p(t)$ is the price in period t, $q(t)$ is the quantity produced (and sold), while $C(q(t),t)$ is the cost function, with C_q denoting the derivative, where $C_q > 0$. We have assumed perfect foresight, such that the price path is perfectly known for the individual producer. Our problem is now to find the extraction path $q^*(t)$ and the termination date T^* that maximizes the present value of the profits.

[7] The following exposition draws on Sydsæter (1990) and Léonard and Van Long (1992)

We have a few constraints that must be respected. If we let the amount of remaining reserves be *S(t)*, we have that $dS(t)/dt = \dot{S} = -q(t)$, which says that in period *t* the remaining reserves are reduced by the amount of production in that period. Further we have the following conditions:

$$S(0) = S_0$$
$$S(T) \geq 0$$
$$q(t) \geq 0$$

These restrictions just tell us that the amount of initial reserves equals S_0, that there cannot be negative amounts of reserves left after period *T*, and that production cannot be negative, that is, we cannot pump oil back into the reservoir. We should note that S_0 here is considered to be the certain level of reserves at time *t=0*, that is, there is no uncertainty related to the true level of reserves.

If we let λ denote the costate variables, we can present the Hamiltonian:

(2.2) $$H(t,S,q,\lambda) = \lambda_0(p(t)q - C(q,t))e^{-rt} + \lambda(t)(-q).$$

λ_0 must take on the values 0 or 1, and λ(t) must be a continuous function such that $(\lambda_0, \lambda(t)) \neq (0,0)$ for all *t*, and such that[8]

(2.3) *q*(t)* maximizes $\lambda_0(p(t)q - C(q,t))e^{-rt} + \lambda(t)(-q)$ for $q(t) \geq 0$

(2.4) $$\dot{\lambda}(t) = -\partial H / \partial S = 0$$

(2.5) $$\lambda(T^*) \geq 0$$

(2.6) $$\lambda_0(p(T^*)q^*(T^*) - C(q^*(T^*),T^*))e^{-rT^*} = \lambda(T^*)u^*(T^*)$$

It can be shown that $\lambda_0 = 1$ is the only possible solution for λ_0. If we denote equation (2.3) as $\pi(q,t) = (p(t)q - C(q,t))e^{-rt} - \bar{\lambda}q)$, and take the second derivative of the profit function with respect to *q*, *π''(q)*, we find that the function is concave, and hence that the second order condition for maximization is satisfied. $\bar{\lambda}$ is here a constant ≥0, due to equations (2.4) and (2.5). (2.3) says that *q*(t)* maximizes *π(q,t)* for *q≥0*. If *q*(t)=0* we must have that

[8] This is according to Sydsæter (1990), p. 383

$\pi'(q^*(t),t) \leq 0$. If $q^*(t) \geq 0$ we must then have that $\pi'(q^*(t),t)=0$. Due to the concavity of π, these conditions will hold.

At a time when $q^*(t) \geq 0$, that is, at a time when production is taking place, we have that

(2.7) $$\frac{\partial \pi(q^*(t),t)}{\partial q} = (p(t) - C_q(q^*(t),t))e^{-rt} - \bar{\lambda} = 0 .$$

If we just manipulate (2.7) slightly, we get that

(2.8) $$p(t) - C_q(q^*(t),t) = \bar{\lambda} e^{rt} .$$

This is the Hotelling rule. On the left hand side of (2.8) we have the marginal profit, and the rule then says that in equilibrium the marginal profit must rise at the rate of interest, *r*, which is the discount rate.

Further, if we replace *t* with *T**, the optimum end date of production, into (2.8), and use relationship (2.6), we find that, as long as $q^*(T^*)) \geq 0$, we must have that

(2.9) $$C_q(q^*(T^*),T^*) = \frac{C(q^*(T^*),T^*)}{q^*(T^*)} .$$

In other words, at the final stage of production, the marginal cost of production will equal the average cost of production. After that stage, it is time to leave the wells. When one knows the final date of production, one can work backward to the initial time *0*, using the price path and Hotelling rule to allocate production over the interval *[0,T*]*. By aggregating these individual supply curves, it is straightforward to find the industry supply curve, and the market equilibrium is found by setting demand equal to supply in every period. Hotelling (1931) showed that this perfect competition equilibrium was socially efficient by showing that the same output path as under perfect competition would have been chosen by a social planner. The Hotelling rule is the basic intertemporal efficiency condition that will determine the producer's allocation of extraction over time, and it says that a producer in equilibrium, since both oil in ground and profits will earn an interest equal to *r*, will be indifferent between a marginal increase in production and leaving this marginal barrel in the ground. Hence, the rule says that there is zero profit from arbitraging the marginal barrel of oil across time *t*.

The basic implication of the Hotelling rule, is that we should expect a price path that is strictly increasing over time, as the scarcity rent increases. This mode of thought dominated thinking during the 1970s and 1980s, with the consequence that moderate variations in the price level of period $t=0$ caused dramatic consequences for net present value calculations of petroleum wealth.[9] The development of the oil price after the two major oil shocks in the 1970s has not given support to Hotelling's rule. It is reasonable, therefore, to ask, not whether the rule is correct per se, but whether it characterizes the market for crude oil in an adequate manner. Before we try to give an answer to this in Part 3, we will discuss how this basic model has been extended in order to capture some of the other characteristics of the oil market.

The model has been extended in many ways. For instance, one may treat extraction costs as rising as the stock of reserves is depleted, and one may also include technological progress, which rises with cumulative production. It should be noted that the Hotelling rule in its most basic form, and under perfect competition, does not necessarily give an exact estimate of the scarcity rent. For instance, if current extraction of an exhaustible resource leaves lower grade resources left and cause production costs to increase in the future, the scarcity rent will be less than the interest rate. I will now illustrate why it must be so.

Following Dasgupta and Heal (1979), let $S(t)$ represent the amount of remaining reserves. Further, let the cost function be $C=C(S(t),q(t))$, with $C_S<0$, as marginal costs increase as reserves are extracted, and $C_q>0$, as marginal costs are increasing in production. We also have that $p(t)$ is the competitive spot price of extracted oil, and $b(t)=\lambda e^{rt}$ is the competitive price of the unextracted product (the scarcity rent). Then, the competitive spot price $p(t)$ must be given by

$$(2.10) \qquad p(t) = b(t) + \frac{\partial C}{\partial Q(t)},$$

and this market price determines the volume of the resource flow that clears the market. Let us now consider the arbitrage condition. The rate of return from holding the marginal unit of the stock now consists of two components. First, we have the capital gains on the unextracted stock itself. Second, we have a new element stemming from the fact that this marginal unit has been stored and hence contributed to a reduction in extraction costs, as one does not have to drill as deep for the next unit as one would have had to if this marginal unit in fact were

[9] For instance, in 1981 the Norwegian petroleum wealth was estimated to NOK 2273 billion, while the estimate had plummeted to NOK 413 billion in 1988.

extracted. A person that now holds *b(t)* units of the numeraire asset will be assured *(1+r)b(t)* at time *(t+1)*. Alternatively one unit of the unextracted resource can be purchased at time *t* before it is sold at time *(t+1)*. At time *(t+1)* he will not only get *b(t+1)* units of the numeraire, he will also be paid for not having increased costs by extracting this unit, and this amounts to *dC* units of the numeraire. Under perfect competition these two alternatives should be equivalent, which means that the following must hold:

(2.11) $$\frac{\dot{b}(t)}{b(t)} - \frac{\frac{\partial C}{\partial S(t)}}{b(t)} = r \, .$$

We now see that the underlying appreciation of the unextracted resource is less than the interest rate. This is because there are gains to be had in leaving the resource in the ground over and above the capital gains. We can therefore conclude that the shape of the cost function may moderate Hotelling's result.

Even though perfect competition models do not seem to reflect the real oil market structure, we should try to explain why we under perfect competition do not expect producers with considerable variations in their costs of production to operate simultaneously. This is interesting because the result we get is pretty much just the opposite of what we experience in the real world.

Suppose that there are two different types of deposits of crude oil, the difference being that the unit extraction cost is higher in deposit 2 than in deposit 1. The constant marginal extraction costs are denoted c_1 and c_2 (with $c_2>c_1$). The quality of the oil is identical in the two deposits, and the extracted oil will sell for the same price p_t in the spot market. Further, let $b_1(t)$ and $b_2(t)$ be the spot prices of the unextracted resources. Suppose now that the two deposits were in use simultaneously. Then the arbitrage principle demands that the two producers should be indifferent between storing and extracting:

(2.14) $$\frac{\dot{b}_1(t)}{b_1(t)} = \frac{\dot{b}_2(t)}{b_2(t)} = r \, .$$

Since the extracted products sell for the same price, this implies that the following must hold:

(2.15) $$c_1 + b_1(t) = c_2 + b_2(t), \forall t,$$

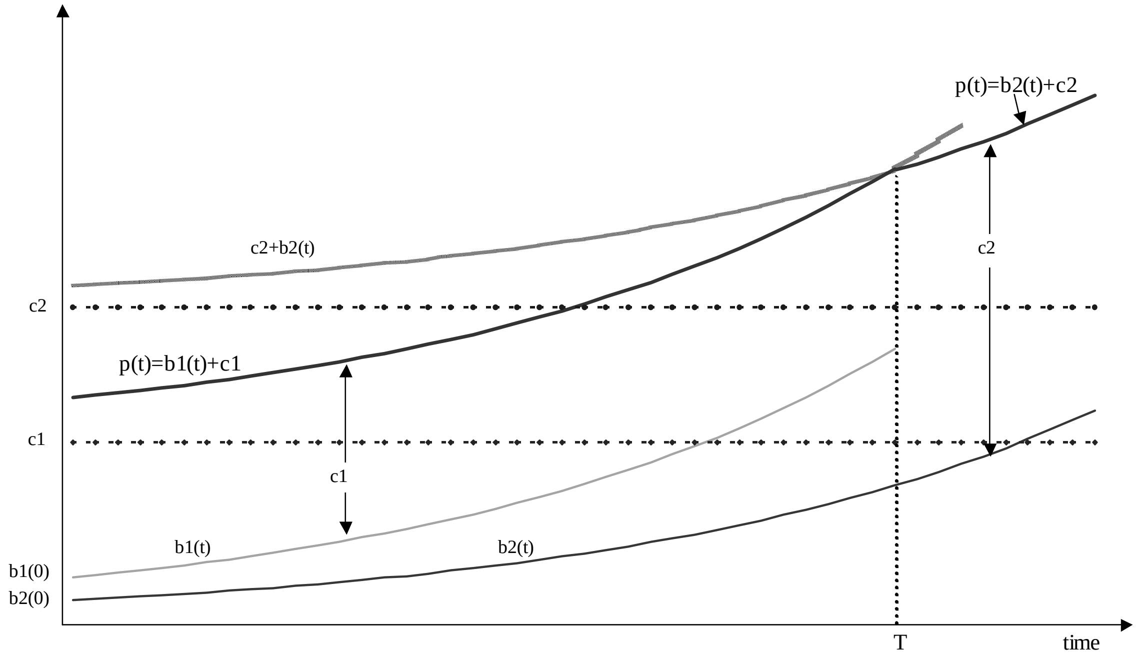
in other words, that the scarcity rent (the value of an unextracted unit) plus marginal costs must be equal at all times *t*. The problem is that (2.14) and (2.15) are not consistent with one another. To explain this, let us assume the following: r=0,05, c_1=20, c_2=40. If in period 0 equation (2.15) shall hold, we must then have that, for instance, $b_1(0)$=40 and $b_2(0)$=20. Then the market price for the extracted product will be 60 in period 1. Let us now move to period 1. According to (2.14), the remaining reserves of the two deposits will now have earned an interest equal to 0,05. This means that $b_1(1)$=42 and $b_2(1)$=21. But now (2.15) will not hold. The conclusion is that the two deposits will not be exploited simultaneously over any interval of time. Rather, they will be exploited sequentially. The precise sequence will be that the low-cost producer's deposit will be exploited first, and when it is exhausted, the high-cost producer will start extracting from his deposit. The sequence must be like this because the low cost producer will always be able to undercut the high cost producer and capture the entire market. So, for an initial period *(0,T)*, the high-cost deposit will not be profitable to exploit, as

(2.16) $$p(t) = c_1 + b_1(t) < c_2 + b_2(t) \text{ for } 0 \le t < T.$$

As the low-cost producer will always be able to undercut the other, the value per unit of his deposit must be higher than that of the high-cost producer as long as $r>0$. With this in mind, we can illustrate the situation as in the figure below.

In the interval *[0,T]*, the low cost producer will be the sole producer. At point *T*, his deposit will be exhausted, and there will then be a smooth shift in *p(T)* at the point of transition between the two producers. It must be so, because if for instance producer 1 expected an upward jump in the prices, he would have stopped production in the periods prior to *T*, just to see the profits rise after *T*. Nor can there be a discontinuous fall in prices, since then producer 2 would have benefited from entering prior to period *T*. Hence, *p(t)* is continuous at *T*. What we also see is that the rate of price growth will decline as the cost of production increases.

Perfect competition – Two groups with different marginal costs



The harmony of this theory is not reflected in the real world. What we see is that prices in the oil market are so high that high-cost producers are allowed to participate in production, with the low cost producers keeping quiet just to get as high a profit as possible. Even though there might be some deposits outside the OPEC countries that would have been exploited even under perfect competition, the development where high-cost producers have been expanding more rapidly than low cost producers at a time when the low cost producers still possess most of the reserves is simply opposite to what theory predicts. Adelman (1987) puts it this way:

> […] we should expect to see production in the lower-cost areas grow faster than the high-cost. This is precisely what we saw before 1973. Then there was an abrupt turnaround. High-cost areas expanded drilling mightily while low-cost cut back. It was water flowing uphill. The only theory that explains it is monopoly, whether of one or a small group trying to act as one, to restrain output to maintain prices.

Monopoly

After the 1973 oil price increase, analysts of the oil market were interested in studying the consequences of the monopolization of the oil market. The easiest models are those that study a market with one single monopolist. We can just alter the framework of the perfect-competition case slightly in order to see what the producer's problem would look like. First of all, the producer is no longer a price taker, but faces a downward sloping demand curve, and this makes both prices and quantities decision variables. Second, one must add one constraint in order to make sure that supply equals demand in every period. Then, the monopolist's problem would look like this (we now assume an infinite horizon, within which the date of exhaustion is determined endogenously):

$$\max_{q(\cdot),p(\cdot)} \sum_{t=0}^{\infty} \left[p(t)q(t) - C(q(t),t)\right] e^{-rt}$$

(2.17)

$$\text{subject to } \begin{cases} S(0) = S_0 \\ \dot{S} = -q(t) \\ S(T) \geq 0 \\ q(t) \geq 0 \\ q(t) = D(p(t),t) \quad \forall t \end{cases}$$

It is not as straight forward as in static models to figure out how the monopolist would act in this situation. We would of course expect him to restrict output and increase prices, but when a resource is exhaustible, this would imply that at some time in the future the price will be reduced simply because more of the resource will be left, and potential supply relatively more abundant. Indeed, Stiglitz (1976) has shown that with isoelastic demand (and with the elasticity of demand greater than 1), and production costs equal to zero, the output path of the monopolist would be equal to that of the competitive industry. How do we explain this? Let's assume that demand takes the isoelastic form $p(q)=Aq^{\mu}$, where the elasticity of demand is $1/\mu$ and μ takes on a value in the range $-1 < \mu < 0$. With zero extraction costs, the marginal revenue will be $m=(\mu+1)p(q)=$. Then the growth in marginal revenue must equal the rate of interest, that is $\dot{m}_t / m_t = (\mu+1)\dot{p}(q_t)/(\mu+1)p(q_t) = \dot{p}(q_t)/p(q_t) = r$. But this is just the simplest form of the Hotelling rule under perfect competition with zero extraction costs. Hence, we have shown that the profit maximizing condition is the same under constant and isoelastic demand as it is under perfect competition, and this means that the price path under these assumptions will be equal under both regimes. However, if production costs are strictly positive and stationary, or demand becomes more elastic over time (but still isoelastic within each period), the monopolist would be more conservationist and limit production. If demand gets more elastic over time, which in our case means that μ increases toward zero, the above condition becomes $\dot{m}_t / m_t = \dot{\mu} /(1+\mu) + \dot{p}(q_t) / p(q_t) = r$, which means that $\dot{p}(q_t) / p(q_t) = r - \dot{\mu} /(1+\mu)$. Hence we see that as μ increases over time, the rate of growth in prices will be slower, and the level initially higher, than under perfect competition. This is because he will choose a price path that allows the marginal revenue, rather than the price, to rise at the discount rate. A price path that follows this logic will tend to have higher initial prices and slower appreciation of prices, which again leaves the date of exhaustion more distant.

We will now try to solve for the optimal extraction path for a monopolist in order to show how optimal production is affected by changes in the interest rate, changes in reserves, and changes in the discount rate. By solving a problem like (2.17), we must simultaneously find both the time of exhaustion, *T*, the shadow price of reserves, λ, before we can find the price and output path of a monopolist. It might therefore be worthwhile as an exercise to go through the process to see how this is done. Most expositions of the theory stops short of this.

If we assume constant marginal costs, *c*, and inverse market demand given by *p(q)=a-q*, the Hamiltonian of problem (2.17) looks like this (we now let $q_t=q(t)$):

(2.18) $$H(t,S,q,\lambda) = ((p(q_t)-c)q_t)e^{-rt} - \lambda q_t = ((a-q_t-c)q_t)e^{-rt} - \lambda q_t$$

As long as $q_t>0$, the first order condition says (after some manipulation) that

(2.19) $$q_t = (a-c-\lambda e^{-rt})/2 .$$

At time T, the transversality condition tells us that the value of the Hamiltonian must equal zero, that is, $H(T,S,q,\lambda) = ((a-q_T-c)q_T)e^{-rT} - \lambda q_T = 0$, which is the same as saying that $q_T = a-c-\lambda e^{-rT}$. We now have two expressions for *q*. We set them equal to each other and solve for λ, and find that

(2.20) $$\lambda = e^{-rT}(a-c)\,.$$

By using this expression for λ, the expression for q_t given by the first order condition (2.19), and also the reserve constraint $\sum_{t=1}^{T} q_t = S_0$ (we assume that all reserves are exhausted), we can solve the problem. If we substitute the expression for λ into (2.19) we get that

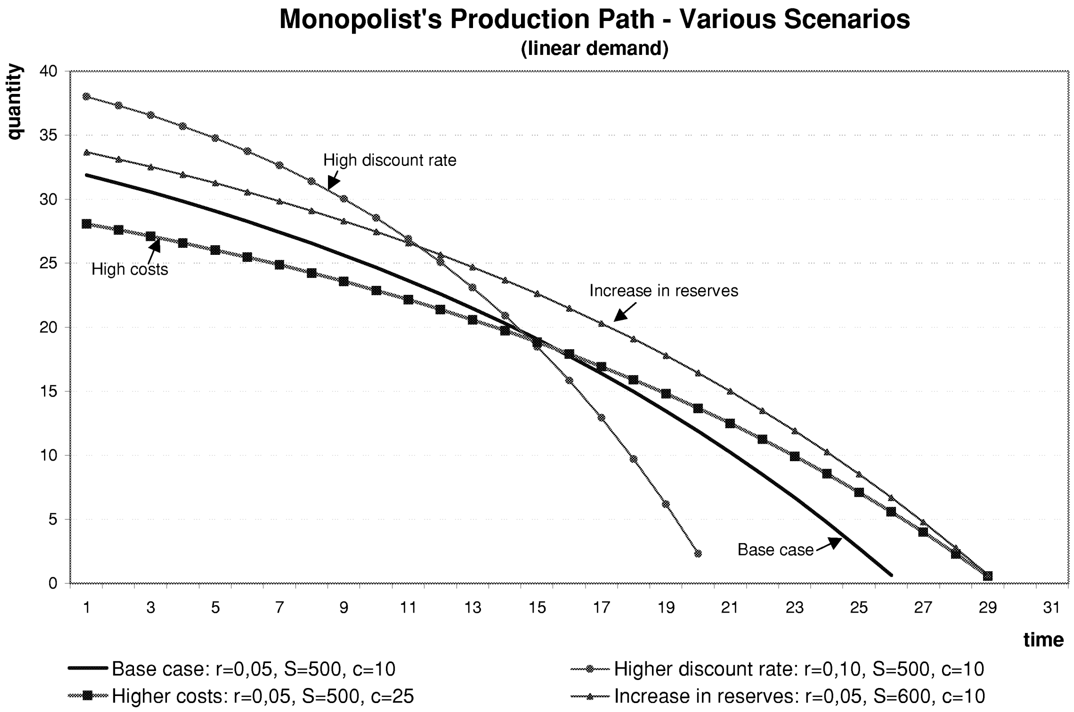
(2.21) $$q_t = \left[a - c - (a-c)e^{r(t-T)}\right]/\,2\,.$$

By summing over all periods T we have that $2S_0 = \sum_{t=1}^{T}\left[a - c - (a-c)e^{r(t-T)}\right]$. Then, by manipulating this expression slightly, and by using the formula for geometric series, we get that

(2.22) $$T(1-e^{-r}) + e^{-rT} = \frac{2S_0}{a-c}(1-e^{-r}) + 1\,.$$

We get a really complicated expression if we want to isolate T in (2.22). However, we can see that the second element on the right hand side converges to zero when T is large, so this is helpful if we want to get a sense of what the solution looks like. When we have got a value of T, it is then easy to find the value of λ (from expression 2.20), and it is also easy to find the values of q (from expression 2.21) and p in each period. Below is an illustration of the production paths of a monopolist when specific values of r, S_0, a, and c are chosen (base case given by $r=0{,}05$, $S_0=500$, $c=10$, $a=100$). We can see that a higher discount rate r increases the quantity produced early in the extraction period, and T is lower than originally. Increased costs delays production, while we see that an increase in the amount of reserves increases production in all periods, but the increases are relatively larger late in the extraction period. (All this can of course be seen from equations 2.19 and 2.21).

It should be noted that the effects from changes in discount rates, amount of reserves, and productions costs all work in the same direction under perfect competitions as they do under monopoly.



Cournot Models

Neither the perfect competition nor the monopoly case fit the oil market in a good manner. Therefore, other modeling strategies have been chosen, especially to illustrate the operation of the cartel and its problems. However, Crémer and Salehi-Isfahani (1991) says that no theoretical work has been developed that really tackles the modeling problems a cartel brings. The second main strategy is to study the relationship between OPEC and the non-OPEC producers. However, these models also face troublesome technical difficulties when the problem is set in a closed loop framework, which means that the optimal value of the control variable during period *(T-t)* is given as a function of the state variable at the beginning of that period (Léonard and Van Long 1992). The closed loop framework is the proper framework for these kinds of models.

There are however extensions of the basic Hotelling model that model the Cournot game between producers in the market.[10] Lewis and Schmalensee (1980) Loury (1986), and Polasky

[10] I have not seen articles on the Bertrand game in exhaustible resource markets. With respect to the oil market, the main reason for this must be that there are clear capacity constraints and most producers are producing more or less at capacity. Kreps and Scheinkman (1983) have shown that a game with quantity precommitment and Bertrand competition gives the Cournot outcome, and this type of game looks like a feasible characterization of the market for crude oil production. Hence the focus on Cournot competition.

(1992) are examples of this. They analyze the Hotelling model with n producers playing a Nash-Cournot game. These models show that an equilibrium exists if we allow for stationary demand and that the elasticity of demand is strictly decreasing in aggregate output. The main conclusions are that prices and profits are falling in the number of producers, and that the equilibrium converges to the perfect competition case as the number of producers grows.[11] The model also shows that the market share of the producers with the largest reserves increases over time.

This description of the oligopoly model to a large extent follows that of Polasky (1992), the main difference being that I will here use discrete rather than continuous time. The results are the same. Let us then assume the following: Each producer *i* possesses an initial amount of reserves, $S_{i0} \geq 0, i = 1,2,...,n$, and there are no uncertainties with respect to the level of this amount. The production level of producer *i* at time *t* is q_{it}. The extraction flow of all producers *i≠j* is given by $Q_{\sim it} = \sum_{j \neq i} q_{jt}$, and the total extraction of all producers at time t is $Q_t = \sum_{j=1}^{n} q_{jt}$. An extraction path for producer *i*, q_{it},, specifies the extraction level over all $t \in [0, \infty\rangle$. An extraction path is feasible as long as $\sum_{t=0}^{\infty} q_{it} \leq S_{i0}$ with $q_{it} \geq 0$ for all *t*. The extraction path for all producers *i≠j* is given by $Q_{\sim i} = \{q_1, q_2, ..., q_{i-1}, q_{i+1}, ..., q_n\}$, and the remaining stock at time t is $S_{it} = S_{i0} - \sum_{\tau=0}^{t} q_{i\tau}$.

We have an inverse market demand curve, $p_t = P(Q_t)$, which is constant over time. We will assume, i), that this market demand curve is downward sloping, ii), the industry revenue $R(Q_t) = P(Q_t)Q_t$ is concave in Q, and, iii), we also assume that value of the elasticity of demand, $\eta_t = -P(\cdot)/(P'(\cdot)Q)$, is strictly decreasing in Q_t. The first two assumptions guarantee that a unique equilibrium exists, while the third assumption gives us oligopoly prices that will differ from those under perfect competition. As mentioned above, Stiglitz (1976) has shown that for constant isoelastic demand over time the behavior of monopolistic and competitive markets will be identical. If assumption iii) holds, the imperfectly competitive producers can earn more than under perfect competition by acting as price discriminators over time.

[11] They are not assuming that the number of producers is growing over time. Rather, they just show that prices and profits are lower in a market with *n* producers than when there are only *n-k* producers.

Let us first look at the case where marginal extraction costs are constant and identical for all producers. We also assume that the discount rate is identical for all producers. At time *t=0* each producer *i* will choose an open-loop extraction strategy that will give him the profits $\pi_i(q_i, Q_{\sim i})$. The time horizon *T* is finite but arbitrarily long.

Producer *i*'s problem is the following:

$$\max_{q_{it}} \pi_i(q_i, Q_{\sim i}) = \sum_{t=0}^{T} e^{-rt}\left[P(Q_t) - c\right]q_{it}$$

subject to

(2.23) $$S_{it} = S_{it-1} - q_{it}$$
$$S_T \geq 0$$
$$q_{it} \geq 0$$

The Lagrangian then becomes

(2.24) $$L = \sum_{t=0}^{\infty} e^{-rt}\left[P(Q_t) - c\right]q_{it} - \sum_{t=0}^{\infty} \lambda_{it}(S_{it} - S_{it-1} + q_{it}) + \mu S_T \ .$$

Lewis and Schmalensee (1980) and Loury (1986) have shown that there exists a Nash equilibrium in open-loop (precommitment) strategies for this problem. We maximize this expression with respect to q_{it}, use the Kuhn-Tucker theorem, and get the following necessary and sufficient first order conditions:

(2.25a) $$\frac{\partial L}{\partial q_{it}} = e^{-rt}(P(Q_t) - c + P'(Q_t)q_{it}) - \lambda_{it} \leq 0$$

(2.25b) $$q_{it} \geq 0$$

(25a) and (25b) hold with complementary slackness.

(2.25c) $$\frac{\partial L}{\partial S_{it}} = -\lambda_{it} + \lambda_{i,t+1} = 0$$

(2.25d) $\frac{\partial L}{\partial S_T} = -\lambda_{iT} + \mu = 0$

(2.25e) $\mu S_T = 0$

If we assume that all reserves are exhausted at some time *T* *(S_T=0)* we ensure that prices will be higher than marginal costs as there is a positive value associated with obtaining more reserves. Then, from these conditions we see that $\lambda_{it} = \lambda_{it-1} = \mu$, which is the shadow price of a marginal increase in reserves at time *T* (the scarcity rent) for producer *i*. Hence we see that for $q_{it} > 0$ we have from (2.25a) that the discounted marginal profit for producer *i* is equal for all *t*, and the level of this profit is equal to the shadow price of a unit of reserves at time *T*, when all the resources is exhausted.

From these first order conditions we get several results. First, we can see, for $q^*_{it} > 0$ that $q^*_{it} > q^*_{jt}$ if and only if $S_{it} > S_{jt}$. This can be shown by the following argument. Suppose $S_{it} > S_{jt}$. Then we must have that $\sum_{\tau=t}^{\infty} q_{i\tau} \geq \sum_{\tau=t}^{\infty} q_{j\tau}$. For some τ we then must have that $q^*_{i\tau} > q^*_{j\tau}$. But from (2.25a), since *P'(Q)<0*, $q^*_{i\tau} > q^*_{j\tau}$ then implies that $\lambda_i < \lambda_j$. Therefore we must have that $q^*_{it} > q^*_{jt}$ for all t such that $q^*_{it} > 0$ and $q^*_{it} = q^*_{jt} = 0$ for any date where $q^*_{it} = 0$.

From this we can see that producers with relative large reserves place a relatively low shadow value on reserves. This means that producers with large reserves will produce more in absolute terms (not relative to reserves) in each period than producers with small reserves. It also means that the producers with a relatively small amount of reserves will have a greater incentive to acquire additional reserves, as the shadow price on these are higher than for large holders of reserves.

Still, even though producers with large reserves will produce more than those with less, their level of production relative to reserves will be lower than for smaller players. Or formally, we will have that $q^*_{it} / S_{it} < q^*_{jt} / S_{jt}$ if and only if $S_{it} > S_{jt}$, for all dates for which $S_{it}, S_{jt} > 0$. The proof can be found in Appendix 1 of Polasky (1992). The intuition behind this result is that large reserve holders will spread out their production over more time periods than smaller players in order not to depress the marginal profits of each period. So, basically, since for $S_{it} > S_{jt}$ we will have that $T_i > T_j$, we will also have that $q^*_{it} / S_{it} < q^*_{jt} / S_{jt}$.

Let us now look at the extraction paths in a market with two players *i* and *j*, and where the inverse market demand curve is linear: $P(Q_t) = a - Q_t$. From (2.25a) we then get that

(2.26) $$P(Q_t) + P'(Q_t)q_{it} = c + \lambda_i e^{rt}.$$

Putting in for $P(Q_t) = a - Q_t$ and solving for q_{it}, we get the following:

(2.27) $$q_{it} = \frac{a - c - \lambda_i e^{rt} - q_{jt}}{2}.$$

We have an analogous function for q_{jt}. Manipulating the latter so that it becomes a function q_{it} of q_{jt}, we can solve for the Nash equilibrium quantities. These are given by

(2.28) $$q^*_{it} = \frac{a - c - 2\lambda_i e^{rt} + \lambda_j e^{rt}}{3}$$
$$q^*_{jt} = \frac{a - c - 2\lambda_j e^{rt} + \lambda_i e^{rt}}{3}$$

From (2.28) we can see that the quantity produced by producer *i* increases in the shadow price of reserves of other producers (that is, a reduction in producer *j*'s reserves, which will increase his shadow price, will increase the quantity produced by *i*). However, a change in producer *i*'s own reserves will have a stronger effect.

We then get that the total quantity produced in period t equals:

(2.29) $$Q^*_t = q^*_{it} + q^*_{jt} = \frac{2a - 2c - \lambda_j e^{rt} - \lambda_i e^{rt}}{3}.$$

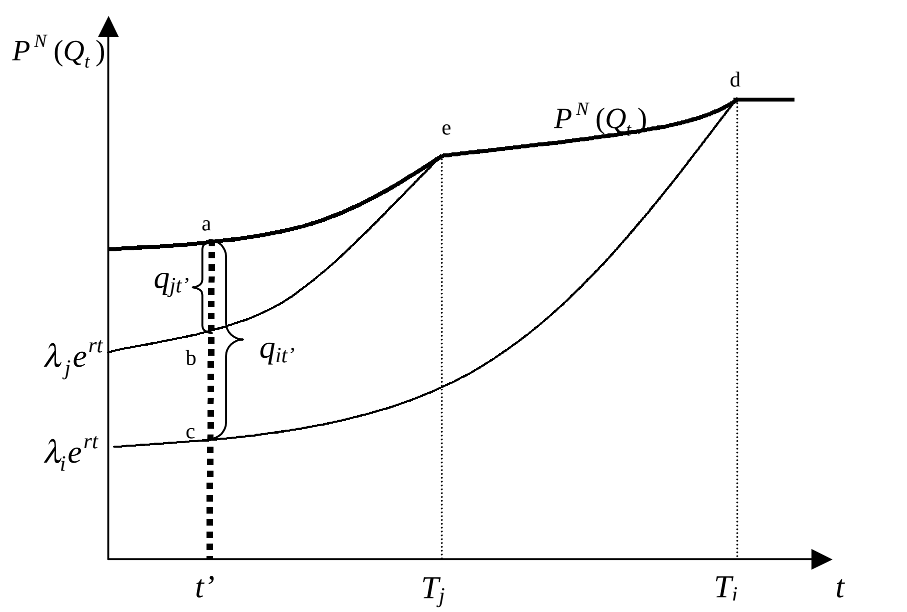
From this we can solve for the equilibrium price path

(2.30) $$P(Q^*_t) = a - Q^*_t = \frac{a + 2c + e^{rt}(\lambda_i + \lambda_j)}{3}.$$

From (2.30) we can see that prices increase in *a* (a positive shift in demand), *c* (increased costs of production), *r* (increased discount factor), *t* (prices increase over time), λ_i (the amount of reserves of producer *i* is reduced), and λ_j (the amount of reserves of producer *j* is reduced). The results can be illustrated graphically.

From (2.25a) we can see that $q_{it}^* = P(Q_t^*) - c - \lambda_i e^{rt}$, and if we let $P(Q_t^*) - c = P^N(Q_t^*)$, in words, the price net of extraction costs, we have that $q_{it}^* = P^N(Q_t^*) - \lambda_i e^{rt}$. With this in mind, we can give the figure below some intuitive explanation. In the figure we have shown the case for $S_{i0} > S_{j0}$. For all periods t we can see that $q_{it}^* > q_{jt}^*$. Up until period T_j, the development in prices is determined by a weighted average of λ_i and λ_j, while between T_j and T_i, when producer *i* is the single producer, prices is determined by λ_i alone. Hence, the growth rate of prices drops after T_j as $\lambda_i < \lambda_j$. After period T_i, when all reserves are exhausted, it is normal to assume that some kind of backstop technology is serving the market. We can also see that the area *abe* corresponds to *j*'s reserves at time t', while the area *acde* corresponds to *i*'s reserves at time t'.

Price and extraction path for homogeneous reserves (equal costs)



We will now move to a situation where we allow for differences in the marginal cost of extraction. We remember from above that under perfect competition the sequence of production is such that the order of production is fully determined by the cost of extraction. First the low cost reserves are extracted, and then production gradually moves towards the more costly reserves. The results are different under oligopoly (Polasky, 1992). Production will then be undertaken from low- and high-cost reserves simultaneously, but higher extraction costs will cause the producer to reduce the amount of production.

Assume again that we have two producers, *i* and *j*, where the marginal extraction costs, *c*, are such that $c_i > c_j$. Otherwise the situation is exactly like it was above. By solving the problem, we now find that an increase in the marginal extraction costs for producer *i* causes him to extract the reserves more slowly relative to his rival. The proof for this goes as follows (see Polasky, 1992):

First, a couple of definitions: Let us assume that we have *n* producers with equal amounts of initial reserves. Define $\alpha_t = e^{-rt} P(Q_t)$, and, for producer *i*, $\theta_{it} = e^{-rt} c_i$. We also define producer *i*'s equilibrium market share: $\sigma_{it} = q_{it}^* / Q_t^*$. We can then transform equation (2.25a) so that we get the following condition (for q_{it}>0)[12]:

(2.31) $$\alpha_t (1 - \sigma_{it} / \eta_t) - \theta_{it} = \lambda_i$$

Then, take the time derivative of (2.31) at time t, when all producers are producing. By doing this we get the following expression for the development in producer *i*'s market share over time[13]:

(2.32) $$\dot{\sigma}_{it} = \frac{\dot{\alpha}_t (\eta_t - \sigma_{it})}{\alpha_t} + \frac{\dot{\eta}_t \sigma_{it}}{\eta_t} - \frac{\eta_t \dot{\theta}_{it}}{\alpha_t}.$$

And by summing over all *N* producers we get

---

[12] It is easy to see that the expression in (25a) is equivalent to $P(Q_t^*)(1 - \frac{q_{it}^*}{\eta_t Q_t^*}) = c_i + \lambda_i e^{-rt}$. Hence, we are just using the definitions of α and θ to obtain (31).

[13] Dots denote time derivatives

(2.33) $$0 = \frac{\dot{\alpha}_t (N\eta_t - 1)}{\alpha_t} + \frac{\dot{\eta}_t}{\eta_t} - \frac{\eta_t \sum_{i=1}^{N} \dot{\theta}_{it}}{\alpha_t}.$$

Substitute (2.33) into (2.32) and simplify to get the following:

(2.34) $$\dot{\sigma}_{it} / \sigma_{it} = \frac{\dot{\alpha}_t \eta_t (1/\sigma_{it} - N_t)}{\alpha_t} + \frac{\eta_t \sum_{i=1}^{N} \dot{\theta}_{it}}{\alpha_t} - \frac{\eta_t \dot{\theta}_{it}}{\alpha_t \sigma_{it}}.$$

If we solve for the time derivative of the ratio of relative market shares we get

(2.35) $$\begin{aligned} \frac{d}{dt}\left(\frac{\sigma_{jt}}{\sigma_{it}}\right) &= \left(\frac{\dot{\sigma}_{jt}}{\sigma_{jt}} - \frac{\dot{\sigma}_{it}}{\sigma_{it}}\right)\frac{\sigma_{jt}}{\sigma_{it}} \\ &= \frac{\dot{\alpha}_t \eta_t}{\alpha_t}\left(\frac{1}{\sigma_{jt}} - \frac{1}{\sigma_{it}}\right)\frac{\sigma_{jt}}{\sigma_{it}} - \frac{\eta_t}{\alpha_t}\left(\frac{\dot{\theta}_{jt}}{\sigma_{jt}} - \frac{\dot{\theta}_{it}}{\sigma_{it}}\right)\frac{\sigma_{jt}}{\sigma_{it}} \end{aligned}$$

Now we decompose c_i into two parts, so that $c_i = c_j + \Delta_i$. By using this expression for *i*'s production costs, we can manipulate the second expression of the right hand side of (2.35) to get the following:

(2.36) $$\frac{\eta_t}{\alpha_t}\left(\frac{\dot{\theta}_{jt}}{\sigma_{jt}} - \frac{\dot{\theta}_{it}}{\sigma_{it}}\right)\frac{\sigma_{jt}}{\sigma_{it}} = \frac{\eta_t}{\alpha_t}\frac{\sigma_{jt}}{\sigma_{it}}\left[\dot{\theta}_{jt}\left(\frac{1}{\sigma_{jt}} - \frac{1}{\sigma_{it}}\right) + \frac{re^{-rt}\Delta_i}{\sigma_{it}}\right].$$

By substituting (2.36) into (2.35) we now get

(2.37) $$\frac{d}{dt}\left(\frac{\sigma_{jt}}{\sigma_{it}}\right) = \frac{\eta_t}{\alpha_t}\left[(\dot{\alpha}_t - \dot{\theta}_{jt})\left(\frac{1}{\sigma_{jt}} - \frac{1}{\sigma_{it}}\right) - \frac{re^{-rt}\Delta_i}{\sigma_{it}}\right]\frac{\sigma_{jt}}{\sigma_{it}}.$$

From this expression we can see the effect of an increase in marginal extraction costs. Since the elasticity of demand is defined in absolute terms and hence is positive, and prices increase over time, we can see that a higher cost of extraction for producer *i* means that producer *j*'s market share declines over time, and hence that the rate of extraction is lower for the high-cost producer *i* than for producer *j*.

Since we have assumed that demand is constant over time, both producers will have declining extraction paths over time (this is the only way prices can increase such that producers on the margin are indifferent between current and postponed extraction). However, the extraction path of producer *i*, the high-cost producer, will decline more slowly than the path of producer *j*. Further, the high-cost producer will reach exhaustion of reserves later than the low-cost producer. This theory might therefore explain why production currently takes place in high-cost areas, which we would of course not expect under perfect competition. However, we are unable to explain that high-cost producers are depleting their reserves prior to low-cost producers, which seems to be the case in the real world.

Again it might be useful to go through an exercise in order to see how one may calculate the extraction paths. There are some benefits from doing these calculations, as one in particular will note one interesting characteristic of the production paths that has not been noted in the literature I have studied. Let us assume a situation with two producers, R (for Russia) and SA (for Saudi Arabia). We assume that the discount rate is equal for both producers, and the same are the marginal costs of production. Further, SA has larger reserves at t_0 than does R. Equation (2.28) gives us the Cournot Nash equilibrium production levels of the two producers. In order to find the termination dates for the two producers, T_R and T_{SA}, we have to make use of a transversality condition, which says that the value of the Hamilitonian at time T_i must equal zero, or formally:

(2.38) $$H_F^i(T_i) = e^{-rT_i}(a - q_{T_i}^i - q_{T_i}^j - c)q_{T_i}^i - \lambda_i q_{T_i}^i = 0,$$

where the subscript F just denotes that we have to do with the value at termination of extraction. Since $S_R < S_{SA}$ we must have that $T_R < T_{SA}$. We also have the first order condition saying that

(2.39) $$\frac{\partial H^i}{\partial q_t^i} = e^{-rt_i}(a - 2q_t^i - q_t^j - c) - \lambda_i = 0.$$

We then set (2.38)=(2.39) and solve for λ_i and then get

(2.40) $$\lambda_i = e^{-rT_i}(a - c - q_{T_i}^i).$$

Since SA is the sole producer left at T_{SA}, this means that the condition for SA is $\lambda_{SA} = e^{-rT_{SAi}}(a-c)$, while the condition for R is $\lambda_R = e^{-rT_R}(a-c-q_{T_R}^{SA})$.

Let us now first find the termination date T_R. First we note that the production of SA at T_R is given from (2.27), saying that

$$q_{T_R}^{SA} = (a-c+\lambda_R e^{rT_R} - 2\lambda_{SA} e^{rT_R})/3. \tag{2.41}$$

If we than substitute (2.41) into (2.40) and by using the expression for λ_{SA}, we get that

$$2\lambda_R = (a-c)e^{-rT_R} + (a-c)e^{-rT_s}. \tag{2.42}$$

We then substitute (2.42) into (2.28) to find the expression for R's production path:

$$q_t^R = (a-c-(a-c)e^{r(t-T_R)})/3. \tag{2.43}$$

This equation gives us the production path for R. In order to find the date of termination, we make use of the (discrete time) resource constraint, which now tells us that

$$\sum_{t=0}^{T_R} q_t^R = \sum_{t=0}^{T_R}(a-c-(a-c)e^{r(t-T_R)})/3 = S_0^R. \tag{2.44}$$

If we want to solve explicitly for T_R the expression gets extremely complicated. However, if we make use of the formula for geometric series, we get the following:

$$T_R - \frac{1-e^{-r(T_R+1)}}{1-e^{-r}} = \frac{3S_0^R}{a-c} - 1. \tag{2.45}$$

If we then have the values of the exogenous parameters in this expression, it is to easy find the value of T_R. It is interesting to see that the terminal date for R is independent of the characteristics of SA (given that SA has more reserves than R).

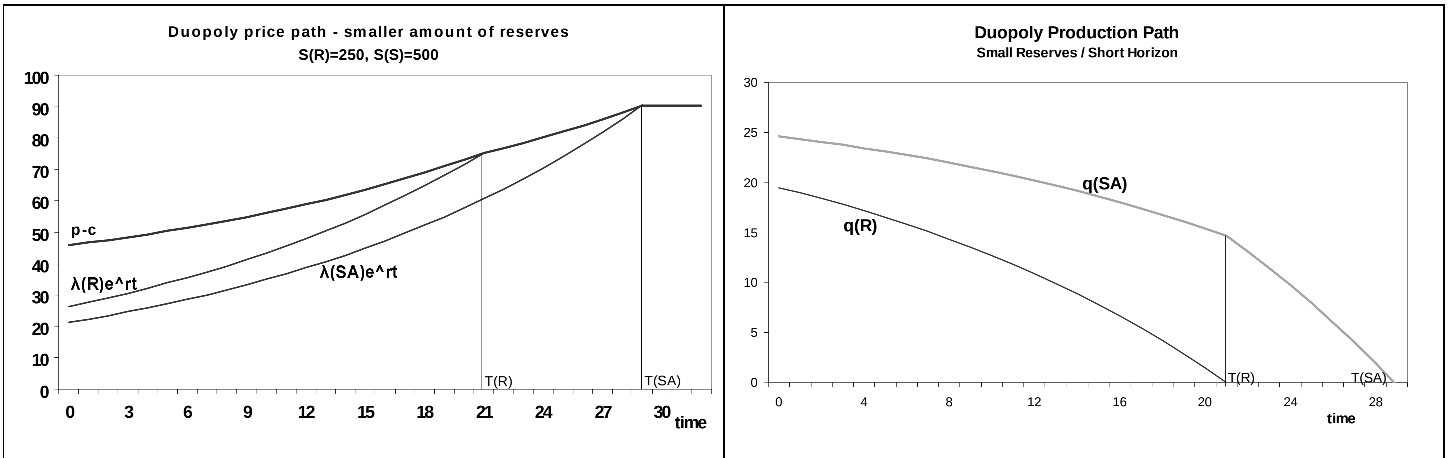
We then want to move on to find T_{SA}. It is then convenient to split the extraction period into two phases for SA, such that (by using equations 2.28 and 2.41) the resource constraint becomes

$$(2.46)\quad \sum_{t=0}^{T_R} q_t^{SA} + \sum_{t=T_R+1}^{T_{SA}} q_t^{SA} = \sum_{t=0}^{T_R} (a - c + \lambda_R e^{rt} - 2\lambda_{SA} e^{rt})/3 + \sum_{t=T_R+1}^{T_s} (a - c - \lambda_{SA} e^{rt})/2 = S_0^{SA}.$$

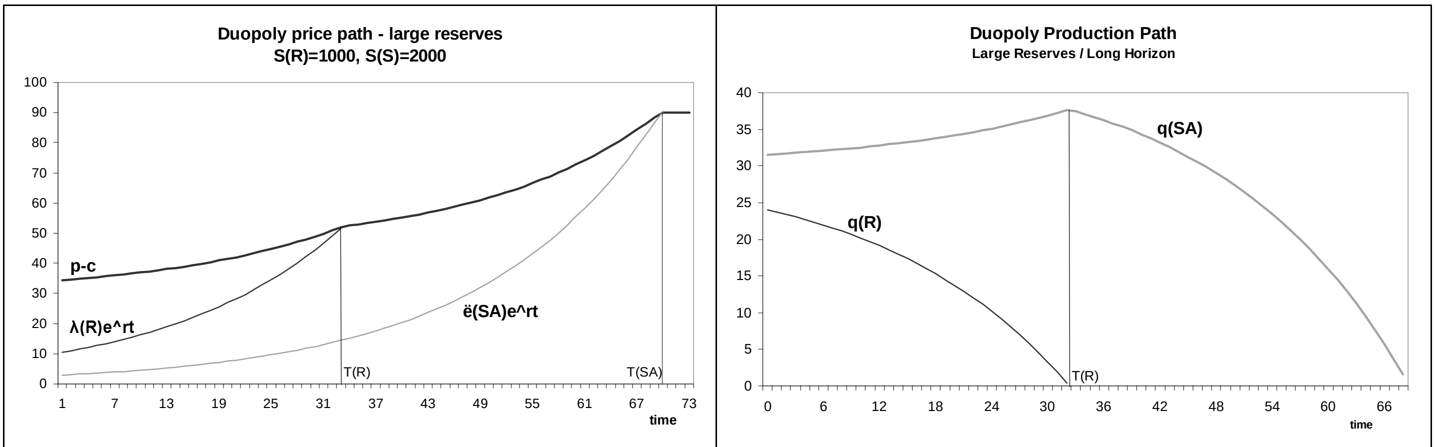
We can substitute for the resource shadow prices for R and SA, and then get an expression for which we can find T_{SA}:

$$(2.47)\qquad 3\left(T_S - \frac{1 - e^{-r(T_S+1)}}{1 - e^{-r}}\right) = \frac{6S_0^{SA}}{a - c} + T_R - \frac{1 - e^{-r(T_R+1)}}{1 - e^{-r}} - 2$$

Since the values of a, c, S_0^R, and S_0^{SA} are exogenuously given, we have then by using (2.45) and (2.47) found the dates of termination. It is then easy to find the remaining unknown variables. I have computed the results in Excel for one hypothetical situation, where r = 0,05, c = 10, a = 100, and the reserve levels varies. Below you can see two set of graphs, describing two different situations. In the first set of graphs, the reserve levels for R and SA are 250 and 500 respectively. We then get the standard solution described above, where the production levels are declining over time for both producers. We also see how there is a kink in the production path of SA at the time when the reserves of R are depleted, as SA is from then on behaving as a monopolist. In the second set of graphs, however, the situation has changed somewhat. The reserve levels are now higher, 1000 and 2000 for R and SA to be specific. The time horizon is now longer for both producers. The production path of SA is now increasing over time, however, all the time when R is still active in the market.



Set 1



Set 2

From equation (2.28) it is easy to derive the condition for when the production path will look the way it does in Set 2 above. From (2.28) we have the following interesting result:

(2.48) $$\dot{q}_t^{SA} = re^{rt}(\lambda_R - 2\lambda_{SA})/3.$$

So if $\lambda_R > 2\lambda_{SA}$ the production of SA will increase over time all the way up to T_R. This means that a producer with large reserves and a long horizon is forced to limit production initially, while it gradually increases production as that of competing producers decline.

The Cournot model has been extended further to include the situation of some large players who take the output path of a competitive fringe as given, and also to study the situation where OPEC acts as a Stackelberg leader, announcing a price path the fringe adjusts to, and then OPEC meets the residual demand. OPEC's problem is then to find the price path that maximizes profits.

A general problem of all of these models is the problem time inconsistency[14]. For instance, in a Stackelberg model with an announced price path, the optimum price path will change over time given the accumulated changes in reserves, and hence it will be optimal to deviate from the previously announced price path. In other words, the strategies are not foolproof with respect to renegotiation. Some of the models tackle these problems, but are subject to another problem: If a producer knows that the equilibrium price path will be changed in some future period, it will in general be in his interest to deviate from the current path. Finally, can you ever be sure that a producer is really following his announced path (if the output is non-observable)?

Crémer and Salehi-Isfahani (1991) conclude that the underlying oil market theory has a hard time developing model that includes the most important characteristics of the market. In order to compensate for this, simulation models can bring important perspectives on long run trends that are not easily understood. The traditional theory has also ignored some of the important questions like that of OPEC stability and credible punishment mechanisms. To this comes the question of whether the Hotelling principle should dominate the theoretical modeling strategies of the market for crude oil. This topic will be discussed now.

[14] For more on this, see Crémer and Salehi-Isfahani (1991)

Part 3 - Exhaustibility - A Relevant Assumption?

Even though the scarcity effect may in principle push prices above marginal costs, there are no consensus as to how large this effect is, and if it is present at all. Of course, geologically there is a fixed amount of oil reserves in the world, so if one wants to argue against a substantial scarcity rent, one must in some way argue that the scarcity is less important than in traditional Hotelling models.

One way to do this is to consider the uncertainty with respect to remaining reserves and argue that the experienced degree of scarcity is not increasing, but rather declining over time. The table below illustrates the development in the proven reserves to production ratio, which tells us in how many years we can expect to run out of crude oil if production continues at a constant pace and if more reserves are not found.

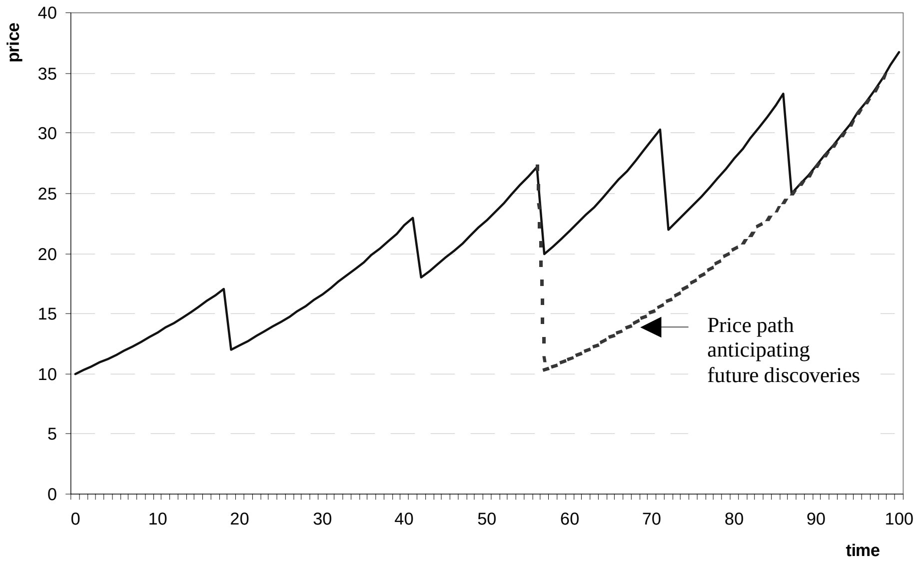
Table: Development Reserves-Production Ratio

Year	Proven Reserves	Production	R/P ratio
1970	533	17,4	30,6
1975	602	20,2	29,8
1980	608	22,8	26,7
1985	699	20,5	34,1
1990	969	23,8	40,7
1993	1007	23,7	42,5

Source: Claes (2001)

The table shows us that time and time again, the crude oil reserves have not only been replenished relative to production, the ratio reserves to production has actually increased over time. The basic assumption of the Hotelling model is the fixed stock of initial reserves, but after steadily increasing world production over the past 30 years, the remaining proven reserves are currently twice as large as they were in 1970. In fact, the more oil that has been extracted, the higher the estimate of remaining proven reserves has become. These basic facts run contrary to the basic assumptions of the Hotelling model, and it therefore seems natural to question its validity for characterizing the oil market. Claes (2001) argue that the large price increases of the 1970s have made academics adopt the Hotelling rule. These price surges were taken to signal scarcity, and hence the prediction was that oil prices would only continue to rise. This prediction has not materialized, however, as the real oil prices are far lower now than they were in the 1970s. In fact, the prices of minerals in general have not risen over the past decades. Schmidt (1988) has studied this more carefully. He finds that the real prices of several important exhaustible resources have not appreciated over the past century. Several factors may explain this lack of appreciation. These explanations do not in principle provide a reason to reject the use of the Hotelling principle when modeling markets for exhaustible resources in general, but they point out in how many ways the principle must be refined if the predictions of the theory are to be reliable. First, as we have seen, mineral deposits with different extraction costs may cause price growth to fall. Still, this explanation cannot explain the absence of price growth, only its slower pace. Second, unanticipated changes in available reserves, a possibility that is ignored in the basic Hotelling model, causes a shift among optimal price paths. As shown by for instance Arrow and Chang (1982), a sudden, unanticipated increase in proven reserves causes the price trajectory to fall in order to assure full resource exhaustion. Observed prices in these models fall sharply when the discovery is made. The evidence with respect to this latter point is mixed, however, but sometimes announcements of large new deposits have caused prices to move, though the immediate response is often small. With a series of substantial unanticipated discoveries we would expect to see a price path like the one represented by the solid line in the graph below. However, if producers after several discoveries would begin to anticipate more discoveries, a price path like the one represented by the dotted line would be the appropriate one.

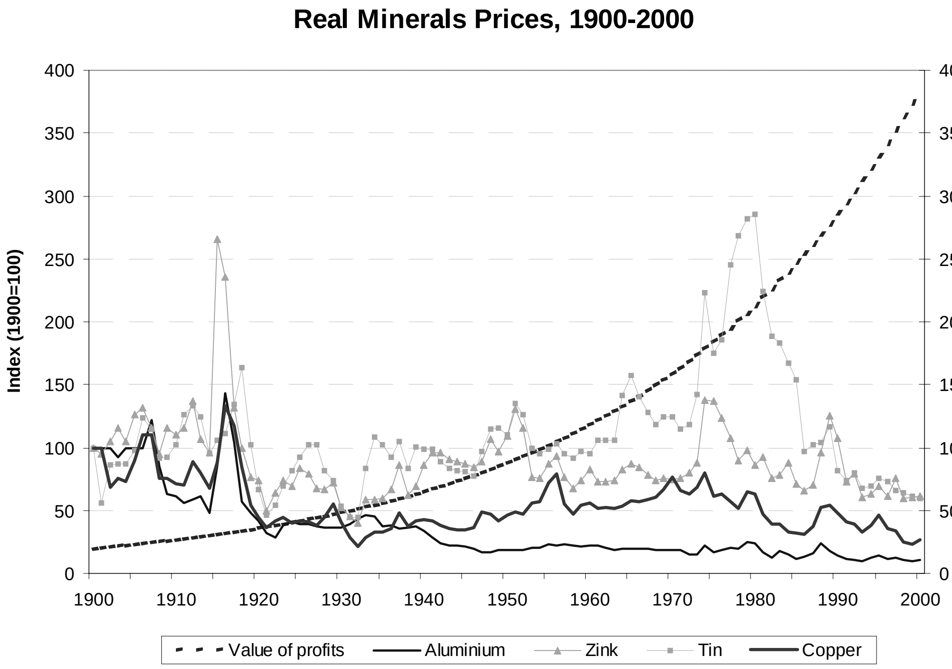
Price Path with Recurrent Unanticipated Discoveries



The third explanation of the lack of price increases relates to informational problems and uncertainty in demand and the supply of substitutes. This may cause both large fluctuations in prices and regular prices declines. For instance, producers of an exhaustible resource must find a price path that maximizes profits subject to the expected response from the producers of a substitute. If the producers underestimate the potential of the substitute, the response of the substitute may force the producer to shift to a lower price path. This also works in the opposite direction, that is, the price path may shift upwards if the producers have overestimated the potential of the substitute. Hence, the absence of perfect foresight may cause prices to fluctuate considerably. Further, miscalculations of the long-run elasticity of demand may cause the price to rise so rapidly as to encourage, i), an increase in capacity in the production of the exhaustible resource, ii), the production of substitutes, iii) and the substitution of demand away from the exhaustible resource. Fourth, changes in the institutional structure of the market may cause movements in the resources prices. For instance, several of the markets of exhaustible resources are cartelized, and in other markets one cannot exclude the possibility that resources that are state-owned are extracted with not only profit maximization being among the chief objectives. We remember from above that the price paths of monopolized markets, except under special assumptions, are different than those under perfect competition. Hence, varying degrees of cohesiveness in cartels may cause shifts between different price paths, and cause considerable price volatility.

All in all, Schmidt’s arguments show us that in principle the Hotelling model may be useful when analyzing the markets for exhaustible resources, but the theory must be refined considerably and take many aspects into consideration. If one believes in the fundamental prediction of continuously increasing prices, predictions are likely to be extremely vulnerable to unanticipated changes in market conditions.

Julian Simon and others have strongly criticized the basic premise of the Hotelling model, namely that the prices of exhaustible resources must rise over time. First of all, the trend of natural resources prices show exactly the opposite of what is predicted by Hotelling’s theory. Theory says that a producer must be indifferent between selling today and tomorrow. I will now go through a simple example to illustrate what a terrible investment it would have been to hold on to a marginal unit of an exhaustible resource over the past 100 years. Let us assume that the scarcity rent is just 20 percent of the market price of the extracted resource. This means that by selling for a price equal to 100 today you are left with a profit of 20. You then invest this for a modest 3 percent real interest rate. The graph below compares actual price paths of various exhaustible resources from year 1900 to 2000 to the implied path of the value of the profits by selling the resource in 1900. The figure shows clearly that an investment in a natural resource would have been a disastrous investment.



Data : United States Geological Service

Hotelling's analysis also bears some of the characteristics of Malthus' analysis of the effects of population growth on welfare. As Malthus did not foresee that mankind was able to make rapid progress, so Hotelling assumes that man is unable, through technological progress and the discovery of substitutes or through discoveries of additional reserves, to find a way around the eventual scarcity of a good. Long-term predictions are often proved invalid due to the labor and creativity of the human mind. When a problem is pressing, man finds a solution that makes things better than they were during the arrival of the problem. One might therefore argue that the exhaustibility of oil reserves are relevant only in one aspect, namely that it will force man to find a solution and make the exhaustibility irrelevant.

There is also a dispute as to whether petroleum products really are fossil fuels. Some researchers, and especially the astrophysicist Thomas Gold, claim that hydrocarbons are derived from materials incorporated in the mantle at the time of the Earth's formation. Gold and others claim that the traces of fossils that can be found in for instance crude oil are coincidences rather than a proof that biological debris is the origin of crude oil and natural gas. This hypothesis is supported by a discovery of oil in Siljan Ring in Sweden in a place deep under ground were there could never have been any fossils. Other discoveries in recent years are also giving this hypothesis increased credibility. For instance, evidence from the Gulf of Mexico suggests that some old oil fields are being refilled by petroleum surging up from deep below. On the other hand, even though there may be some truth in these claims, the question remains as to whether they imply a substantial increase in the amount of recoverable reserves.

Replacement Cost

In any case, fossil fuels as the origin of hydrocarbons or not, it is hard to justify modeling the market for crude oil with the assumption that oil is a fixed scarce resource. Investments in exploration activities are continuously replacing all and even more than the oil that is extracted. Adelman (1993b) argues that it might be more fruitful to think of proven oil reserves as inventory rather than as a fixed stock, and that the inventory should be measured in how many years of production at the current rate it provides. He then argues that the replacement cost of this inventory should be the true measure of scarcity, and not until the replacement cost surpasses the value of finding reserves will the fixed stock be the relevant point of departure. “The early warning signal of scarcity is a persistent rise in development costs and in the in-ground value of oil reserves” (Adelman, 1993a, 276). In some region it may be fair to say that these warning signals have arrived, for instance in the US and in Norway, where the exploration activities are becoming less successful and the size of recently discovered oil fields are smaller than they used to be. In theses cases it is fair to talk of an approximately fixed stock. Here, the replacement costs are increasing over time, as the average size of newly found deposits is decreasing, and as the better resources are extracted first. However, as Adelman argues, this is not the case for the most important class of producers over the long term, that is, the OPEC countries. For a group of ten OPEC countries, there are no signs yet that replacement costs are increasing. The table below, based on estimates from OPEC countries, indicates that costs actually may be falling in the regions where most of the world’s known oil reserves are located.[15]

[15] The data is a bit old, due to problems of collecting new data of this kind. Adelman (1993b) reports that the former providers of this type of data are no longer operating.

Table: Investment per Additional Daily Barrel of Capacity ($)
Ten OPEC Nations, 1985-1990

Country	1985	1989/90
United Arab Emirates	2912	2033
Algeria	4570	3533
Indonesia	5908	5637
Iran	NA	1433
Iraq	NA	145
Kuwait	1455	895
Libya	24199	NA
Nigeria	3179	2611
Qatar	1331	1408
Saudi Arabia	692	373

Source: Adelman (1993b)

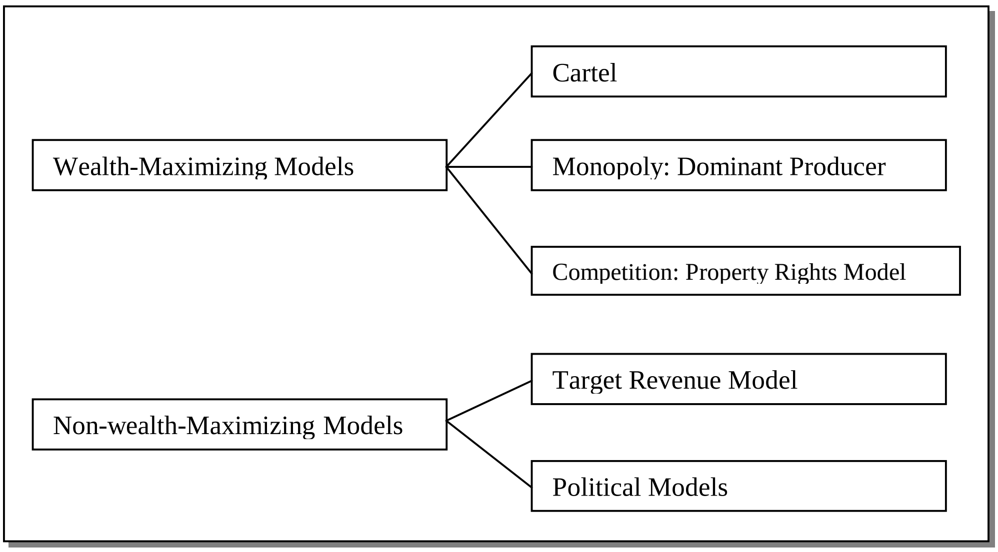
Even though the current situation does not look too troubling with respect to the costs of replenishing the reserves, a time will of course come, as long as oil continues to be a commodity with substantial demand, when the fixed-stock effect will start to kick in and contribute to price increases. Even though the hydrocarbons are far more plentiful than hitherto assumed, the cost of exploration and extraction of the less accessible oil will be far higher than for instance the costs in the Middle East. The information and arguments presented here simply illustrate that it does not seem likely that we are nowhere near this point yet, and, hence, that it does not seem reasonable to take the Hotelling rule as the one and only truth with respect to crude oil production- and price paths. In the long run, when replacement costs are becoming prohibitive, one would expect the Hotelling rule to reign. Until then, we must try to figure out the factors that are determining the market dynamics under the current environment.

Adelman has proposed new modeling strategies as to how to include this concept of replacement cost. I will not go into these models here, but the underlying idea can be found in Cairns and Davis (2001).

Part 4 - 'Real World' Models of the Oil Market

It is hard to avoid the suspicion that the jump in prices during the 1970s stems from OPEC's gradual transition to an effective cartel, and that the member countries managed to hold each other's expansion impulses in check. The price increases during the 1970s have been thoroughly analyzed, and the development of OPEC and changes in the market power of the producers in the industry are indeed commonly used as explanations. However, things are not necessarily as straight forward as this, which we will see below. In the next sections of this paper the behavior of OPEC will be analyzed in more detail. In this section we will describe some of the main models that have been used to explain the sudden jump in prices in the 1970s. We will currently leave the non-OPEC producers to themselves, and just assume that they operate as a competitive fringe. The models of OPEC behavior differ from the preceding theoretical models in that the theories are less formal, and more concerned with the institutional details of the market. As these models avoid the technical difficulties of rigorous mathematical formulations, they can address directly the questions that puzzle the analysts of the oil market. It should also be noted that these models are less concerned with the implications of exhaustibility.

Griffin and Teece (1982) have classified the models of OPEC behavior according to the figure below[16]:

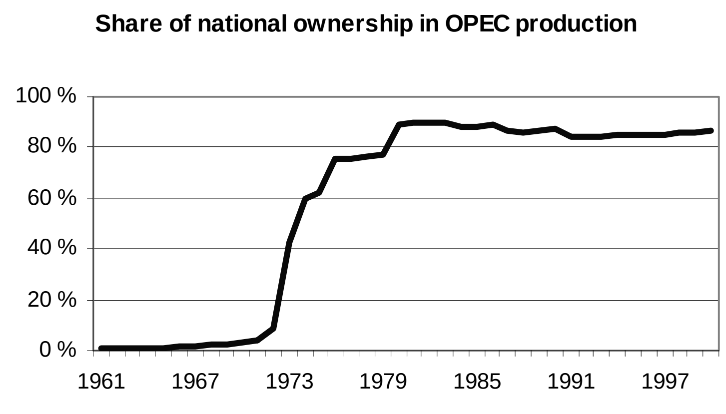


[16] Actually, this is not completely true. The Cartel variant of Wealth Maximizing Models has been added. The illustration by Griffin and Teece was made prior to OPEC's implementation of output quotas, and a traditional cartel theory did not seem likely as an explanation of the price movements during the 1970s.

The cartel and dominant-producer models will be thoroughly discussed in the sections below. By just remarking that the Dominant Producer Model regards Saudi Arabia as the dominant producer that is acting as a swing producer, the basic intuition behind these two models should be clear. However, before discussing these quite familiar conceptions of OPEC's behavior, we should illustrate the intuition behind the remaining three models of the framework above. These models contain one common denominator, namely that the participants in OPEC in some aspects differ from regular profit maximizing firms. First of all, one cannot accept the notion of OPEC producers as agents with profit maximization as the sole objective without apology. Second, differences in the discount rates of private and sovereign owners may cause large fluctuations in the price paths when the ownership shifts from the private to the public, or vice versa. In a context where we are dealing with countries located in one of the most conflict-laden regions in the world, economists might well do itself a service if they are at least open to suggestions that other things than pure profit maximization are among the relevant objectives for the producers.

Property Rights Model

Let us first look at the Property Rights Model. The hypothesis presented in this model is that it is not the monopoly power of OPEC but rather the alteration of property rights that caused the spike in prices in the 1970s (Johany, 1979). The argument supporting this hypothesis is that prior to the increase in prices a de facto nationalization of the private oil companies' crude oil resources in all Middle East countries took place (see figure below).



Source: OPEC Annual Statistical Bulletin

In October 1973 the oil producing countries decided to set the price of their oil unilaterally rather than through negotiations with the oil companies, which was what had been done in the past. This happened after the property rights in crude oil deposits had shifted from the companies to the countries, which were now able to decide both how much to produce and what price to charge. The underlying argument of the Property Rights Model consists in this transition in which the host countries suddenly had obtained control of the relevant decision variables. Let us assume that both regimes, that is, before and after nationalization, maximize the discounted present value of profits by choosing a time path for output in the same manner as argued under the Hotelling rule. Let us further assume that marginal costs of production are constant *(dC/dq=c, dC/dt=0).* Then, in the absence of uncertainty, the arbitrage condition states that the marginal barrel in ground will appreciate at the same rate as the discount rate, that is

$$b(t+1) = b(t)(1+r).$$

However, Johany argues that the oil companies prior to 1973 knew that there existed some risk, dependent upon country and regime, that their deposits would be expropriated. Let us fix the probability that the property rights are preserved at *p*. The expected value of the marginal barrel of oil left in ground is then:

$$E[b(t+1)] = pb(t+1) + (1-p)0$$
$$\Rightarrow E[b(t+1)] = pb(t+1)$$

This means that positive amounts of oil will be supplied by the companies in the two periods only if

$$pb(t+1) = b(t)(1+r)$$
$$\Rightarrow b(t+1) = b(t)\frac{(1+r)}{p}$$

As long as $p<1$, this uncertainty with respect to property rights increases the effective discount rate of the private companies. For instance, if *r=0,1*, and *p=0,75*, the effective discount rate will be 47 percent. This increase in the discount rate will make the oil companies follow paths where output is extracted more rapidly due to the impatience implied by the high discount rate. The effect is, in other words, that the risks of expropriation kept the oil output artificially high prior to 1973. This means that a shift from the oil companies' regime to the countries' regime would mean that the rate of extraction would be significantly reduced and cause the prices to rise sharply. There would have been no need for collective action for this to happen, indeed, Johany argues against this being the cause of the price increases.

Even though this model seems to explain some of the dynamics in 1973/1974, the experience after this first oil price shock does not seem to support the theory. First of all, the theory assumes perfect competition. How do you then explain, as Adelman puts it, why the water is flowing uphill? Second, oil prices have not settled on a new stable path after 1973. The price volatility has been substantial. This volatility can partially be explained by disruptions due to political events, but it is hard to argue that large producers like Saudi Arabia does not have the ability to influence prices. In some way or the other one must allow for some sort of market failure.

The Target Revenue Model

The Target Revenue Model retains the assumption of perfect competition. Its basic assumption is almost like an inverse of the oil companies' fear of expropriation in the Property Rights Model. In both of these models there are contract uncertainties. The situation in the Target Revenue Model is that the OPEC countries fear that financial investments abroad run the risk of being confiscated by the recipient, alternatively that they will be used for political pressure [see for instance Teece (1982) or Alhajji and Huettner (2000)]. This fear of expropriation may lead to a backward bending supply curve. The logic behind the Target Revenue Model can be described as follows. First, let us assume that the government is in control of both the resources and the production. Let us further assume that oil is the dominant export industry of the country. As foreign financial investments cannot work as a buffer if oil export revenues are either too high or too low, the receipts must be invested in the domestic economies. However, one does not want to see these investments reaching a level that pushes the marginal productivity of capital too low. Hence, the first step in the decision making process is to determine the optimal level of investments given a threshold rate of return level. Once these investment needs are identified, an isorevenue curve showing the different combinations of prices and quantities giving a constant amount of oil export revenues can be constructed. Remember that the theory assumes perfect competition, so the producers are not individually affecting prices by changing production, they are just adjusting production levels to existing market prices. The point is that the aggregate effect of many producers responding in the same manner to changes in prices may cause the supply curve to be backward bending. If we let the investment needs be represented by *I**, the isorevenue curve will look like the hyperbola *P=I*/Q* in figure 1 below. This curve will also represent an individual producer's supply curve for output levels less than full capacity. Further, if we let the demand curve facing an individual producer be represented by d_x^0, we see from figure 1 on page 47 that the producer's desired output level will be *q**. Due to some degree of quality and location differences the demand curve will be downward sloping, but as we regard the producers to be small relative to the market, demand will be highly elastic.

Figure 2 illustrates the situation where the target revenue is high, as represented by *I***. Production will then be adjusted to the capacity level $\overline{q}$. At this output level revenues will fall short of the target revenue by the amount of the shaded area. With a constant capacity level, production will then remain at capacity, and the (short run) supply curve will be vertical. The target revenue can only be reached if the world supply and demand conditions changes such that the demand faced by the individual country, d_x^0, increases.

The target revenue model implies that the individual producer's supply curve is backward bending for prices above those that are needed to support the target revenue. Hence, as seen in figure 3 below, production will fall if demand and world prices increases. Figure 4 shows the case where an increase in demand is coupled with an increase in absorptive capacity, and this is the only case for which both output and prices can increase in this model. The curve S_1 here represents the long run path of supply and demand, where positive shifts in absorptive capacity and world oil demand over time makes this path upward sloping.

An important implication of the target revenue model is that the market supply curve may be backward bending. This will be the case if production decisions in a sufficient number of countries are well described by this model. This in turn creates the opportunity for multiple equilibria, as shown in figure 6. This, however, demands that demand is quite inelastic. With more elastic supply a single equilibrium will be ensured, as shown in figure 5. An elastic demand curve for the individual country seems to be the most realistic case, as the demand a producer will face if he were to set his price above the world price would disappear rapidly. The main point here, however, is that the nature of the supply curve makes the price of oil subject to substantial instability, as small changes in demand translates into considerable fluctuations in prices.

The Target Revenue Model; Figures 1-6

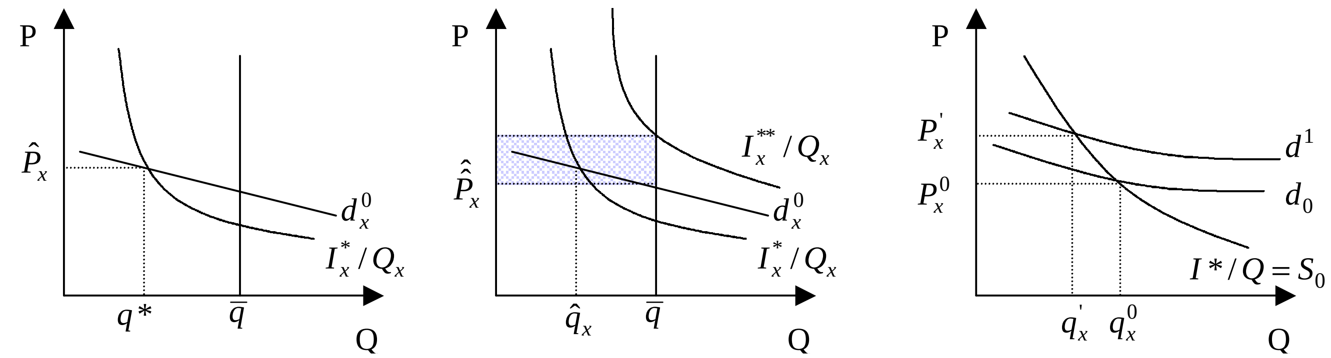


Figure 1 **Figure 2** **Figure 3**

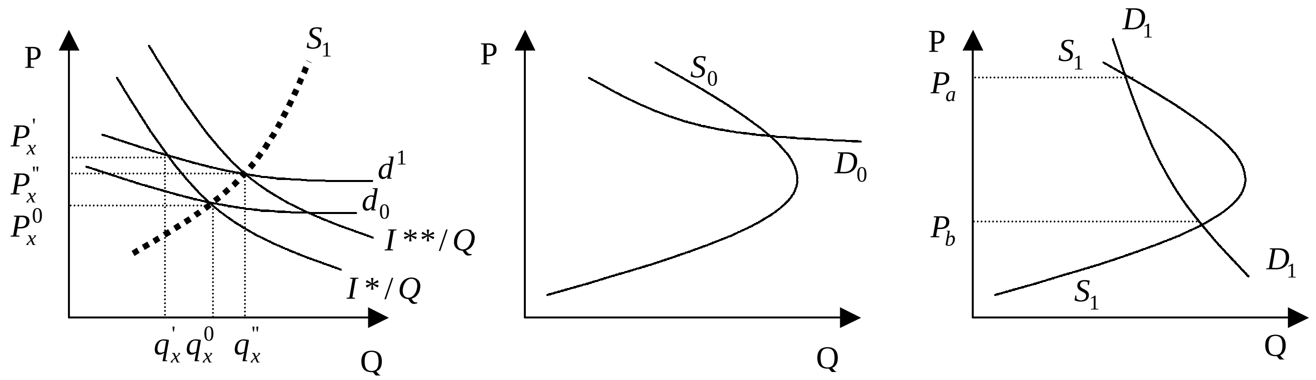


Figure 4 **Figure 5** **Figure 6**

Another implication of the target revenue model is that world oil prices could rise without any concerted effort by the OPEC countries to limit supply. In other words, no cartel is needed for oil prices to increase, but as production is increased and decreased at the same time by target revenue producers, OPEC might look like a cartel even though this is not the case. For the producers following the target revenue principle the objective will not be to maximize profits. Profit maximization would make cheating on price levels or production quotas more likely, but since in this model high prices do not stimulate to higher but rather lower output, no effort is needed in order to preserve the prevailing price level.

This theory might explain why the world price of oil remained high from 1973 to the early 1980s. During this period OPEC tried to administer prices rather than output. The Saudi Arabian Light served as the marker crude, and all prices were calculated from this marker price by using formulas that specified price differences due to quality and location differences. However, in reality the countries were free to sell their crude oil for whatever price they found appropriate. There was no mechanism that could police the pricing decisions of the individual OPEC member countries. Hence, it seems farfetched to argue that it was the success of a cartel that drove the price movements in the oil market during the 1970s. Rather it seems likely that other factors explain the price movements. Aside from political factors, the target revenue model seems like a feasible explanation. This is not to say that we would expect all countries to be restricted by absorption limits. Rather, the constraint will be more relevant the lower the level of economic development and the smaller the population relative to oil revenues. Teece (1982) categorize the OPEC countries into three groups according to these two criteria. Teece finds evidence for the hypothesis that the group of countries with low absorptive capacity (Libya, Kuwait, Qatar, United Arab Emirates, and Saudi Arabia) during the 1970s behaved according to the target revenue model. Saudi Arabia is the exception, and this is hardly a surprise due to its importance as a stabilizing force during the 1970s. For the rest of the group, the Target Revenue Model is supported by several findings. First, these countries did not give price discounts in order to increase sales. Second, public statements by OPEC ministers indicated that budgetary needs to a large extent determined production decisions. Finally, the financial assets abroad did not grow as rapidly in these countries as in others. Kuwait is an exception here, but its production relative to proven reserves was the lowest among the OPEC members, with a reserve ratio of above 100 years.

Alhajji and Huettner (2000) have reviewed the evidence of the target revenue model. The experience from the 1980s does not seem to support the model as for instance the decline in prices during the mid 1980s did not lead to increased production levels. One might also suspect the assumption that financial investment abroad not is an option, but the danger of confiscation exists at least in the cases of Iran, Iraq, and Libya. On the other hand, all OPEC countries have financial investments abroad, but the amounts are in most cases limited. By performing various statistical tests over the period 1970-1994 for all the OPEC countries and for some non-OPEC countries with centrally controlled production, they find limited evidence of the model. The common denominator of the tests seems to be that Libya, Nigeria, and perhaps also Algeria act according to the model. For these countries the tests show that the supply curve is backward bending. One interesting finding is that centrally planned economies are far more likely to follow the behavior the target revenue model predicts. By using the Heritage Foundation’s Economic Freedom Index, they find that a low degree freedom is highly correlated with behavior corresponding to the predictions of the model.

All in all, one might conclude that the target revenue approach seems suitable for some of the OPEC countries. The results indicate that profit maximization alone is not the sole objective of oil producers, and that no single model is able to explain the behavior of all oil producers.

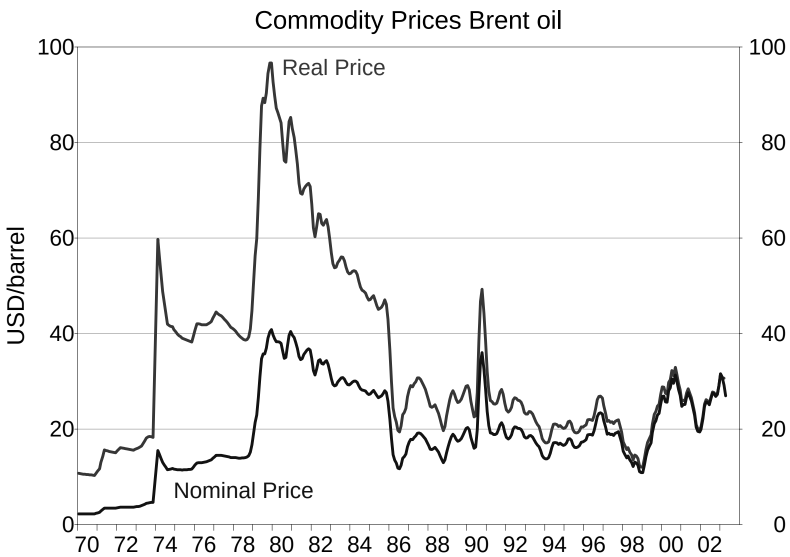
Political Models

We will now take a look at the second of the two non-wealth-maximizing models of the oil market, that is, models that take political factors into consideration. Political Models argue that, first of all, one must accept that crude oil is a resource of great strategic importance. It has been used as a political weapon in several instances, and secure access to oil is of great importance for most economies in the world, not the least the United States. This means that oil producers do not have the luxury to operate as producers with only a single objective in mind. Rather, wealth maximization must be combined with achieving political influence and to assure national security. To the extent that the other constraints are binding, production will be influenced by these considerations, and the behavior will not be in full accordance with narrow-minded profit maximization.

It might be argued that one should be careful about not characterizing the behavior as maximizing. For instance, a producer that maximizes profits subject to the constraints that he faces is doing everything he can do. The fact that an oil producer is facing constraints an ordinary firm does not face, does not by itself allow us to say that profits are not maximized. On the other hand, when sovereign nations are acting as producers, the producing strategy must be part of a strategy aimed at maximizing welfare. If the risk of being attacked by a despot is large, and such an attack will be extremely costly in terms of lost revenues and damaged facilities, the producer must behave according to the existence of this threat. It might be worthwhile to make concessions to the dangerous party in order to avoid an attack, even though this alters production relative to the non-constraint case. Or, a producer might face an opportunity of getting favors in return for certain behavior, which seems to have been the case with the relationship between Saudi Arabia and the United States. If this exchange of services increases welfare relative to some alternative, it will be preferable relative to a strategy aimed at maximizing profits in isolation. On the other hand, behavior that for instance is guided by a desire to achieve power for power’s sake cannot even be described as welfare maximizing.

The problem with political models is that they are qualitative in nature. It is hard to make predictions with models that include issues of security and political power. For instance, political scientists have argued that the economic factors determining production operate within a band created by political and security factors. The question then arises as to how these bands are determined. Still, it should not be regarded as hopeless to include at least some political factors into models of the oil market. If one for instance thinks of standard political economy models where politicians set policies according to the public goods bundle that maximizes the chances of staying in power, it seems clear that the freedom of maneuver is quite limited, especially when oil revenues provide the governments with the main source of funding for these public goods. In the latter parts of this paper I will discuss how risk aversion can influence producer behavior. Political factors are to some extent determining the level of risk aversion. For instance, I will claim that the ability of the regime of Saudi Arabia to provide its people with public goods is an important factor that limits the country's feasible set of production strategies.

We should now describe some of the political factors that have influenced the production decisions by OPEC producers and the path of crude oil prices. If we first take another look at the graph below, showing the development of nominal and real Brent crude oil prices, it is clear that the following political factors have influenced the movements in oil prices dramatically:



Data: Ecowin

First, the initial oil price shock of 1973 (which was precipitated by the Arab-Israeli War and the embargo of all crude oil shipments to the US and the Netherlands, but which we remember also can be explained by the Property Rights Model) is one example. This shock has been succeeded by the Iranian Revolution in 1979 (which removed about 2 mbd of production capacity from the market and came just as secret price discounting was becoming more prevalent), the outbreak of the Iran-Iraq war in 1980 (which removed an additional 4 mbd of OPEC capacity from the market and brought a new sense of panic), US sanctions against Iran (which until very recently have discouraged investment needed to rebuild Iran's pre-revolution export capacity), the Iraqi invasion of Kuwait in 1990 (which removed nearly all Kuwaiti capacity from the market), and the ensuing UN sanctions against Iraq (which to this day have prevented Iraq from fully restoring its pre-1990 export capability). If we add Saudi Arabia's 1985-departure from its swing-producer policy, and the Venezuelan pre-Chavez regime's commitment to increase production followed by the aggressive Saudi Arabian response, we have accounted for the main fluctuations in crude oil prices over the past three decades.

We are now about to move to the next part of this paper. However, we should bear all these political developments in mind when we discuss the existence of a cartel in the crude oil market. Perhaps the hardest job of any cartel is to keep the members disciplined in order to prevent cheating. The political developments during the past decades have in many instances worked in the OPEC countries' interest, and this could be a reason why the stability of OPEC as an organization has been sustained. With recurrent political events disrupting the smooth functioning of the oil market we should ask ourselves whether this perhaps helps sustain an established cartel by reducing the temptation to cheat. On the other hand, external shocks might reduce the need for an effective cartel altogether. Still, an organization called OPEC has been with us for about four decades now, and most people, both economists and others, seem to be followers of the belief that the concerted action of this organization is the major driver of production and price trends.

Part 5 – Collusion in the Oil Market

> In the April 14 (1969) floor debate, Senators Proxmire of Wisconsin and Long of Louisiana crossed swords thusly: Long: "Does the senator know what OPEC is?" Proxmire: " Would the senator from Louisiana like to say what it is?" Long: "First, I'd like to know if the senator knows. What is OPEC? Does the senator know?" Proxmire: "Offhand, no, I do not." Long: "Well…" Proxmire: "Without burdening the senator further with the full name of this organization, which frankly escapes me at the moment, would the senator like to know who they are?" Proxmire: "I'm glad that neither the senator from Louisiana nor the senator from Wisconsin know what OPEC means. That should be clear by now." Long: "May I tell the senator who they are?" Proxmire: "I should be delighted to hear." Long: "They are the Arab countries that sell us oil out of the Near East and Libya. That does not include Venezuela. But Venezuela does business with them. It works hand in hand with them."[17]

Today most people today will know that OPEC's full name is the Organization of Petroleum Exporting Countries[18], and it must be reckoned as a major force in world politics and in the world economy. The notion that OPEC is an effective cartel is widespread. Even Robert Pindyck, who has studied the oil market thoroughly, does not hesitate to state that OPEC is a cartel in his introductory textbook, *Microeconomics*, co-authored with Daniel Rubinfeld. Pindyck and Rubinfeld (2001) state that a market is cartelized if 'some or all firms explicitly collude, coordinating prices and output levels to maximize joint profits'. The oil market fits this definition, doesn't it? Producers of a substantial amount of the industry's output quite explicitly and voluntarily set production quotas in order to gain the most from their position as both suppliers of an important exhaustible commodity and as owners of a great part of the industry's remaining reserves. Still, is this the complete picture? First of all we should note that producers might collude without doing so explicitly. We will see that OPEC lacks many of the characteristics of an explicit cartel, but this does not prevent the producers from

[17] Oil and Gas Journal, May 5, 1969, p. 93. I owe this quote to Loderer (1985)

[18] OPEC was founded in 1960, but was not very effective until 1973-74. Current members are (year of entry in parenthesis): Saudi Arabia (1960), Kuwait (1960), Iran (1960), Iraq (1960), Venezuela (1960), Qatar (1961), Libya (1962), Indonesia (1962), Abu Dhabi/United Arab Emirates (1967), Algeria (1969), Nigeria (1971). Equador left OPEC in 1992, and Gabon followed suit in 1995. Iraqi oil exports are approved by the United Nations under the oil-for-food program for Iraq established by Security Council Resolution 986 (April 1995) and subsequent resolutions. As a result, Iraqi production and exports have not been a part of any recent OPEC agreements.

colluding implicitly. Still, in order to get a sense of the workings of OPEC we have to combine theory with some data. In this section we will first take a look at the underlying theory for collusion and punishment mechanisms, and this will be followed by an evaluation of the ‘OPEC-as-a-cartel’ hypothesis.

Collusion – Underlying Theory

All producers that are trying to maximize profits want to limit the amount of competition they face in the market. This can be done in several ways, for instance through product differentiation, patent rights, or simply by joining forces with other producers in the market and, tacitly or explicitly, negotiate prices and output levels so as to maximize the joint profit of the group of producers. Of course, explicit collusion among companies is prohibited by antitrust laws in all developed economies. Still, nothing prevents countries, or companies owned or controlled by sovereign nations, from forming cartels. International cartels dominate the activity in several global markets, for instance the markets for tin, rubber, diamonds, coffee, bauxite, and supposedly also crude oil. These products are quite homogenous by nature, and in the absence of collusion one would then expect competition to be effective in pushing prices closer to marginal costs. Even though in theory the degree of product differentiation has an ambiguous effect on the possibility of collusion, in practice we see that most successful cartels are found in markets with homogenous products.

Most cartels break down, and most are predicted to collapse when they are formed. For instance, when OPEC effectively started their cooperation in 1973/74, economists’ general reaction was that a break up was imminent. However, OPEC is seemingly still going strong. In order to evaluate the history of OPEC and also to be able to make some predictions about its long-term sustainability, we need to know some of the factors that facilitate and hinders collusive behavior. In general, we know that factors facilitating collusion include: i), relatively few competing producers, ii), a high degree of patience/ low discount rate, iii) rapid detection of defection, iv), the (credible) response to defection is aggressive, v) mutual understanding of each others strategies, and, vi), high barriers to entry. These characteristics are important both for explicit and tacit collusion, and in this section we will see why.

In joining forces in order to maximize joint profits, the members of a cartel face several difficulties. Osborne (1976) classifies these problems into two groups; the external problem and the internal problems. The external problem, which we will ignore for the moment, is to predict and possibly discourage nonmember production, and then to adjust cartel-production to non-cartel production. There are four internal problems. First, the cartel will have to locate the contract surface, and then, secondly, they will have to agree upon a point on this surface. The third problem is how to detect cheating, and the fourth problem how to deter and punish cheating.

Let us first discuss these problems in a static model, with a cartel composed of just two members. This should make us able to get an understanding of the basic dilemmas facing colluding players.

How to detect the contract surface and agree upon a point on this are serious problem. As long as the member have different profit functions, it is hard to solve the production-sharing problem in a manner everybody can accept. Even before opportunistic cheating comes into play, some member will feel that they have not got their fair share of production, and they will be tempted to cheat. However, even if the sharing problem is solved adequately, the members will still have strong incentives to cheat due to the positive marginal profits (that is, given that the other members' keep their promises, it is in every member's interest to cheat on his own promise). We can explain this in figure 1 in the following manner: x^q_i is the output quota for member i $(i=1,2)$, and q is the point that the cartel has selected as the best on the contract curve. $I^1_{i}, \ldots, I^5_i$ are parts of the two members' isoprofit curves, with $I^1_i > I^2_i > I^3_i$,and so on. The profit function of member i at the cartel point q is $\Pi_i\ (x_1^q\ ,x_2^q\ )$. At this point the marginal profit of increasing production is positive, that is

(5.1) $$\frac{\partial \pi_i (x_1^q, x_2^q)}{\partial x_i} > 0$$

This can be understood by studying the figure. If we are in the cartel equilibrium q, and member 1 expects member 2 to stick to his quota in the following period, member 1 has the opportunity to maximize his profit by producing x_1^c. This means that

(5.2) $$\begin{aligned} \pi_1 (x_1^c, x_2^q) &> \pi_1 (x_1^q, x_2^q) \\ \pi_2 (x_1^q, x_2^c) &> \pi_2 (x_1^q, x_2^q) \end{aligned}.$$

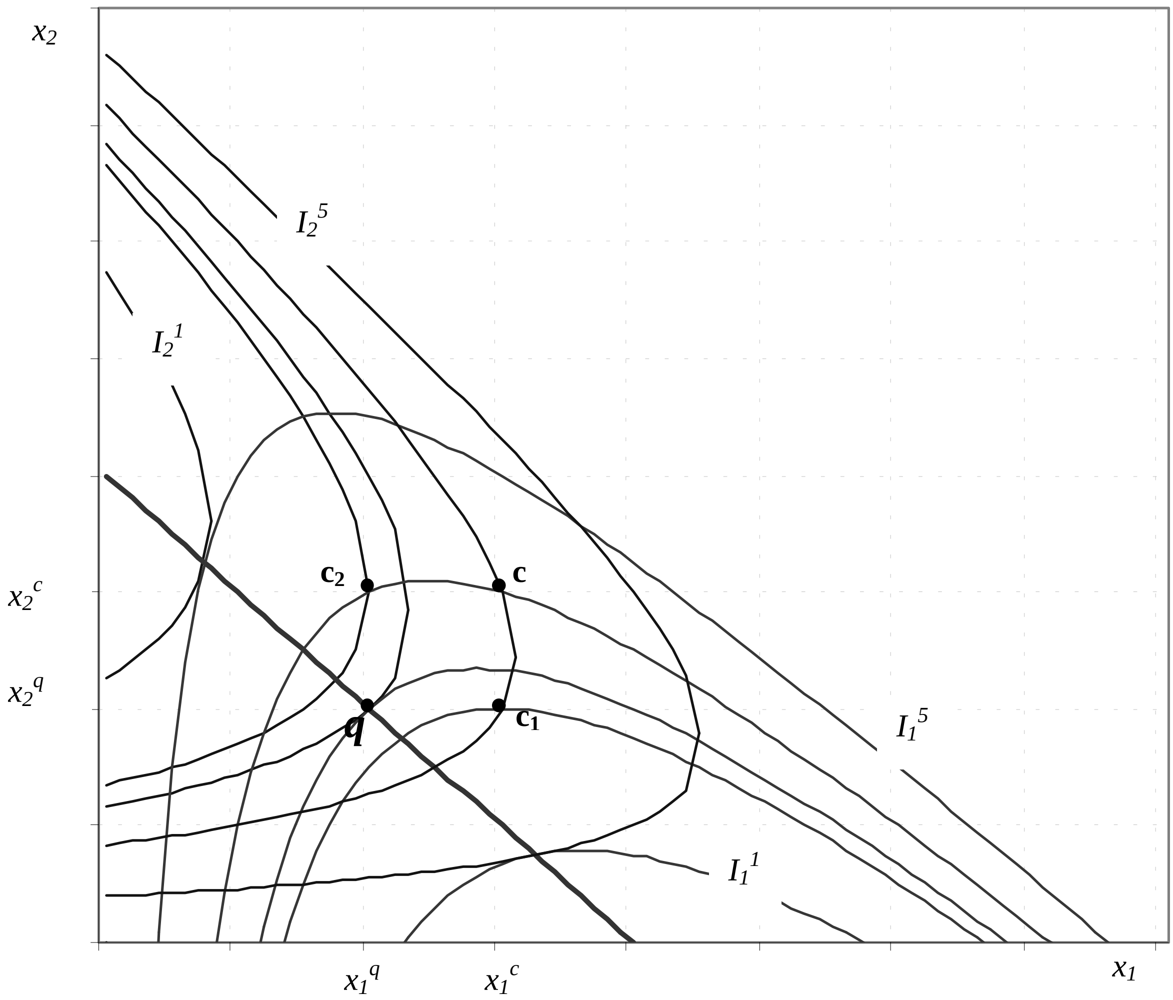


Figure: Cartel Problems

Hence, if both members expect the other member to stick to the agreement in the subsequent period, they will both end up cheating, that is in point c in the figure above, which of course makes them worse off than if both of them had stuck to the agreement (here we end up with a total output of twice the production of a Stackelberg leader). Still, it is also true that if member *i* expects member *j* to cheat in the next period, the best thing to do is for him to cheat as well, as we assume that the following holds:

(5.3) $$\begin{aligned}\pi_1(x_1^c, x_2^c) &> \pi_1(x_1^q, x_2^c) \\ \pi_2(x_1^c, x_2^c) &> \pi_2(x_1^c, x_2^q)\end{aligned}.$$

This situation seems to be like a classical prisoner's dilemma. To observe the quotas are dominated strategies for both players, and we end up in the worst scenario, where the final outcome is Pareto dominated by all other outcomes. It should also be noted that the situation with twice the Stackelberg output cannot be a stable equilibrium, since the producers always has the Cournot equilibrium as an alternative to the collusive equilibrium.

Things do not have to be this bad, however. The prisoner's dilemma could be solved by an enforceable contract, which makes the cartel able to act as a unit and collect and distribute, according to some rule, the monopoly profits. On the other hand, if no such contract is possible, we would expect the cartel to dissolve.

<u>A quota rule</u>

Osborne (1976) suggests one such mechanism that might prevent cheating.[19] Osborne first establishes that an interior production sharing point that maximizes joint profits must have the ray property, that is, the tangency line of the members' isoprofit curves at the production point must also pass through the origin (this can also been seen from the figure above). The intuition is that if this is not the fact, and demand for the cartel's product plummets, at some point one of the members will lose all his production, while the other gets all. The cartel then assigns a quota rule to each member. First let us define member *i*'s market share at the cartel point, *q*:

$$s_i = \frac{x_i^q}{\sum_{j=1}^{n} x_j^q}$$

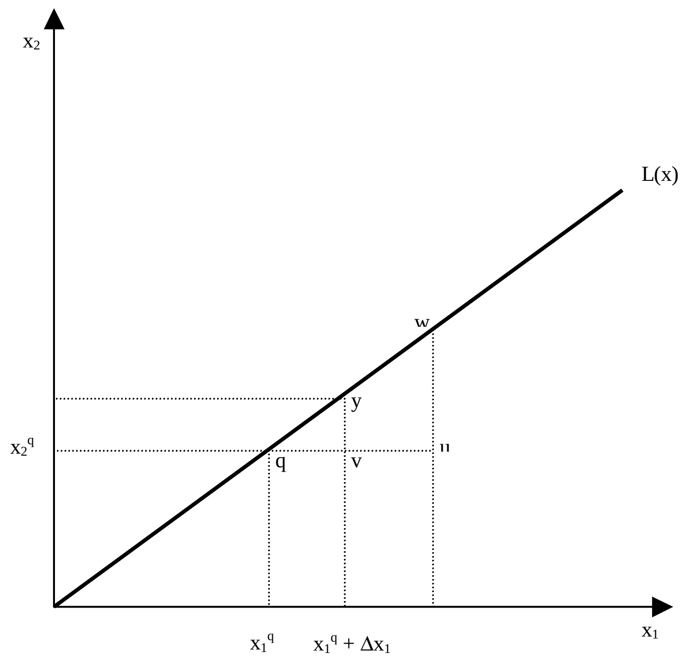
We then set the following quota rule for each member:

$$\max\left\{x_i^q, x_i^q + \frac{s_i}{s_j}\Delta x_j\right\} \quad (5.4)$$

This rule says that if nobody else cheats, you should produce your assigned quota. If, on the other hand, other members are cheating, you should increase your own production by the same amount as the cheaters, adjusted for your original market share relative to the cheater's. This means first of all that the market shares are preserved, but, most importantly, the rule is

[19] This punishment mechanism is included in order to illustrate some of the problems facing OPEC, but also to illustrate what OPEC is not doing.

also a credible threat, as the cheater will have every reason to expect the other members to follow their rules and retaliate accordingly. Of course, this must be because profits are higher by following the rule than not. This can be shown in the following figure.



The figure above shows the situation for a two-member cartel. The cartel point is *q*, and this point has the ray property, which means that the line *L(x)* passing through the tangency point of the isoprofit functions in q passes through the origin. Then member 1 cheats, and we move from *q* to *v*. How should member 2 respond to this? From the figure Cartel Problems, we can see that at least partial retaliation is beneficial for producer 2. Still, if member 1 would not just cheat, but cheat on a massive scale, this result does not necessarily hold, because then a full retaliation would have a powerful negative effect on the product price. In the oil market this argument seems relevant, as demand is quite inelastic in the short run, so the loss in short term profits would be great as oil prices would decline substantially. On the other hand, even though member 2 will not retaliate in full, it will be in his advantage to retaliate to some extent. We can see this clearly from the figure Cartel Problems above. Let as assume that they have allocated quotas such that the two players start out producing in q. The figure then shows that, for linear demand, a player's optimal cheating is to produce the Stackelberg output, while the betrayed party is indifferent between responding in full and not responding at all. If we call the cartel output q^M and the Stackelberg output q^S, the betrayed player will gain from producing anywhere in between q^M and q^S relative to staying at the cartel production level. Hence, retaliation is an effective deterrent.

Osborne's proposal for a quota rule lacks one fundamental aspect of the game between cartel members, namely the repeated interaction over time. Game theory, and especially the theory of repeated games, makes us better able to consider how this affects the incentive constraint facing a cartel member. We will now take a look at how cartel stabilization is affected by repeated interaction, and whether we can find possible credible punishment mechanisms.

A trigger strategy

We will now imagine OPEC and Saudi Arabia using strategies that are not forgiving, for instance a trigger-strategy of the following kind: In the first period every member i produces according to the quota, q_i^C. The aggregate production should then equal to the cartel output. If, ex post, all members have produced according to the quota, q_i^C will be produced also in the next period. However, if some member cheats in any single period, all the other members will retaliate and produce according to the Cournot equilibrium in all the subsequent periods. We assume that the horizon of the game is infinite, that is, the game is repeated an infinite number of times, or, alternatively, there is uncertainty as to when the game ends. By adopting this assumption we avoid that the game collapses into the static equilibrium of the game.[20] Formally, we can represent this situation as:

$$\begin{aligned} &q_{it} = q_i^C \text{ if} \\ &\text{H}_{ti} = (q_{1i} = q_{2i} = \ldots = q_{t-1,i} = q_1^C) \quad \forall i, i = (1,\ldots,n) \\ &q_{it} = q_i^P \text{ otherwise} \end{aligned} \tag{5.5}$$

This system tells us that if the history of the game is such that all players have followed the collusive strategy in all the previous periods (q^C), each player will follow this strategy also in the current period. However, if some member has defected (q^D) in a preceeding period, the strategy tells each member to produce his Cournot output (q^P). The length of each period is the amount of time a cheater will have before it is know that he has defected. Hence, if this

[20] A frequently used example of how finite games collapses is the 'Chain Store Paradox'. This is a situation where an incumbent is faced with numerous potential competitors trying to enter each segment of the incumbent's market sequentially. As it is more costly to deter an entrant than to revert to the duopoly equilibrium in each single case, the incumbent will not deter the last entrant. But then, in the next to last round of the game, the incumbent is faced with the same decision as in the last, since he already knows what he will do in the last round, and so the game collapses to the duopoly equilibrium.

discovery is made quickly, the retaliation will also come quickly, and the period in which a defector can enjoy the increased profits will be short.

If we assume linear (inverse) demand ($p=A-bQ$, $Q=\Sigma q_i$), constant and equal marginal costs for all producers, and a Cournot game with identical players, we know that the profit in the Cournot equilibrium will be (superscript *P* is short for 'punishment')

(5.6) $$\pi_i^P = \frac{(A-c)^2}{(N+1)^2 b}$$, (*N* is the number of players in the market)

and the quantity produced is then

(5.7) $$q_i^P = \frac{(A-c)}{(N+1)b}.$$

As we currently assume a cartel composed of identical players, the cartel profits for each player will be (superscript C here means 'cartel')

(5.8) $$\pi_i^C = \frac{1}{N}\frac{(A-c)^2}{4b}.$$

In order to evaluate whether it is advantageous for a member to cheat, we will also have to know the profits during the defection period. The quantity produced can be found by evaluating the defectors first order condition, given that the other members of the cartels stick to their quoats. We then get the following quantity and profit functions:

(5.9) $$\pi_i^D = \frac{(A-c)^2(N+1)^2}{16N^2 b}, \quad q_i^D = \frac{(A-c)(N+1)}{4N}$$

All the members of the cartel will now be able to evaluate his own incentive constraint, telling him whether cheating will be advantageous or not. The sustainability of the cartel is dependent on that the follwing condition holds:

(5.10) $$\pi_i^C \frac{1}{1-\delta} \geq \pi_i^D + \pi_i^P \frac{\delta}{1-\delta}.$$

This condition tells us that for a cartel to be sustainable, the present value of member i's cartel profits must be at least as large as what he can achieve by defecting for one period and then getting the Cournot profit for all remaining periods. The discount rate, δ, is given by $\delta = e^{-r\gamma}$, where r equals time preference rate, and γ represents the length of each period, that is, how long it takes for the other cartel members to respond to the defection. This enables us to see that a low r, that is, a large degree of patience, and a low γ, that is, rapid punishment, both works in the direction of making cheating less favorable, and hence the cartel more stable. If we solve (5.10) for δ, we get the following expression

(5.11) $$\delta \geq \frac{\pi_i^D - \pi_i^C}{\pi_i^D - \pi_i^P}.$$

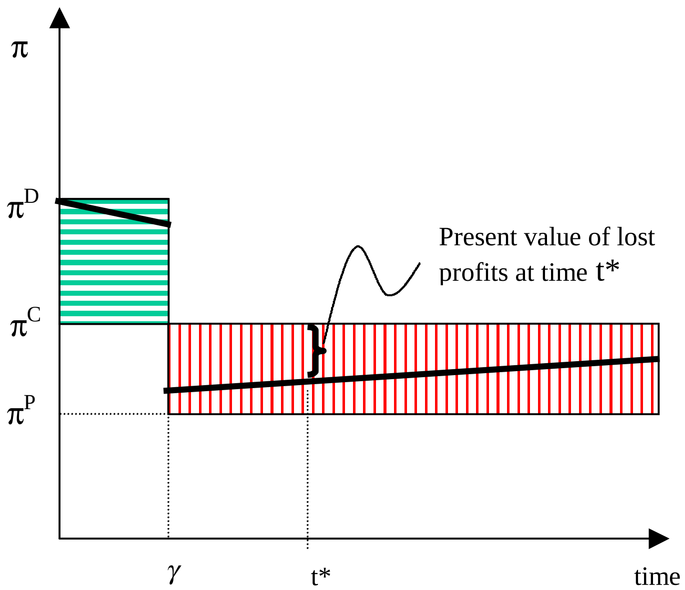
This inequality gives us a numerical expression of the lowest level of the discount rate consistent with cartel bahavior. By inserting for the expressions for profits in the various states, and manipulate the expression slightly, we get that the critical value of δ is given by

(5.12) $$\delta \geq \frac{(N-1)^2}{(N+1)^4 - 16N^2}.$$

This expression provides with us with one clear implication: The number of colluding partners is of critical importance for the sustainability of a cartel. By setting in for $N=2$, and $N=3$, we get that $\delta \geq 9/17$ and $\delta \geq 4/7$, respectively. This shows us that a larger number of producers implies that the discount factor must be higher if collusion is to be sustained.
The logic of collusive behavior and the use of trigger strategies is a bit different in a Cournot situation than it is in a Bertrand setting. In the latter, the gains from cooperation are large, and hence the losses of defection are also relatively large, as the Bertrand static equilibrium gives price equal to marginal cost. In a Cournot setting however, the alternative to the cartel profit is the Cournot profit, which does not have to be much lower than the former. The flip side of the coin is that cheating is relatively less attractive in the Cournot setting, as the gain from

defection is lower. All in all, and all else equal, we expect Cournot cartels to be (marginally) less stable than Bertrand cartels.

The basic dilemma, then, facing all members of a cartel can be illustrated in a simple manner, like it has been done in the figure below. The figure shows how one must trade off the extra profit one may earn by defecting with the profits one loses by being punished. The length of each period influences the desirability of defection by determining the size of the upper and lower shaded areas, and the time preference rate influences the decision by determining the present value of the shaded areas.



This kind of trigger strategy does not look realistic with respect to the oil market. The main argument against it is that it is not a forgiving strategy, and therefore it will be too costly for the ones that enact the punishment to follow this strategy. Alas, the punishment mechanism will be very costly and it will often lack credibility. A strategy where one is not just punished for bad behavior, but also rewarded for good behavior, like the tit-for-tat strategy, looks more likely to serve the members the best.

A tit-for-tat strategy

The Osborne's quota rule and the trigger strategy are just two out of several possible punishment mechanisms. An important element in Osborne's suggestion is that all members of the cartel are responding to the defections in the same manner, and that everybody takes responsibility. This requires of course that all producers have got unused capacity that can be put in use quickly when needed. This does not describe the oil market very well, as most of the producers are producing close to capacity, and only one producer, Saudi Arabia, more or less continuously sits on a substantial amount of free capacity. Therefore it seems natural to think of punishment mechanisms where Saudi Arabia acts as OPEC's 'watchdog'. Another drawback with Osborne's quota rule is that it does not tell us what to do if the defectors return to the agreement. We would then like to develop a new punishment mechanism where these elements are included. We have got more tools available for the analysis of these questions now than Osborne had in 1976. Especially a discovery made in 1984 regarding repeated games is helpful in this respect. That year Axelrod (1984) published a book called *The Evolution of Cooperation*. After having examined numerous different punishment mechanisms for repeated games, Axelrod arrived at the conclusion that a successful strategy cannot only punish deviations from the Pareto efficient strategy combination. It must also be forgiving if the defecting players repent and show renewed commitment to the agreement. One such strategy is the tit-for-tat strategy. Let us apply this strategy to the situation where Saudi Arabia is the sole player punishing defections within OPEC. This strategy then involves punishing other OPEC-members for production in excess of the quotas specified in the agreement, but if the cheating members return to their quota-assigned production, then Saudi Arabia will do the same. Formally, this would look like this:

(5.13) $$Q^{SA} - Q^{SA}_{quota} = \gamma(Q^{OO} - Q^{OO}_{quota}) \ .$$

This rule says just that cheating by other OPEC-countries (OO) will be met by production increases by Saudi Arabia (SA). The parameter γ tells us how aggressive the response from Saudi Arabia will be. For instance, if γ=1 over-production of one barrel by other OPEC-countries will be met by a one-barrel production increase by Saudi Arabia. This strategy looks more realistic than those discussed above, and it is also operational, as Saudi Arabia has the needed spare capacity.[21] A test of the realism of this model will be undertaken below. Before we do that, two more cartel-enforcing strategies will be discussed.

Optimal Punishment

Compared to a grim-trigger strategy, collusion can be more effectively sustained by threatening to administer the strongest credible punishment should deviation occur (Abreu,

[21] In some periods this might not be the case, if, for instance, other disruptions (wars, strikes) have occurred such that the spare capacity is already in use.

1988). This type of punishment is defined as the subgame-perfect Nash equilibrium of the infinitely repeated game that yields the lowest payoff of all such equilibria for the player who deviated (Gibbons, 1992). In most games, switching forever to the stage-game Nash equilibrium is not the strongest credible punishment.

We will now take a look at how, by employing one type of optimal punishment, the carrot-and-stick strategy, collusion can be sustained even with a discount factor, δ, equal to ½ (which is smaller than $\delta \geq 9/17$ for the grim trigger strategy with two players).[22]

Let us assume that there are two players with equal marginal cost, *c*, and that demand takes the form $p(Q)=a-Q$, $Q=q_1+q_2$. Consider now the following carrot-and-stick strategy:

> Produce half the monopoly quantity, $q_m/2$, in the first period. In the t^{th} period, produce $q_m/2$ if both firms produced $q_m/2$ in period $t-1$, produce $q_m/2$ if both firms produced x in period $t-1$, and otherwise produce x.

This strategy involves a one-period punishment phase in which the firm produces x and a potentially infinite collusive phase in which the firm produces $q_m/2$ (equal to $(a-c)/4$). We can see that if one firm deviates, the other will produce x. In the next period, then, the other player will produce x, and this period will then be followed by both producing $q_m/2$. We will now have to check when the collusive phase actually will be infinite, that is, collusion in each period is a subgame-perfect equilibrium.

First, some notation. If both firms produce x, they will each earn the profit $\pi(x)=(a-2x-c)x$. Then, let $V(x)$ denote the present value of getting $\pi(x)$ this period, and $\pi^m/2$ forever after, that is, $V(x)= \pi(x)+ (\delta/(1-\delta))*\pi^m/2$.

If firm 1 will produce $q_m/2$ this period, the quantity that maximizes profits for 2 solves $\max_{q_2} (a-q_2-q_m/2-c)q_2$. The solution here is $q_2=3(a-c)/8$, with the profits (from deviation from collusion) $\pi_d=9(a-c)^2/64$.

If firm 1 will produce x this period, the quantity that maximizes profits for 2 solves $\max_{q_2} (a-q_2-x-c)q_2$. The solution here is $q_2=(a-x-c)/2$, with the profits (from deviation from punishment, *dp*) $\pi_{dp}(x)=(a-x-c)^2/4$.

[22] This exposition follows Gibbons (1992).

If both players follow that strategy above, then the game will consist of two types of subgames, i), the collusive subgames, and, ii), the punishment subgames. The strategy can only be a subgame-perfect Nash equilibrium if it is a Nash equilibrium to obey the strategy in both classes of subgames. Therefore, we have to check the condition for this being the case.

Collusive subgames:

For collusive subgames the outcome in the previous period was either (q_m, q_m) or (x, x). In these subgames, each firm must prefer to receive half the monopoly profits forever rather than receiving π_d this period, and the punishment present value $V(x)$ forever after. Remember that cheating today will be followed by both producing x tomorrow, which again will be followed by the collusion outcome forever after; hence the following condition:

(5.14) $$\frac{1}{1-\delta}\frac{1}{2}\pi_m \geq \pi_d + \delta V(x)$$

By substituting for $V(x)$ we then get that

(5.15) $$\delta\left(\frac{1}{2}\pi_m - \pi(x)\right) \geq \pi_d - \frac{1}{2}\pi_m,$$

which says that the value of having stuck to the collusive outcome and hence having avoided the punishment must be (weakly) larger than current profits from deviating.

Punishment subgames:

The punishment subgames are given by the subgames in which the outcome of the previous period was neither (q_m, q_m) nor (x, x). For these subgames each firm must prefer the punishment present value from the value of deviating from the punishment ($\pi_{dp}(x)$) and begin the punishment again next period. This means that:

(5.16) $$V(x) \geq \pi_{dp}(x) + \delta V(x).$$

If we again substitute for $V(x)$, we get the following condition:

(5.17) $$\delta\left(\frac{1}{2}\pi_m - \pi(x)\right) \geq \pi_{dp} - \pi(x),$$

which says that the net present value of resorting to the collusive outcome in the next period must be (weakly) preferred to the net value of deviating from the punishment today.

Results:

Let us now assume that the punishment phase involves producing *x=(a-c)/2*, such that *p(Q)=a-2x=a-(a-c)=c*, that is, the punishment involves producing until price equals marginal cost.

By substituting for the profit expressions, we then get that (5.15) demands that $\delta \geq 1/8$, while (5.17) demands that $\delta \geq 1/2$. Then, for it to be a subgame-perfect equilibrium to obey the strategy in both subgames, we must have that $\delta \geq 1/2$. Hence we see that by enforcing punishments that are tougher than the static Nash-equilibrium outcome, the opportunity for collusion gets bigger. The attractiveness of deviation is reduced by 'quick and ugly' punishment. This makes the other characteristic of the strategy possible, namely that it is forgiving. A problem with grim-trigger strategies is that in present-value terms, distant punishment is not an effective deterrent relative to the 'carrot-and-stick'-type of punishment. With optimal punishment you can leave the punishment behind you relatively rapidly, and then invite to a return to collusion.

The swingproducer strategy

Saudi Arabia is frequently referred to as a swing producer among the OPEC countries. This means that Saudi Arabia commits to defend a certain price of oil in the market, and adjusts its production unilaterally so that this price target is achieved. Such a strategy seems quite vulnerable to exploitation by the other members of the cartel. We can describe this oil-market situation in a three-stage game: First, the non-OPEC suppliers decide on how much to supply, given OPEC's official price target. Second, OPEC other than Saudi Arabia must decide on its own production, given that other co-members are tempted to exploit the situation, but it must also be taken into consideration that Saudi Arabia must not be pushed to far, because then it will withdraw from the strategy. Third, Saudi Arabia has a choice between, i), deciding upon how much it may supply given all other producers' decision, and, ii), deciding to withdraw from the strategy, and increase its production in order to restore its market position. Formally, we can describe this strategy, as long as it is pursued, in the following manner:

(5.18) $$Q^{SA}(P^A) = Q^W(P^A) - Q^{NO}(P^A) - Q^{OO}(P^A)\,.$$

This says that Saudi Arabia's supply for the price P^A equals world demand, less non-OPEC supply, less other-OPEC supply, all evaluated at the price P^A.

The only way this strategy can be sustained is if Saudi Arabia's profits are at least as high under swing production as it is in the Cournot equilibrium. Whether this will happen, will in turn depend on whether the other OPEC member manages to limit production so that Saudi profits do not fall below the Cournot level.

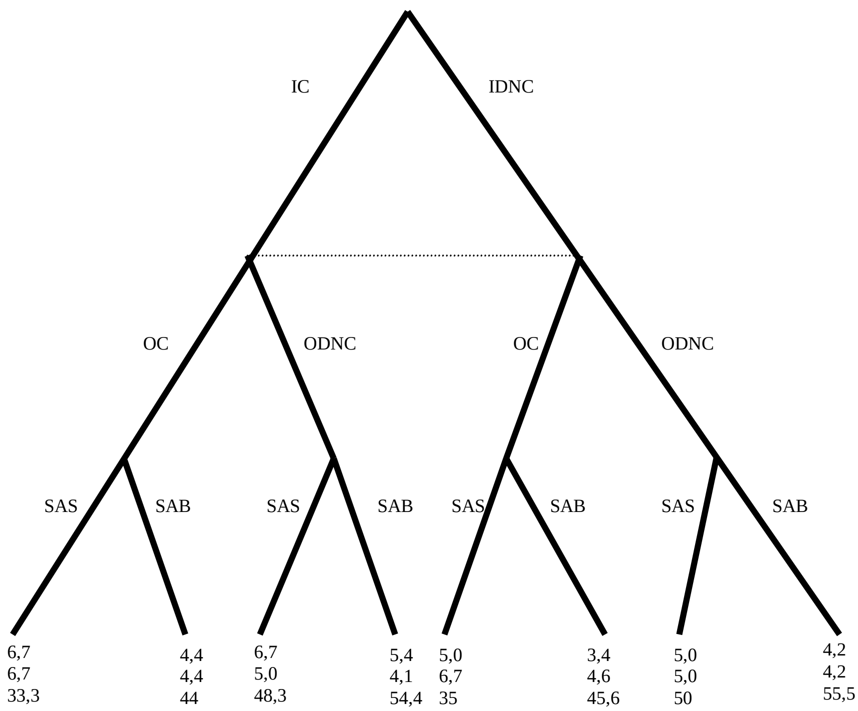
Griffin and Neilson (1994) discuss whether this strategy can be an equilibrium strategy. They make, in my opinion, one very important simplification, as they treat all non-Saudi producers as one single producer, so that there are only two players in the game. The game is based on a trigger strategy. Initially, the producers make a nonbinding agreement specifying each part's output and the price target. Implicit in the agreement is a range of production levels of the non-Saudi producer that will not trigger a defection by Saudi Arabia. However, if production is above the limit of this range, Saudi Arabia will produce according to its Cournot reaction function for the rest of the game. Fudenberg and Maskin (1986) has shown that there can exist equilibria in both finite and infinite games with both producers acting within the boundaries of the swing producer strategy. However, this equilibrium does not necessarily exist for the entire game.

I would like, with a simple example, to provide an argument why I think it is wrong to reduce the number of players in this game to just two. The basic element of the argument is that it is in all single members' interest to increase his own production, because this will not hurt Saudi Arabia enough for the swing-producer agreement to break down. Further, even if the individual member sees that the agreement will break down if the other members increase production, it will never be in his interest not to increase production, as he will lose the short term profits that arises in the period before Saudi Arabia break out. Hence, we end up in a traditional prisoner's dilemma outcome. The situation is described in the sequential game below.[23]

[23] Explanation of abbreviations: IC= I cheat, IDNC= I do not cheat, OC=Others cheat, ODNS=Others do not cheat, SAS=Saudi Arabia stays calm, SAB= Saudi Arabia breaks out.

In the game, the small players first play a static game deciding on output, and then Saudi Arabia decides whether to stick to the agreement or to break out. We assume that Saudi Arabia only breaks out of the agreement if the swingproducer strategy yields lower profits than the static Cournot game.

Swingproducer game



The payoffs below the end nodes indicate the profits of the individual small producer, the profit (per producer) of the other nine small producers, and the profit of Saudi Arabia, respectively. The aggregate cartel profit is 100, and the aggregate Cournot profit is 88,[24] and Saudi Arabia's share is fifty percent in both instances (as long as there is no cheating in the cartel outcome), which means that Saudi Arabia breaks out if the swing producer profit is less than 44. There are eight possible scenarios, based on the strategy profiles where either:

- the individual member increases production (IC), or does not (IDNC)

[24] I have assumed that demand is p=200-Q, that Saudi Arabia gets half the quantity in the cartel and Cournot equilibria, that costs equal zero, and that ten producers share the other half of the quantity in the cartel and Cournot equilibria. The small producers cannot increase production by more than thirty percent, from 5 in the cartel, to 6,7 when cheating. This means that the outcome when everyone is cheating is the same as the Cournot equilibrium outcome of the same game with just two players.

- the other members increase production (OC), or do not (ODNC)
- Saudi Arabia responds aggressively to cheating (SAB), or does not (SAS)

For simplicity, all players except Saudi Arabia have the same individual profit functions, and hence each player is faced with the same payoff matrix. Since Saudi Arabia protects the prices by adjusting production, profits will increase proportionally with production increases as long as Saudi Arabia does not break out of the agreement. Each member decides whether or not to increase production by 30 percent, which by itself will have negligible effects on prices and cartel stability. We see that whenever all the other members increase production in this way, Saudi Arabia will respond by going to ‘war’, because the long term gain from reestablishing its market position is larger than the profits stemming from the defense of the swing producer strategy (that is, we end up in (IC, OC, SAB)). The net present value of the industry profit is reduced substantially when this happens, as the order and discipline that made cartel profits possible has disappeared.

The individual member knows all this, but he also sees that if he does not increase production himself and all the others do it, he will with all realizations of the other players’ actions be worse off than if he had increased production. In addition, if the other members do not increase production, it will be advantageous for the individual member to do so, as Saudi Arabia will then not respond aggressively. Hence, the dominating strategy for the individual member is to increase production. As every member is faced with the same situation, all members will increase production, erode Saudi Arabia’s market position, and eventually see Saudi Arabia respond aggressively by increasing production substantially in order to regain its foothold. This means that the subgame perfect equilibrium will be (IC, OC, SAB) with profits (4.4, 4.4, 44), which is lower than the cartel profit of (5, 5, 50) The profits of OPEC as a whole are reduced substantially in this prisoner’s dilemma outcome.

One might object to this by saying that the other members might increase production by less than the indicated thirty percent. In this case, Saudi Arabia will not be pushed to defection. However, this just delays my argument one period. After stepwise increases in production, they will at some point face the situation presented in the matrix above, and the agreement will collapse. Collapse is also what the agreement eventually did do in 1985, perhaps not so much due to production increases by the other OPEC countries, but primarily because they let Saudi Arabia carry the entire burden of reduced OPEC market shares as non-OPEC production surged.

Collusion With Exhaustible Resources

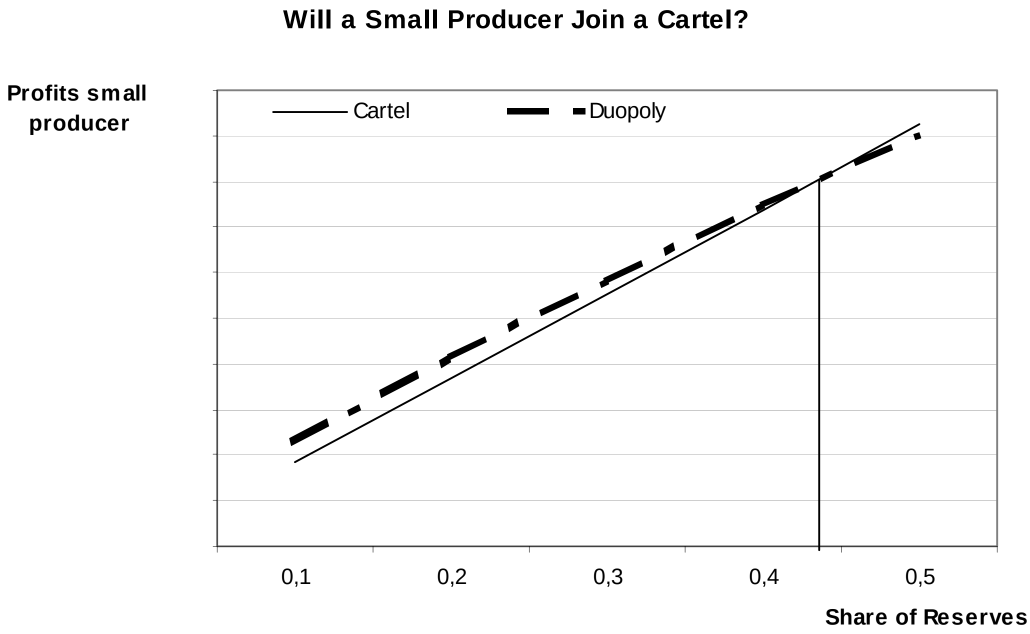
How does exhaustibility affect the possibility of collusion? Markets for exhaustible resources are different from the infinitely repeated games from above in at least two respects. First, the horizon cannot be infinite, as the resource stock is finite. From above we have seen that repeated games with a finite horizon is prone to collapse into the equilibrium of the static game (the Chain Store Paradox). Could it be the case that collusion may not be supportable in markets for exhaustible resources? Second, the nature of the game is not purely repetitive, it should rather be described as dynamic. From each period to the next, the state variable, the remaining resource stock, is changing, and this makes the nature of the game dynamic.

What one might suspect would pose a problem for supporting collusion with exhaustible resources is that it may be impossible to enforce the punishment. Just think of two producers who are colluding, one with larger reserves than the other such that the latter will exhaust his resource stock prior to the former. In this setting, the incentives for the small producer to cheat in the periods prior to exhaustion are clear. Not only will he benefit from increased current production, but he will also avoid the punishment as he leaves the market in the period following the defection. However, we must remember that a small producer in the second to last period is confronted with a problem. If he cheats today, he must produce less tomorrow. This trade-off is not present in standard finite repeated games, and we would suspect that this would be a factor that compensates for the negative impact the finite horizon poses.

The literature on collusion with exhaustible resources is surprisingly small. In fact, I have only found one paper (Hogan and Holland, 2002) formally trying to study the possibility of colluding in these kinds of markets. The main contribution of this paper is that Hogan and Holland develop a model that is not relying upon pre-commitment (that is, they are considering a closed loop rather than an open loop). In this paper, the authors show that for states with remaining stocks close to zero, there will in fact be a unique subgame-perfect equilibrium that does not support collusion. However, this unique non-colluding equilibrium for states with small stocks is not the only subgame-perfect equilibrium for the game as a whole. This is shown by constructing a simple model, with one producer with a limited stock (the small producer), and one producer with an unlimited stock (the large producer). Then, the authors show that when the stock of the small producer is close to zero, the equilibrium is unique and non-colluding. However, at some critical level of remaining stock, S^*, the small producer will be indifferent between colluding and deviating. This means that for periods where $S \geq S^*$, there will be multiple equilibria, and this means that some degree of collusion can be supported by a subgame-perfect equilibrium in these periods. However, in this model, the monopoly payoff cannot be supported by a subgame-perfect equilibrium. The folk theorem says that, with reasonable discount factors, there are many possible equilibria that can support the monopoly payoff through different ways of dividing the monopoly profits among the producers. This does not seem possible with exhaustible resources. However, for the producers, the results might not be as negative as they first look. As the initial stock of the small producer increases, the authors find that the (average) payoff that can be supported for $S \geq S^*$ approaches the monopoly profits. This result is intuitive, as increases in initial stock levels leaves date when S^* is reached very distant, and the game will approach a game with an infinite number of periods. Hence, the conclusion is that exhaustibility hinders a cartel's enforcement of perfect collusive agreements but will not prevent the cartel from supporting prices above the competitive level.

From the distribution of crude oil reserves presented in the first part of this thesis, one could clearly see that OPEC is composed of all the largest reserve holders, and most of the relatively large owners of crude oil reserves. This can readily be explained by using the same framework that was used to develop the production paths under duopoly. If we again look at a situation with two producers, and let the share of total reserves for the smallest resource owner vary between ten and fifty percent, we can deduce from the profits resulting from the production paths under duopoly and a perfect cartel (producing monopoly output) when the small producer will benefit from joining a cartel. The framework is exactly the same as that of the monopoly and Cournot models of part 2. Let us also assume that the cartel follows what we may call a proportional sharing rule, that is, each producer gets a share of total production equal to his share of total reserves.

What we find by computing the total discounted profits for market shares varying from 0,1 to 0,5 for the two different regimes (Cournot and monopoly) is that the small producer will only benefit from joining the cartel if he is not too much smaller than the large producer. This can be seen from the graph below.



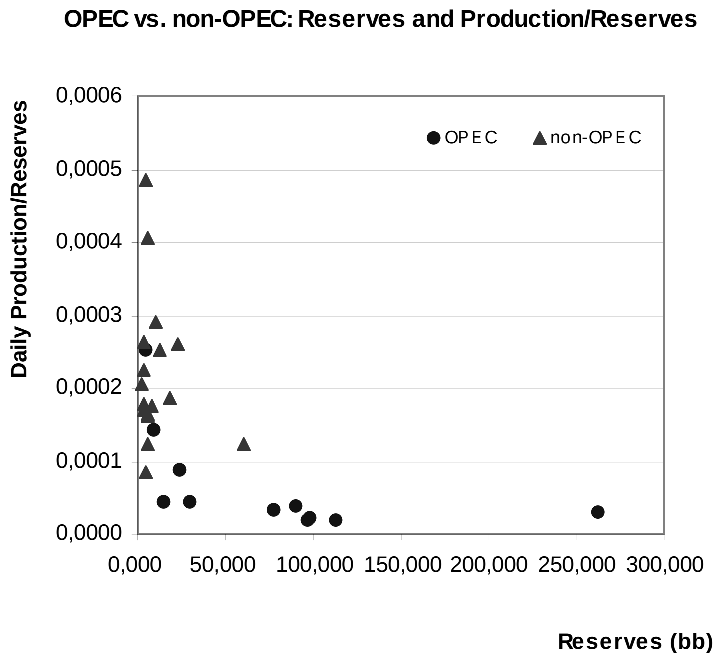
In our scenario, only when the small producers possesses between 40 and 45 percent of initial reserves, will he benefit from joining the cartel. This result generalizes to situations with more than two producers. The intuition is that in a duopoly small producer will exhaust his resources many periods prior to a the large producer. Even though industry profits may be increased by the creation of a cartel, the small producer will now see a bulk of his profits coming from periods where he would not have been producing were the cartel not formed. As the time interval between the Cournot time of exhaustion for the small producer and the monopoly time of exhaustion increases as the reserve share of the small producer declines, this trade off between higher prices but increased discounting works in the disfavor of the cartel. Further, in a duopoly the market share of the small producer is initially bigger than his share of total reserves. Hence, in a cartel the small producer will lose out when only given his proportional share of production and profits. All this means that given a proportional sharing rule, we should expect only quite homogeneous producers to form a cartel in markets with exhaustible resources.

We might also note that as industry profits are lower under Cournot competition with homogeneous producers (equal reserves) than with heterogeneous producers, the gain from cartelization is bigger when producers are homogeneous. Luckily, this is exactly when we would expect collusion to be feasible.

A proportional sharing rule is of course not the only feasible sharing rule. For instance, one might give a larger than proportional share to small producers in order to improve their incentive constraint (we might call this a concave sharing rule). With a concave sharing rule collusion would be supportable for a larger range of distributions of market shares. This would be a sensible sharing rule since the distribution of profits in the non-cooperative game is concave in the share of reserves. Still, even with a concave sharing rule, large heterogeneity among the producers' reserve levels may still make it impossible to support collusion.

Who Wants to Be an OPEC Member?

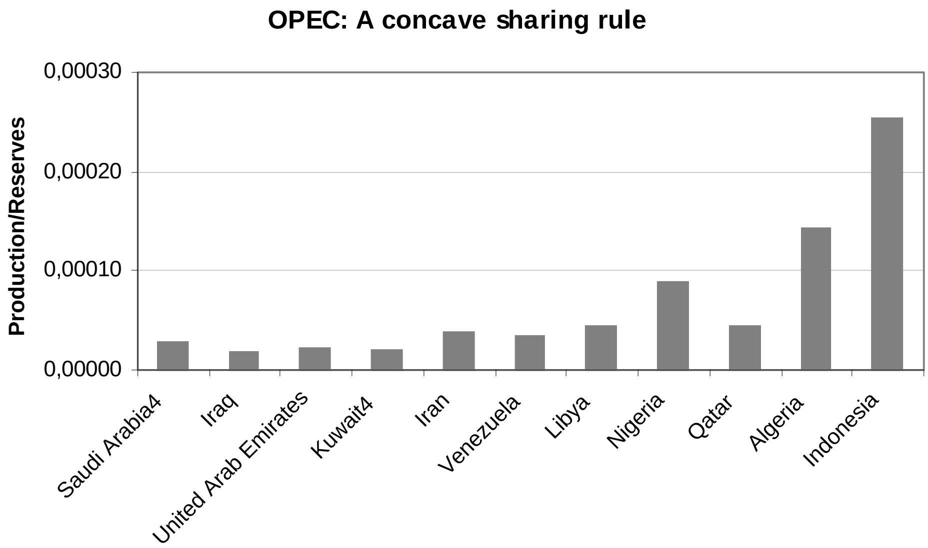
Looking at data on how world oil production is shared between OPEC and non-OPEC countries one gets quite puzzled. Take for instance a look at the following plot with proven reserves (in billion barrels) on the *x*-axis, and production as a share of remaining reserves on



Data: EIA (2001 production), Oil and Gas Journal (2001 reserve levels)

the *y*-axis (daily production as a share of reserves). The triangles are representing non-OPEC countries, while the circles are representing OPEC countries. From this plot we can see that eight out of eleven OPEC members have lower production relative to reserves than ALL the non-OPEC members. The OPEC members are voluntarily giving up current profits, and giving non-members a free ride. This is bound to happen when a market is only partially cartelized, but we are still faced with a question: Why do the small reserve holders within OPEC continue their membership? We have already noted that Equador and Gabon left the organization without any sort of punishment at all. Why are not other members following Equador and Gabon's example?

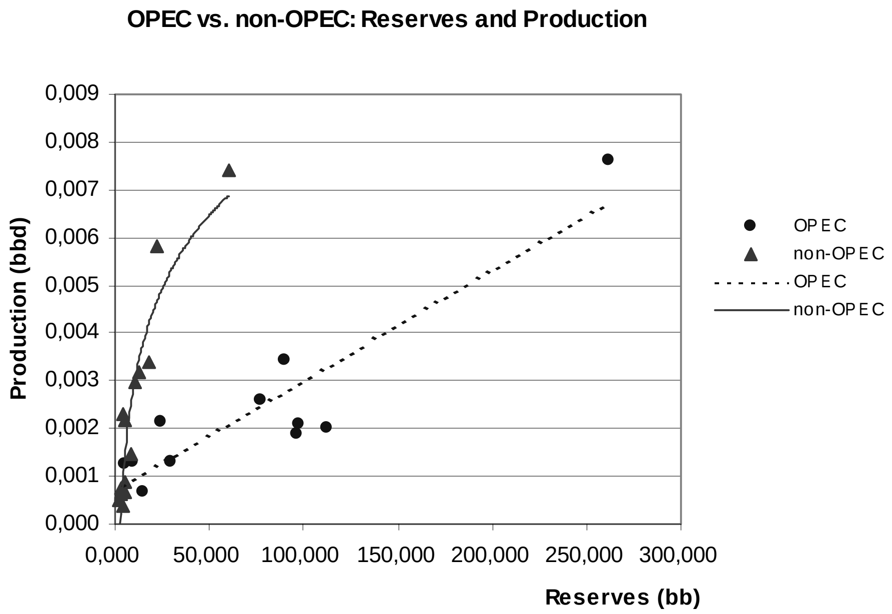
As predicted above, the plot also tells us that OPEC in fact is an organization that consists of all the biggest reserve holders. Still, there are also some relatively small members (in terms of reserves). Above we could also see that it is difficult to support collusion when member heterogeneity increases. This plot tells us why. Small non-OPEC producers are currently producing far more as a share of reserves than large OPEC members. This means that small reserve holders will find a proportional sharing rule very unattractive. OPEC has understood this, and the concavity of the sharing rule is apparent from the following figure, where the OPEC members are ordered from the largest to the smallest.



Data: EIA, Oil and Gas Journal

Still, we have a puzzle to explain: Why do small OPEC members seem to be content with a deal that apparently is worse than the one they would be able to get outside the cartel?

This puzzle can be illustrated in another way. The plot below again has the reserve level (in billion barrels) on the *x*-axis, but now just the production level on the *y*-axis (in billion barrels per day). I have also added trend lines for the two groups. We can see that increases in reserve levels for non-OPEC members implies a large expansion in output, while the relationship is far weaker with respect to the OPEC members. We can see that the production sharing rule in OPEC, even though it is concave, is not fully compensating the small members for being



Data: EIA

members of the cartel. Again: Why continue with a membership when there apparently are large gains from terminating it?

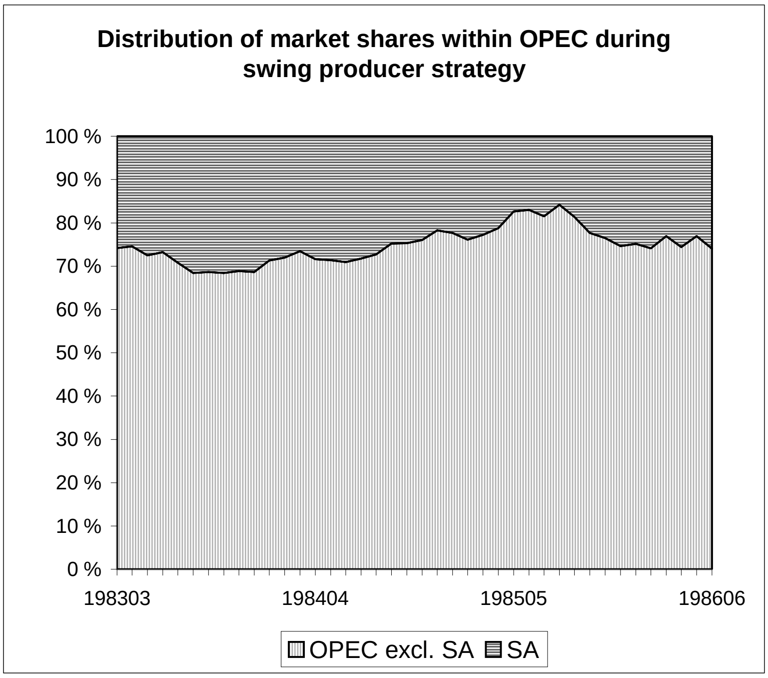
Empirical Results: How does OPEC behave?

The theory of repeated static games has shown us that collusive outcomes can be supported in infinitely repeated games if the players have sufficiently high discount factors (low discount rates). In markets for nonrenewable resources, however, we cannot rely on simple repeated static games. As the remaining stock of reserves changes from period to period, the game must be dynamic. Further, the available punishment mechanisms are limited by the amount of remaining reserves, and especially for owners with small amounts of remaining reserves we would expect that it is difficult to deter cheating. On the other hand, above we have seen that most members of OPEC do have large amounts of remaining reserves, even though there are exceptions. These exceptions (like Qatar and Algeria), however, are too small to attract considerable importance. Since there are no certain termination dates for the repeated interaction among the major producers within OPEC, we might suspect that the possibility of supporting collusion is real.

We have now looked at some of the underlying fundamentals of cartels. We have established that credible punishment strategies are important for the sustainability of cartels, and various possible punishment mechanisms have been described. It is now the time to look at whether OPEC really has behaved according to any of these strategies, and to use this information to evaluate OPEC as a cartel. What is noteworthy is perhaps not that OPEC never has behaved according to Osborne's rule or the trigger strategy described above. Rather, it is a bit surprising that OPEC never even has had a rule stating how defections from quotas would be punished. We will now see that Saudi Arabia has had to take on much of the responsibility single-handedly.

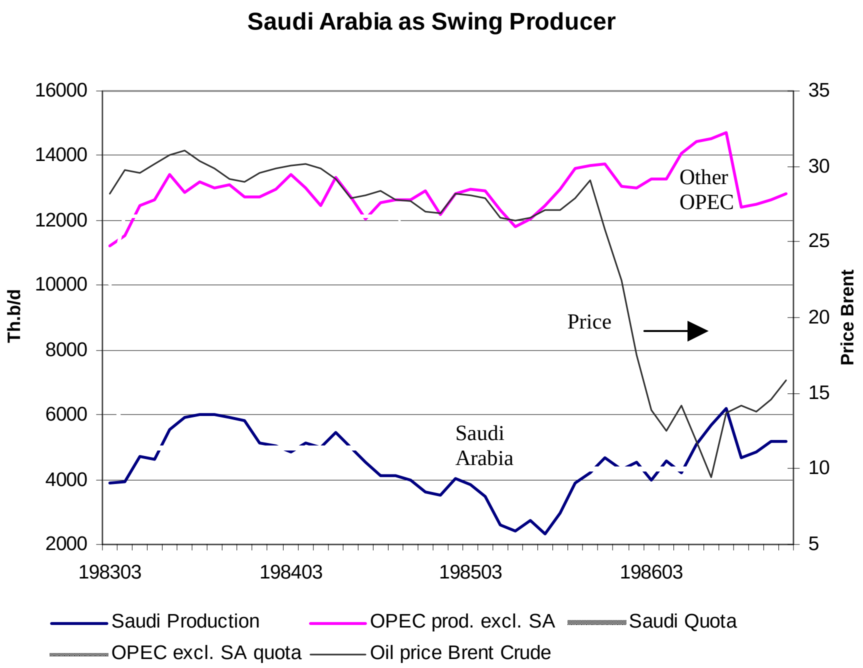
The oil price shocks of the 1970s are usually attributed to the effects of an effective cartel. Still, OPEC did not have a quota system until 1982, 10 years after its claimed success as a cartel, and 23 years after its establishment. Even though one might attribute the price increases during the 1970s to collusion, I will start my analysis in 1982 with the introduction of output quotas. As there were no explicit production agreements before 1982, we may start our evaluation of the OPEC strategies that year. The high oil prices during the 1970s had caused massive capacity investments in non-OPEC countries, and combined with depressed demand resulting especially from a recession in the US, this forced OPEC to rethink its strategy. Increased fringe production and a long-run demand elasticity higher than expected would make the situation harder than it had been during the 1970s.

From 1982 to 1985 OPEC behavior was governed by a non-binding agreement that gave Saudi Arabia the role as swing producer, whereas the other members were assigned production quotas. What happened in this period seems to fit the hypothesis presented above quite well. After an initial slump in the OPEC-ex-SA market share, it rose gradually, up to a peak of 84 percent of total OPEC production in September 1985. The Saudi share of OPEC production was then down from 27,2 percent in March 1983 to 16 percent in September 1985 (see the figure below).



Data: EIA

The market shares of the other OPEC countries increased due to one basic reason: cheating on quotas. In the figure below, you can see how Saudi Arabia continuously was squeezed as the others cheated on their quotas. In the middle of 1985, Saudi production was barely fifty



Data: EIA, Ecowin

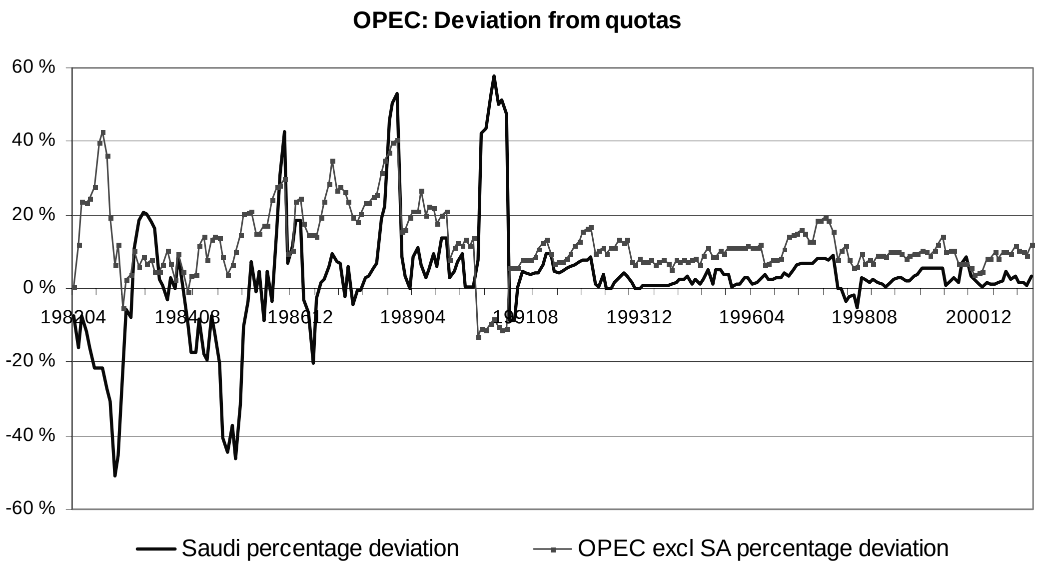
percent of its official quota. Still, Saudi Arabia kept sticking to the agreement, at least for a while. The nominal price of Brent Crude was quite stable in this period, moving within a $27-

$32 price band (real oil prices, however, was declining substantially during this period). On the other hand, we can see that the price of oil over time gradually moved lower. Then, in the fall of 1985, Saudi Arabia reneged on the agreement, increased production by 250 percent within one year, and the era of Saudi Arabia as a swing producer had ended.

It is worth noting the dramatic effect on oil prices after the collapse of the agreement. The prices of Saudi Light Crude fell from $28 to $10 in the period between November 1985 and July 1986, after Saudi Arabia alone had increased production by close to 4 million barrels per day. As other OPEC members also had increased production by more than 2 mbd, and non-OPEC production were 1 mbd above its recent through, the total increase in world oil production was around 7 mbd. This amounted to a 14 percent increase in world production, while the prices fell by around 60 percent. This example illustrate an important facet of the oil market: The short term elasticity of demand is very low, in absolute terms, that is, demand is very price inelastic in the short run. A typical estimate is that the short run elasticity of demand is around –0,3, and this contributes to the volatility in prices. This makes it possible to punish deviations even with a relatively small spare capacity.

Since the oil price collapse in 1986 there has been no official punishment mechanism within OPEC. However, as the organization is still alive, and the OPEC members regularly set new production quotas in order to stabilize the oil market, there must be some mechanism or another that prevent the member countries from trying to increase production at the cost of someone else. One interesting candidate was presented in the last section, namely the tit-for-tat strategy. As a start, we can try to check whether this strategy has been used by studying graphs of production and quota levels for Saudi Arabia and the other OPEC countries. If the tit-for-tat strategy is used, we would expect to see Saudi production increasing and decreasing with the cheating and repentance of the other members, perhaps with a response lag.

The figure below tells us several interesting things. First of all, over time, the volatility in production relative to the production quotas has declined considerably over the past decade. Even though there has been an average overproduction of about 2 mbd during the 1990s and early the 2000s, the volatility of this cheating is far lower than it was during the 1980s. This taken alone seems to indicate one of two things: First, the cartel has become more successful in controlling production and creating stability among the member countries. Or, second, it could be that the cartel is not working at all, and that the member countries, except Saudi Arabia and possibly a couple of other countries, are given irrelevant quotas close to capacity, and that this is the factor explaining the stability. However, one should not forget that things were different during the 1990s than they were during the 1980s and the early 1990s, when especially the Iran-Iraq war and the Gulf crisis were creating instability in the industry. Such shocks were largely absent during the 1990s, except for a negative demand shock stemming from the Asian financial crisis, and the aggressive production increases undertaken by Venezuela during the mid-to-late 1990s.



Data: EIA, CERA

The second thing the figure tells us is that Saudi Arabia seems to have responded to cheating by other OPEC members in the manner prescribed by the tit-for-tat strategy. Increased cheating by other members is followed by increased cheating by Saudi Arabia, and the same applies to situations where the sign of the change in cheating is negative. This gives us some preliminary support for the tit-for-tat strategy, but in order to feel more certain, we will have to perform a statistical test. A candidate test is the following:

(5.19) $$Q^{SA} - Q^{SA}_{quota} = \gamma(Q^{OO} - Q^{OO}_{quota})$$

This is the same as equation (5.13). However, we would expect Saudi Arabia to react more forcefully to large deviations from the quotas than from small deviations, and Griffin and Neilson (1994) suggests that the following specification might be more appropriate:

(5.20) $$Q^{SA} - Q^{SA}_{quota} = \gamma_0 + \gamma_1(Q^{OO} - Q^{OO}_{quota}) + \gamma_2(Q^{OO} - Q^{OO}_{quota})^2$$

This allows non-linear responses from Saudi Arabia. If we expect large deviations to be punished more harshly than smaller ones, we would expect γ_2 to be positive, and γ_1 to be negative. I have checked these hypotheses by running some simple OLS regressions. The results are provided in the table below.

Table: Estimation of tit-for-tat behavior

Test: $Q^{SA} - Q^{SA}_{quota} = \gamma_0 + \gamma_1 (Q^{OO} - Q^{OO}_{quota}) + \gamma_2 (Q^{OO} - Q^{OO}_{quota})^2$

Sample period	Constant	$(Q^{OO} - Q^{OO}_{quota})$	$(Q^{OO} - Q^{OO}_{quota})^2$	R sq., adj.	SE
1) 1983:03/1985:08	-496,9 (-1,34)	0,663 (0,87)	-0,00049 (-1,00)	0,0357	1028,3
2) 1985:10/1990:03	835,8 (1,84)	-0,875 (-2,71)	0,000238 (4.32)	0,592	392,2
3) 1990:04/2002:01	603,9 (10,84)	-0,572 (-16,52)	0,000198 (11,87)	0,672	330,7

Data: Production: EIA, Quotas: CERA

The results provide quite strong support for the tit-for-tat hypothesis from late 1985 and onwards. The tests of the periods running from 1985:10-1990:03 and 1990:04-2002.01 give use the expected sign for γ_1 and γ_2, whereas the test of the period where Saudi Arabia acted as a swing producer gives us the opposite signs. The adjusted R-squared for the two tests after 1985 are as high as 0,592 and 0,672, and this shows us that the model can explain a high degree of the variation in Saudi cheating. Further, all the coefficients of tests 2 and 3 are significant, while those of test 1 are not. We also see that, though not significantly, the coefficient in column 4 is negative for test 1. This means that during the swing producer period Saudi Arabia did not punish cheaters, but instead responded by reducing production.

These results improves our confidence in the hypothesis that Saudi Arabia plays an adjusted tit-for-tat strategy, where it leaves small quota deviations alone, but responds more aggressively when the deviations are large. However, cheating within OPEC is a constant phenomenon, and one might then wonder why the organization tries to share production through quotas in the first place. There is something wrong with the cartel hypothesis when the members never respect their quotas.

OPEC vs. Other Explicit Cartels

Alhajji and Huettner (2000) have studied a number of international quantity-fixing commodity cartels[25]. It should be noted that they are considering markets with explicit collusion. Even though the conditions facilitating explicit conditions are lacking, we cannot rule out that producers are cooperating tacitly. Alhajji and Huettner first specify the general characteristics of cartels in general, and then discuss whether the various international commodity cartels share these characteristics. The main characteristics of a cartel can be summarized in the following way:

- i) A cartel will divide the market by having a quota system
- ii) This quota system will be monitored to identify violations and violators
- iii) Cheaters and quota violators will be punished by creating and implementing a punishment mechanism.
- iv) It will be ensured that the cartel, not the member countries, has authority by enforcing cartel authority
- v) The cartel members should have cash and buffer stocks to support prices and to prevent high prices that lead to substitution and erosion of the market for the cartel
- vi) The cartel should have a large market share to be able to control the market.

How then does OPEC fit these characteristics?

First, OPEC has had a quota system since 1982. Note however that this was implemented 21 years after the birth of the organization. All the other cartels implemented a quota system from day one.

Second, cartels should have a monitoring system to oversee production and shipments. Usually this monitoring takes the form of requiring certificates of origin for all shipments. OPEC has established a Ministerial Monitoring Committee, but this is only used for short periods of time when violations are suspected. Ultimately the organization will get to know if someone cheats as trade data are collected, but this does seemingly not prevent cheating.

Third, as discussed, OPEC has no official punishment mechanism for the organization as a whole. Saudi Arabia may act as the organization’s watchdog, but OPEC has never agreed

[25] They looked at the cartels in the markets for the following products: Diamonds, coffee, bauxite, tin, rubber and oil.

upon a common punishment mechanism, which all other cartels in the sample have done. In the tin cartel, for instance, violations lead to lower quotas in the future, while in the diamond market De Beers penalizes quota violators by using the buffer stock to force down prices.

Fourth, with respect to cartel authority, OPEC has not gone very far. The member governments ultimately retain production autonomy. As seen from the production data above, cheating is a constant phenomenon, and this shows that even though the organization tries to enforce its authority upon the member countries, this effort is quite ineffective.

Fifth, OPEC has never used or even had side payments and buffer stocks in order to force oil prices into the preferred territory. Side payments can be given to a producer to keep his production low in order to sustain the price level, while buffer stocks can be used to increase the quantity of the product in the market in order to lower prices from an undesirably high level. Alternatively, when prices are below the price floor, the cartel can use the assets from a cartel fund to buy the commodity in the open market in order to push up prices. Looking at price data it does not seem like OPEC has been very successful in preventing that prices fall below some minimum price level. In other cartels, on the other hand, the floor price has effectively been sustained by using these mechanisms. To some extent, excess capacity in the major member countries works as a buffer stock, as it frequently has been used to influence prices. We have already seen that with inelastic demand even relatively small amounts of spare capacity can force prices down and work as an effective deterrent. However, just three OPEC member account for the bulk of the excess capacity (Saudi Arabia, Kuwait, and the UAE), and these countries' excess capacity is based on the interest of the individual member, not on the interest of OPEC as a whole. Saudi Arabia, for instance, keeps its excess capacity in order to prevent huge shocks to the market and perhaps also in order to gain some goodwill from the US. This, of course is costly for Saudi Arabia, and it does so not due to benevolence, but to serve its self-interest.

Sixth and finally, compared to other international cartels, OPEC does not have a large market share. It controls less than 40 percent of the market, while other cartels controls about 80 percent of the market and never below 70 percent.

Another problem for OPEC is that quota allocations are based on reserve and capacity data. These are inherently uncertain, and there are great incentives for all parties to inflate their numbers. This makes it hard to find a real substance as the basis for the quota allocation. Further, the fact that previous production is not the basis for quotas shows that deviations will not necessarily be punished. Quite the contrary, producers who expand their capacity and cheat on their quotas might actually be rewarded for this as quota allocations are partially based on production capacity.

One basic fact characterizing profit maximizing monopolists and cartels is that they operate on the elastic part of its demand curve. If this is not the case, the cartel either does not maximize profits or it cannot be a dominant producer. If the elasticity is less than one (in absolute terms), the cartel will not be maximizing revenues (which happen when the elasticity equals one). Whether OPEC follows this basic principle of profit maximization is possible to test with a simple model. If we assume that OPEC is a cartel, its demand will be the residual demand, denoted

(5.21) $$Q_o = Q_w - Q_n ,$$

where Q_o is the demand for OPEC oil, Q_w is the world demand, and Q_n is the non-OPEC supply (assumed to be a competitive fringe). Changes in OPEC prices will lead to changes both in world demand and non-OPEC supply, or

(5.22) $$\frac{\partial Q_o}{\partial P_o} = \frac{\partial Q_w}{\partial P_o} - \frac{\partial Q_n}{\partial P_o} .$$

If we then multiply both sides by P_o/Q_o and the first and second term on the left hand side by Q_w/Q_w and Q_n/Q_n respectively, we get

(5.23) $$\frac{P_o}{Q_o}\frac{\partial Q_o}{\partial P_o} = \frac{P_o}{Q_o}\frac{\partial Q_w}{\partial P_o}\frac{Q_w}{Q_w} - \frac{P_o}{Q_o}\frac{\partial Q_n}{\partial P_o}\frac{Q_n}{Q_n} .$$

This can be rearranged

(5.24) $$E_o = \frac{Q_w}{Q_o}\beta - \frac{Q_n}{Q_o}\delta_n ,$$

where E_o is the elasticity of demand for OPEC oil, β is the world elasticity of demand, and δ_n is the non-OPEC elasticity of supply. Alhajji and Huettner (2000b) have estimated this relationship for OPEC and Saudi Arabia with quarterly data from 1973 to 1994, using prior estimates on the world elasticity of demand and the non-OPEC elasticity of supply. The results show that the elasticity of demand for OPEC is less than one for the entire period, while it is higher than one for Saudi Arabia for the entire period. For OPEC the elasticity is relatively stable over time, fluctuating between 0,2 and 0,4 (in absolute terms). For Saudi Arabia, on the other hand, the elasticity fluctuates considerably, from 1,5 to 6.[26] These results does not seem compatible with the notion of OPEC acting as a profit maximizing cartel. They indicate however that Saudi Arabia operates on the elastic part of the demand curve, and that its role in the market, not surprisingly, is different from that of other producers. However, we must remember that these tests are performed with short-term data (quarterly). As the long-term elasticity of demand has proven higher then the short-term elasticity, one should test whether the long-term elasticity of demand for OPEC oil is larger than one (in absolute terms). I have not checked this myself, but my feeling is that it probably is. The extent of substitution away from oil after the 1970s price increases surprised the OPEC members, and this has lead to increased awareness of the relatively elastic long-run demand.

Statistical Tests

As an extension of these results, I will present the main findings of Alhajji and Huettner (2000b), where the cartel hypothesis is tested more thoroughly. Their primary objective is to test whether the world crude oil market is characterized by the existence of a dominant producer, and whether this dominant producer is OPEC, the OPEC core[27], or Saudi Arabia. They model variants of the competitive and Cournot models were also tested, but all of them were rejected. Their dominant firm model considers OPEC as a unified cartel and assumes that the members of the dominant firm have unified goals and collectively set the price of oil. Non-OPEC oil producers are the competitive fringe, and the demand for the dominant firm is the residual demand. Prices will be set where the dominant firm’s marginal revenue equals marginal costs, and the fringe will then supply up to the point where price equals marginal cost. The model includes specifications for demand and non-OPEC supply. Estimates for cost of extraction and user cost (measured by the price of replacing the oil stock) are included, and the authors also includes above normal defense and security expenditures into the cost of oil

[26] These results might seem surprising at first, but the explanation is intuitive. A high elasticity for Saudi oil supply is combined with a low elasticity for aggregate OPEC supply because the correlation coefficient between Saudi production and rest-of-OPEC production is significantly negative.

[27] The OPEC core includes Saudi Arabia, Kuwait, UAE, and Qatar

production. Further, a dummy variable for the US price regulation from 1973 to 1980 is included. The US imposed price controls on domestically produced oil in an attempt to lessen the impact of the 1973-74 price increase. The result of the price controls was that U.S. consumers of crude oil paid almost 50 percent more for imports than domestic production, and as production declined in the US, demand shifted to other producers.

The general predictions of the model are in accordance with theory. World demand responds positively to GDP growth and negatively to price increases. Further, non-OPEC producers exhibits competitive behavior because supply increases in prices, and declines in costs. We will discuss this more in detail below, but we might now note that even though this response is in accordance with competitive behavior, it is also in accordance with non-competitive behavior. The results also show that the effects of US price regulation were on the one hand to increase world demand, and on the second hand to decrease non-OPEC supply as production in the US fell. The main benefactor of the US price controls was Saudi Arabia, the only country that had substantial amounts of spare capacity making it able to meet the increased demand.

The results of the statistical tests do not support the model of OPEC as a dominant producer. The reason is the same as above, namely that OPEC does not meet the profit maximization condition which says that the elasticity of demand for OPEC oil must be larger than 1 (in absolute terms). According to the authors, OPEC did not meet this condition for one single period between 1973 and 1994. But again we might reply that tests based on short term elasticities might be the wrong method when long-run demand is far more elastic.

Neither the behavior of OPEC core is fully consistent with the dominant firm model, even though the profit maximization condition is met for the period as a whole. For the period 1971 to 1980 Saudi production was strongly negatively correlated with that of Kuwait and Qatar, and insignificantly positively correlated when UAE is included[28]. In this group, however, the elastic demand for Saudi oil, which composes 70% of the demand for the core's oil, to some extent compensates for the inelastic demand for the others, and hence the results are closer to being consistent with the dominant firm model.

As expected from the above results, the dominant firm model cannot be rejected when Saudi Arabia is considered the dominant firm. Saudi Arabia has operated on the elastic part of its demand curve for the entire period, and this seems to suggest that it is the dominant producer in the world crude oil market.

[28] This seems to indicate that Saudi Arabia was a de facto swing producer also during the 1970s.

Economists have for several decades now been troubled by their lack of understanding for the oil price increases during the 1970s. Of course, political factors influencing exporting countries' production (like embargoes, revolutions, and wars) were important, but they do not seem to provide us with the complete picture. Alhajji and Huettner provide us with one quite intriguing explanation of how the price increases were sustained. Their statistical results show that the US oil price control suppressed US production, increased the world demand for oil, and raised Saudi Arabia's output. The US Congress had implemented regulations that put a ceiling on the domestically produced oil, and they also capped refining and marketing profit margins so that the consumers would benefit from the lower prices. A basic problem with this solution was that there would be arbitrage opportunities from buying cheap oil in the US and selling it on for instance the European market. The structure of international oil trading made it possible for OPEC countries to benefit from this. In the international oil market oil is sold under contracts between producers and major oil companies or in the spot market. While the spot prices surged after the start of the embargo in 1973, contract prices remained lower and increased more slowly, as OPEC gained confidence that the market would support higher prices. The OPEC countries, and especially Saudi Arabia, took advantage of this situation, by selling expensive spot market oil to the US and gradually canceling long-term contracts. Remaining low-priced contract oil was sold to the European market up to the point where prices equalized (the weighted average of imported and domestically produced oil in the US equaled prices in Europe, eliminating arbitrage opportunities). This process can account for a large share of the wealth transfer from oil importers to oil producers. Based on the statistical results, Alhajji and Huettner estimate that Saudi Arabia's extra revenues from price control in the US amount to $508 billion (current, that is year 2000, dollars), that is, about seven times Saudi Arabia's current GDP. After the substantial decline in real oil prices since the early 1980s, the economic rent earned by oil producers over and above the resource rent has dropped considerably.

This explanation of how the oil price increases were sustained after the 1973 embargo and how wealth was transferred from oil importers, primarily the US, to oil producers does not depend on the existence of a powerful cartel. Rather, a combination of political events, misplaced economic policies, and the nationalization of petroleum resources seems to explain much of the development.

Other tests of whether OPEC behaves as a cartel have found more support for the cartel hypothesis. Griffin (1985) is one example. He states that shares of cartel members should be expected to fluctuate with the price level in accordance with the differential elasticities of members' individual marginal cost schedules. By testing this for OPEC data, he cannot reject this relationship, and concludes that this supports the existence of a cartel. A problem with this and many other tests is that they test the behavior for some characteristic that support the cartel hypothesis, but this characteristic is rarely unique for cartels. Smith (2002) has stressed this point, noting that a test of elasticities may support both a perfect competition model and a dominant firm/cartel model. This critique seems valid for Griffin's test above. Smith further says that this weakens the argument of Alhajji and Huettner described above, and he claims that the oil market might as well be perfectly competitive. I have a hard time understanding this. If the market for crude oil were perfectly competitive how could then some producers have inelastic demand and others not? And, further, how could possibly high-cost and small-reserve producers substantially improve its market shares in a perfect-competition environment? Even though Smith does not answer this, he adds a more damaging critique to Alhajji and Huettner's tests. With the reported elasticities of demand for oil and non-OPEC oil, the elasticities of OPEC demand reported would imply that OPEC's market share must have been larger than 1 in several periods. When Smith corrects for the true market shares and takes the reported β and δ_n as true, he finds that the demand elasticities for OPEC then ranges between –1,66 and -0,67, with the elasticity less than one in more than fifty percent of the times. This means that Alhajji and Huettner's evidence suddenly looks a lot weaker.

Other tests have checked whether, as expected, low-cost producers (which within OPEC primarily also are large holder of reserves) have higher production than high-cost producers. This would be the case in a profit-maximizing cartel, but it would also in theory be the case in any other market structure, whether it be perfectly or imperfectly competitive. In this respect, Smiths critique is in place. It is important for statistical tests to have high power, that is, that they do not easily fail to reject the null hypothesis if it in reality should be rejected. Statistical tests in the oil market lack power not only due to uncertain elasticities of demand, but also due to uncertain marginal costs and uncertain scarcity rents. This has made it important to construct tests with results largely invariant to these uncertainties, with the consequence that the results often are inconclusive with respect to the true structure of the market.

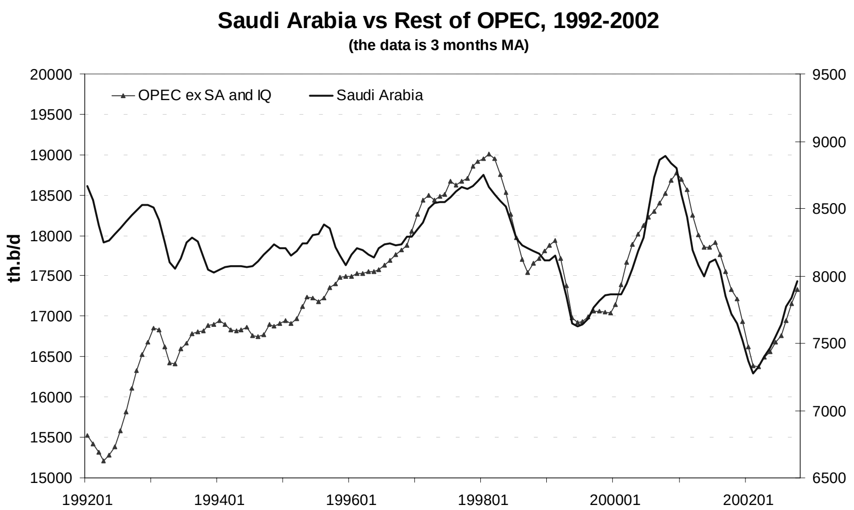
To sum up, we are left with the problem of how to explain prices far above user cost (marginal cost of production plus scarcity rent) of the most efficient producers without having the opportunity to claim that the functioning of a pure explicit cartel is the main cause of this.

How can one unite significant and continuous cheating on quotas with, especially during the 1990s, production behavior that overall, especially over the past five years, has moved in tandem? It seems clear that OPEC does not work as a pure group of explicitly colluding partners, but even though the phenomenon of quota violations is a constant, there seems to be a tacit understanding that some extent of cooperation is both beneficial and possible.

From Hogan and Holland's analysis we remember that exhaustibility makes collusion more difficult than what is the case for infinitely repeated games. Perhaps OPEC has understood this, and is content with a level of collusion that perhaps does not give them monopoly profits, but at least something better than the competitive alternative?

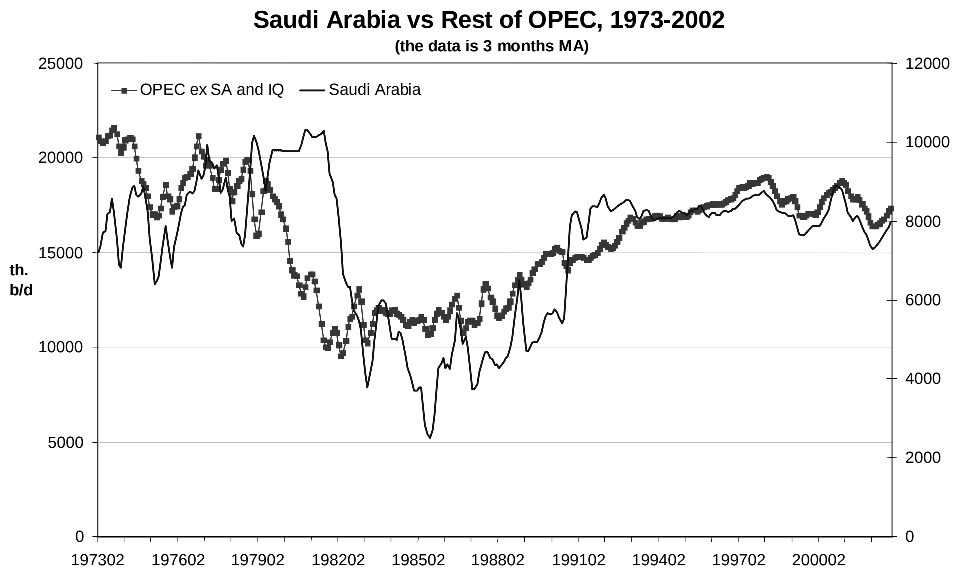
What is more surprising than quota violations is the continued membership of small reserve holders that apparently has a lot to gain from leaving OPEC. What complicates the picture is that many of the OPEC members are controlled by authoritarian regimes where standard economic optimization might be less important than political factors. The political climate is turbulent in several countries (Algeria, Nigeria, Libya, Indonesia, Venezuela). Perhaps, the countries will rethink their membership (if and) when political stability settles down?

The graph below shows how the production level of Saudi Arabia and the other countries of OPEC excl. Iraq have developed over the past decade. We can see that after conditions had normalized after the Gulf War, the discipline of OPEC seems to have gained strength. Even though quotas are nothing but pro forma, the member countries seem to adjust production levels according to the need of the group as a whole. Of course, Venezuela was a major exception during the 1990s, but in aggregate the action seems to be concerted even though none of the prerequisites of cartel success are present.



Data: EIA

We should also contrast the development during the late 1990s and the early 2000s with the one during earlier period. The fluctuations have been erratic and substantial, and the noteworthy fact of OPEC behavior during the 1990s is its relative stability.



Data: EIA

To a large extent this newfound stability must be attributed to relative political stability. However, is this all there is to it, or are other factors contributing to this recent trend? The rest of this paper will be concerned with possible answers to this question.

Part 6 - Saudi Arabia: Accommodation Due to Public Finance Difficulties

Saudi Arabia has been regarded as the most influential and powerful player among the oil producers over the past decades. Due to its position as the producer with the world's largest proven oil reserves, Saudi Arabia has a long-term interest in preserving the relative stability of the world's oil markets. The task of capitalizing on these enormous resources is not an easy one, however. Saudi Arabia must consider several factors when developing its oil extraction strategy. First, there is the strong interest in sustaining demand in the long run. This made the Saudi Arabian leaders oppose further price increases during the 1970s, as they increased production in order to compensate for the low production of other OPEC countries. The current official strategy of trying to force the oil prices into a \$22-\$28-band is also publicly justified on these terms. Even though world oil demand is projected to increase steadily over the coming decades, experiences especially from the 1970s have shown that long-run oil demand is relatively elastic. Further, one might expect that a sustained high price of oil would not only trigger substitution to other established sources of energy, but that it would also trigger a dramatic increase into research in alternative sources of energy, especially as the leading research powerhouse of the world, the US, would seek to reduce the costs of being dependent upon oil imports. It is therefore in the interest of Saudi Arabia to avoid triggering these effects. Saudi surplus production capacity has been rationalized on these grounds, as it makes the Saudis able to compensate for production shortfalls due to political or other factors.

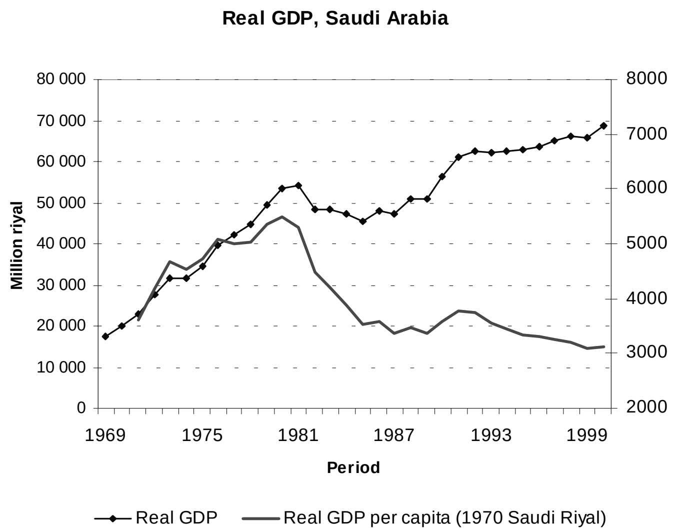
A second factor Saudi Arabia must consider is how to best make its own goals compatible with the strategies of the other producers in the market. While trying to keep the price of oil artificially high relative to the perfect-competition level, the low-cost producers in the Middle East are inviting increased production from higher-cost areas. This directly reduces demand for Saudi oil. An official policy of keeping prices high increases investment in high-cost areas, and further reduces demand directed towards the players who are trying to cooperate in order to sustain the high prices. Further, the stability of a cooperative solution is always under threat, as the marginal profit of deviating always is positive. In order to keep a cooperative solution stable, a credible threat must be developed in order to deter this deviation. But to follow up these threats are costly, especially in the short run. Saudi Arabia has experienced this in several instances, for instance in the 1985-86 and 1998 price collapses. Without any official coherent threat of OPEC as a whole, Saudi Arabia has been the only producer with the capacity to really punish the deviators, and this combined with its long-term horizon, have put a large amount of responsibility upon the Saudis when it comes to controlling the producers' behavior. What all this means is that Saudi Arabia must balance the threat of declining market shares that results from the high-price policy against the considerable, at least short-term, costs of increasing production in order to regain market shares.

The first and second factor would have been no different if Saudi Arabia were a regular oil-extracting company. Of course, this is not the case. That is not to say that a producer of an exhaustible resource under public ownership cannot act as a regular profit-maximizing firm. As long as the producer is allowed to develop its market strategy without regard to any constraints external to the market itself, there is no reason why the national ownership should influence the strategy. One might suspect that this would have been possible for instance in Norway, where budgetary needs to a large extent are independent upon oil revenues. What seems to be the fundamental factor is that Norway (or Great Britain) had climbed the ladder of economic development prior to the discovery of the petroleum wealth. Hence, a sound system of public finance was already developed, and the era of oil revenues did not fundamentally change the way the government collected its revenues. This has not been the case in Saudi Arabia. There annual revenues were considered exactly as that, that is, instead of treating the oil extraction as a process of moving wealth from the ground to financial accounts, they were treated as a cash flow allowing the government to dramatically improve and expand the welfare system. Government revenues in Saudi Arabia are inexorably tied to oil revenues, and there have only been reluctant attempts to find other sources of government revenues. So, because oil revenues largely is the source of being able to provide the people with welfare services, I will argue that we get an constraint external to the oil market when we want to consider the possible strategy set available to Saudi Arabia as an oil producer.

The Saudi Economy

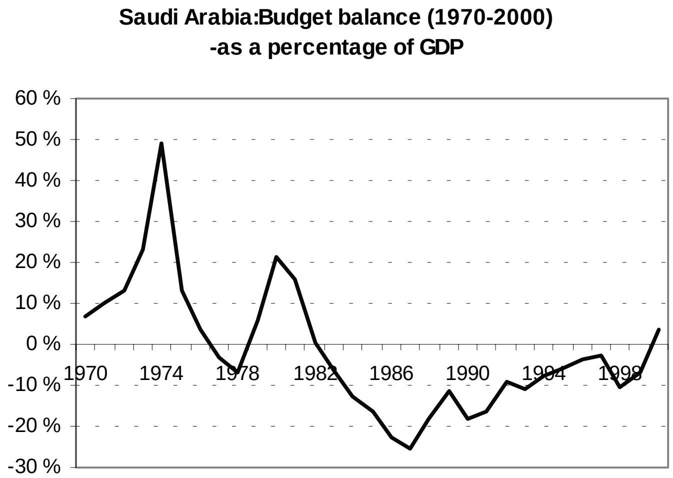
First however, we should take a look on how the Saudi economy has developed over the past couple of decades. This will make it apparent that there may be looming economic troubles ahead. I should first warn you that the quality of the collected data is questionable. Saudi Arabia is notoric for its lack of openness into public affairs. Still, the IMF has encouraged Saudi Arabia for improvements on this area over the past years, and this at least gives us some confidence in the collected numbers.

Let us first take a look at the overall economic picture. The graph below shows the development in Saudi Arabian real GDP. We can see that the spectacular growth during the 1970s was followed by a slump during the 1980s when Saudi Arabia was acting like a swing producer trying to defend the oil price. Since the late 1980s growth has resumed, even though it has been relatively moderate during the 1990s. In per capita terms the picture looks bleak. Real GDP per capita has nearly fallen to half the level of the early 1980s, both due to slow growth in aggregate economic activity itself but particularly due to a high rate of population growth. In fact, the population growth rate in Saudi Arabia is among the highest in the world.



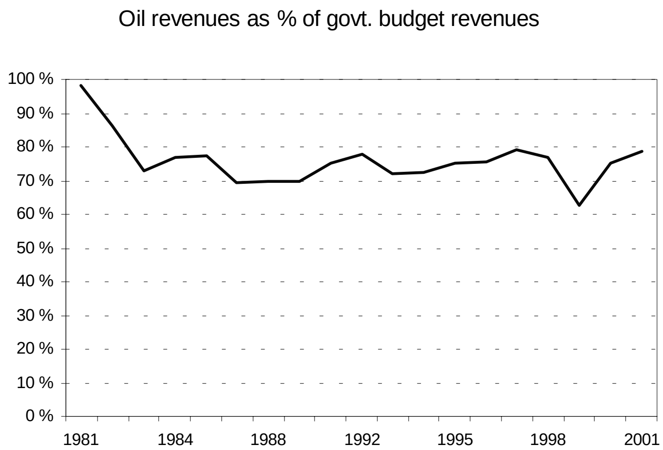
Source: Saudi Arabia Monetary Agency (The Central Bank of Saudi Arabia)

After the oil price increases during the 1970s Saudi Arabia accumulated foreign assets at a rapid pace. Only heavy investments especially in infrastructure around 1980 prevented the budget surplus from being even higher (se graph below). The era of budget surpluses came to halt in 1982 however, and since then there has been a deficit in every single year but year 2000. The large deficits during the 1980s are primarily due to the decline in Saudi oil production (from 1982 to 1986) and later due to the decline in oil prices in 1986. Further, heavy subsidies of Iraq during the Iran-Iraq war in the 1980s added to the problems. According to Gause and Gregory (2000) Saudi Arabia 'subsidized Iraq's war effort against Iran to the tune of nearly $26 billion – not, in retrospect, history's wisest investment.' The Gulf War added to the fiscal problems, and the official cost of this war has officially been reported to be $55 billion, or about the value of one year of public expenditures. During 1990s, the picture has not improved much, even though the political climate has been relatively calm, and the external environment favorable also in other respect. The exception is of course the decline in oil prices during 1998, when a combination of weak Asian oil demand and increased Saudi production in order to punish deviation from Venezuela caused the prices to plummet. We can see that this move by the Saudis was costly in terms of the public financial position.



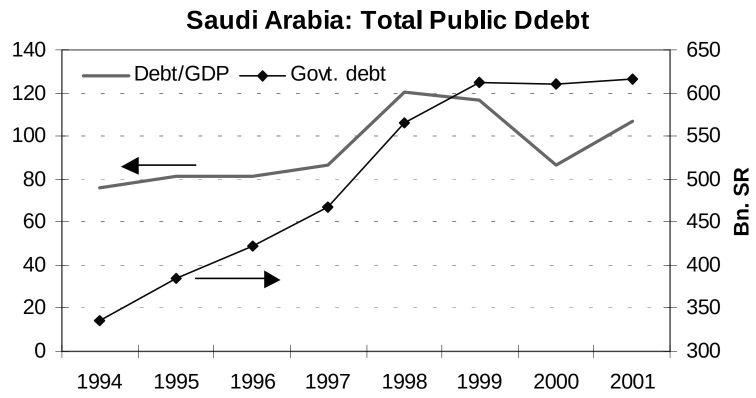
Source: Saudi Arabia Monetary Agency (The Central Bank of Saudi Arabia)

The reliance on oil revenues for covering public expenditure is extremely high, and is not diminishing. Between 70% and 80 % of the public revenues stems from oil production, and the vulnerability of the public finances due to production cutbacks or oil price declines remains high.



Source: Saudi Arabia Monetary Agency (The Central Bank of Saudi Arabia)

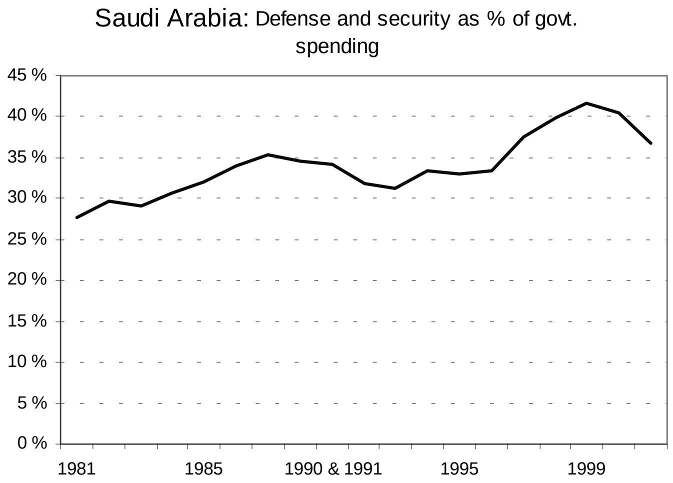
Government borrowing has a short history. Deficits prior to 1987 were financed through sales of foreign assets, but since then the deficits have been financed by selling government bonds. The debt is owed entirely to domestic creditors, after foreign borrowing was terminated in 1994. Most of the public debt is owned by the state pension funds. In 1998 the public debt levels in Saudi Arabia reached new heights. The exact levels are uncertain, but there seems to be a consensus that it reached a level of about 120 % of GDP that year. The pressure against the Saudi currency, the riyal (which is pegged to the US dollar), was high, and the government used a credit line through the state oil company Saudi Aramco to get a USD-loan to defend the currency. Devaluation was avoided, but the vulnerabilities had become apparent.



Source: Estimates made by SAMBA, The Saudi American Bank

(They are in line with the number the Saudi finance minister publicly has indicated)

I am just going to make one final remark before we leave the macroeconomic situation of Saudi Arabia. Below the level of public expenditure on defense and security is illustrated. As you see, Saudi Arabia is allocating enormous amounts of resources to sustain peace and security. Part of the reason why this is so is the desire to prevent hostile powers to feel tempted to take control over the Saudi petroleum wealth. Hence, it is reasonable to include parts of these expenditures in the costs of oil and gas production. Increases in the capacity to extract petroleum raises the risks of being attacked by hostile powers. Further, the long term interests Saudi Arabia has in the markets for petroleum products make it important to sustain some level of peace in the Middle East in order to avoid market shocks that negatively affects Saudi Arabia’s interests. This order comes at a large cost, however, and this adds to the fiscal problems.



Source: Saudi Arabia Monetary Agency (The Central Bank of Saudi Arabia)

In the following I will use some theory of risk aversion in order to illustrate how a weakened wealth position might make a producer less aggressive. We can use this theory to explain how Russia has managed to increase production dramatically without being punished by Saudi Arabia and OPEC. In fact, OPEC tried to force Russia to slow down its output expansion by threatening to force down prices by production increases. Russia, however, saw through this non-credible threat and continued to expand output without being punished.

Theory: Risk Averse Firms in Oligopoly

Just as risk aversion and uncertainty in conjunction creates outcomes favorable for the stability of a cartel, they also influence the intensity of competition between parties in an oligopoly. Competition is usually analyzed in the absence of uncertainty. Still, after the initial articles on the effects of risk aversion on competition by Baron (1970) and Sandmo (1971), this topic has received more interest. The assumption of risk neutrality does not seem valid in the market for oil production, especially as the fiscal situation of the dominating player, Saudi Arabia, has deteriorated over the past years. Outcomes with low prices have gradually become more costly for Saudi Arabia to bear, and I will show that, under assumptions of decreasing absolute risk aversion, this could be modeled as an increase in its risk aversion.

I will now present some underlying theory of how risk aversion affects competition in a Cournot oligopoly. The exposition is to a large extent based on Asplund (2002). Following this, I will apply the theory to the situation in the oil market.

First, assume that each firm wants to maximize its von Neuman-Morgenstern expected utility, $V^i = E(U^i(W^i))$, where the final wealth is given by $W^i = w^i + \pi^i - f^i$. Here, w^i is the initial endowment, π^i is uncertain net operating profits (the uncertainty can relate to either demand or costs), and f^i is fixed costs of production. The utility function is twice continuously differentiable and concave. We will consider a market with two firms, where all parameters are familiar to both players, even the players' attitude towards risk. The profit function is a twice continuously differentiable concave function of its strategy, s^i. We here assume that the strategic variable is the amount of quantity to produce ($s^i=q^i$). We can then denote the best response functions as $b^i(q^j)$. The Nash equilibrium strategies, q^{i*} and q^{j*}, must jointly satisfy $q^{i*}=b^i(q^{j*})$ and $q^{j*}=b^j(q^{i*})$. Quantities are among the class of strategies that are strategic substitutes. This means that $\pi^i_{q^i q^j} < 0$ [29], that is, the best response functions (the reaction curves) are downward sloping.

From the objective function, we can see that the first order condition must be

(6.1) $$V^i_{q^i} = E(U^i_{W^i} W^i_{\pi^i} \pi^i_{q^i}) = E(U^i_{\pi^i} \pi^i_{q^i}) = 0 .$$

We can apply the property $E(U^i_{\pi^i} \pi^i_{q^i}) = E(U^i_{\pi^i}) E(\pi^i_{q^i}) + Cov(U^i_{\pi^i}, \pi^i_{q^i})$ to illustrate how risk aversion affects the choice of best-response strategies. This means that we can write the first order condition as

(6.2) $$E(\pi^i_{q^i}) + \frac{Cov(U^i_{\pi^i}, \pi^i_{q^i})}{E(U^i_{\pi^i})} = 0 .$$

In (6.2) the denominator of the second term on the right hand side is positive (the marginal utility is positive), so the question of how risk aversion affects the reaction functions then depends upon the sign of the covariance. We know that higher profits increases utility and hence make the marginal utility decrease. Hence, there is a negative covariance between profits and marginal utility. This means that we have to establish the relationship between profits and marginal profits in order to establish the sign of the covariance in (6.2). In order to analyze this we will make some assumptions about the source of uncertainty. There is one

[29] Subscripts here denote derivatives

single source of uncertainty, with a continuous or discrete distribution, bounded by upper and lower limits, that is $\varepsilon \in [\varepsilon_L, \varepsilon_H]$. We then assume that profits and marginal profits are monotone in ε for all q^i, q^j. This means that

(6.3) $$\begin{cases} \pi^i_\varepsilon(q^i, q^j, \varepsilon) > 0, \forall q^i, q^j \text{, and either} \\ \pi^i_{q^i\varepsilon}(q^i, q^j, \varepsilon) > 0, \forall q^i, q^j \text{, which is denoted } \rho^i>0\text{; or} \\ \pi^i_{q^i\varepsilon}(q^i, q^j, \varepsilon) > 0, \forall q^i, q^j \text{, which is denoted } \rho^i<0. \end{cases}$$

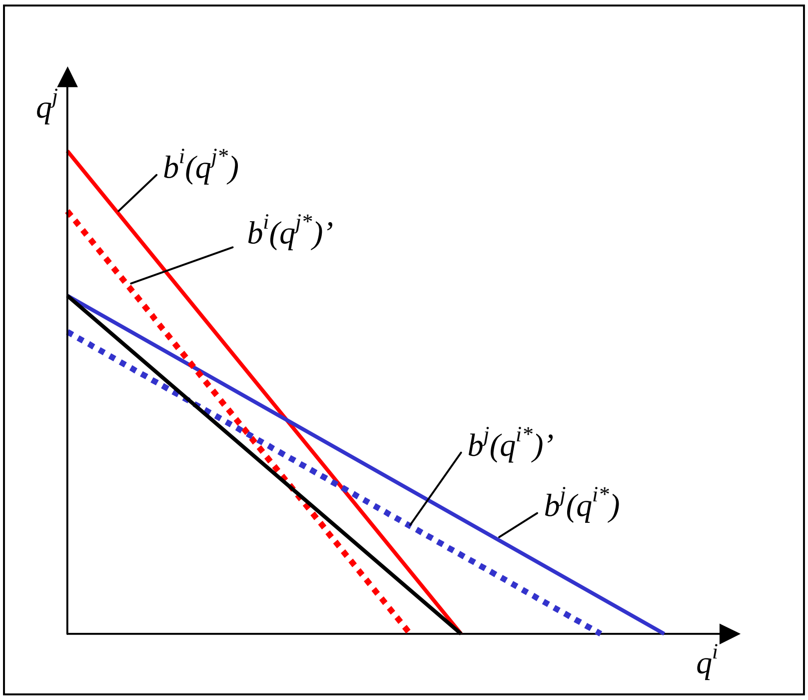
What this says is that profits and marginal profits are positively correlated if the uncertainty affects both of them positively, and negatively correlated if the marginal profits are affected negatively (since we have assumed that the uncertainty affects profits positively). We then see that marginal profits are negatively (positively) correlated with marginal utility if $\rho^i>0$ ($\rho^i<0$). We also see that if (6.2) is to be satisfied, the expected marginal profits must be of the same sign as that of the correlation between profits and marginal profits, as it must be the opposite of the correlation between marginal utility and marginal profits. If the second term of (6.2) is negative, we must have that the expected marginal profits are positive. Under Cournot competition, this means that quantities are lower than under risk neutral behavior. Still, we want to know if we can say something specific about the sign of ρ^i. Let us assume that the profit function is given by $\pi^i(q^i, q^j, \varepsilon) = (a - b_1 q^i - b_2 q^i - c)q^i + f$, where the parenthesis on the right hand side represents price minus (a constant) marginal cost. ε may now refer to uncertainty with respect to either a, b_1, b_2, c, or f. If we now take the derivatives of the profits and marginal profits with respect to the uncertain variable, we find that the following must hold:

Uncertain variable	$a \in (\varepsilon_L, \varepsilon_H)$	$b_1 \in (\varepsilon_L, \varepsilon_H)$	$b_2 \in (\varepsilon_L, \varepsilon_H)$	$c \in (\varepsilon_L, \varepsilon_H)$	$f \in (\varepsilon_L, \varepsilon_H)$
Correlation between π^i_ε and $\pi^i_{q^i\varepsilon}$	$\rho^i>0$	$\rho^i>0$	$\rho^i>0$	$\rho^i>0$	$\rho^i=0$

We see that the correlation is positive for all variables except for the fixed cost, *f*. This means that *uncertainty* in the level of fixed cost is irrelevant for the choice of strategy (as long as they are not so high as to prohibit participation). Asplund (2002) discusses the general validity of (6.2). With cost uncertainty (6.2) will hold as long as the uncertainty changes total variable costs and marginal costs in the same direction for all quantities produced. With demand uncertainty proposition (6.2) may be violated, for instance in the case of uncertain degree of product differentiation, in which case the effect of uncertainty on demand and marginal demand, which must have the same sign for (6.2) to hold, might have the opposite sign for some combination of quantities.

We can now use the above analysis to see what the consequences of risk aversion on Cournot competition will be. For uncertainty in all variables except *f*, the second term on the right hand side of (6.1) will be negative, as marginal profits are negatively correlated with marginal utility. This must imply that the first term on the right hand side of (6.1) is positive, which means that the expected marginal profit must be larger than zero. The effect upon the best response strategies will then imply that the reaction functions move towards the origin, and that total output is reduced. This softening of competition is illustrated in the figure below. Here we can see the effect upon the reaction functions for the Cournot-players *i* and *j* for an increase in both players' risk aversion. We can see that player *i*'s reaction function shifts from $b^i(q^{j*})$ to $b^i(q^{j*})'$ while the effect on player *j*'s best response is a shift from $b^j(q^{i*})$ to $b^j(q^{i*})'$. In equilibrium the quantity produced declines for both producers, and we can see that the equilibrium moves closer to the cartel production path.

Risk aversion affecting Cournot reaction functions



The intuition behind this result is that increases in risk aversion means that the best response functions must give higher profits and marginal profits in bad states, and lower profits and marginal profits in good states. The way to accomplish this is by producing less for each quantity produced by the competitor. This serves as a guarantee for the risk averse firm, as profits and prices are protected by a relatively low output in states when realized costs are high, or realized demand low.

Before we apply this theory to the developments of the world oil market, we should add one more element into the discussion on how risk aversion affects producer behavior. We know that with risk neutral players the choice of strategy in a static game is independent of fixed costs, f^i, and initial endowment levels, w^i. These elements have 'disappeared' when arriving at the first order condition and never come into play. This is not in general the case under risk aversion. Uncertainty with respect to fixed costs or initial endowments will not influence the strategy, but the level of these factors will. To explain this we should go through some basic theory regarding risk aversion.

There are two main measures of risk aversion; absolute and relative risk aversion, both related to the concavity of the utility function, or, in other words, how rapidly the utility of marginal increases in wealth declines. Absolute risk aversion is defined by the negative of the ratio between the second derivative and the first derivative of the utility function, or formally

$$r_A = -\frac{u''(W)}{u'(W)}.$$

Absolute risk aversion is suited to the comparison of attitudes towards risky projects whose outcomes are absolute gains or losses from current wealth. Relative risk aversion, on the other hand, takes account of how risk aversion changes with percentage gains or losses of current wealth. The measure of relative risk aversion is given by

$$r_R = -W\frac{u''(W)}{u'(W)}.$$

As Mas-Colell (1995) says, it is a common contention that wealthier people are willing to bear more risk than poorer people, probably because richer people to a larger extent can afford to take a chance, as the consequences of bad outcomes are not 'too' bad. This implies that decision makers should exhibit behavior in accordance with decreasing absolute risk aversion (DARA).

If we accept the hypothesis that risk-averse decision makers have utility functions with DARA, this gives us some clear implications with respect to the equilibrium strategies of our Cournot players. In our case I find it most interesting to discuss how the level of initial endowment influences the equilibrium strategies. It should be noted that the effects of equal absolute changes in the levels of initial endowment and fixed costs with respect to behavior are equivalent.

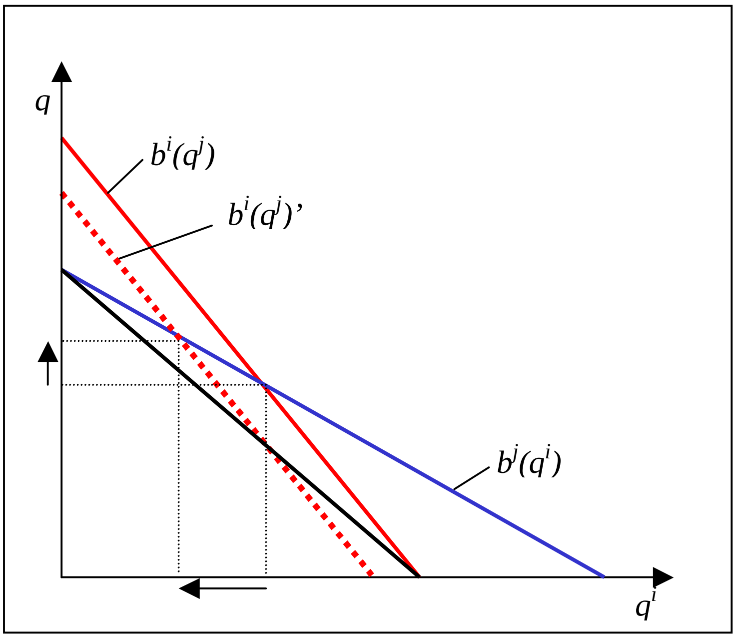
We first have to show how increases in risk aversion affect the best-response strategies of Cournot players. This can be done by defining a new utility function $\hat{U}^i(W^i) = G(U^i(W^i))$, where G is a positive, increasing, and concave function. This implies that player i is more risk averse with utility function $\hat{U}^i$ than with U^i. The first order condition now becomes

(6.4) $$\hat{V}^i_{q^i} = E(G_{U^i} U^i_{W^i} W^i_{\pi^i} \pi^i_{q^i}) = E(G_{U^i} U^i_{\pi^i} \pi^i_{q^i}) = 0$$

With marginal profits positively correlated to profits, we now have that realizations with low marginal profits will carry a greater weight with $\hat{U}^i$ than with U^i, since G_{U^i} will take on higher values for low marginal profits than for high. This means that the best response strategies for bad outcomes will carry greater weight for more risk-averse producers. We also know that the value of the strategic variable quantity takes on lower values in the bad realizations (reduced demand, higher costs), and this means that $\hat{b}^i(q^j) < b^i(q^j)$, that is, an increase in risk aversion makes the producers less aggressive, and in a Cournot game this means that the quantity produced for a given production level of the competitor's will decline.

This result can be combined with our assumption of DARA. Since DARA implies that a decline in the level of initial endowment will increase the absolute risk aversion of a producer, we must have that a fall in w^i (or an increase in f^i) will imply that the Cournot-player reduces his best-response production levels such that the reaction function moves toward the origin. This is shown in the figure below. We can see how an increase in risk aversion for player i leads to a shift in his best-response function towards the origin. The result is increased production for player j, and reduced production for player i.

An increase in risk aversion for player *i*



Now we are finally in a position to apply this theory to the oil market. Recently, I think we really could see how reduced endowment levels might affect the best-response strategies of producers in the oil market. In early 2001 Saudi Arabia and OPEC told Russia that they would only cut production if Russia did the same thing. Prices were then falling toward $20 per barrel, and especially Saudi Arabia relied on high oil prices in order to stabilize the level of public debt, which we have seen had increased to unsustainable levels. The situation was quite different for Russia, which recently had seen it currency being devalued by 75 percent, making USD-income from oil exports quite lucrative. High income from oil exports assisted Russia in regaining some of the lost confidence of the violent market forces. To this came the fact that Russia's oil industry is very fragmented, with little government control over production and export levels. All this meant that Saudi Arabia's threat of responding to increased Russian production by further production increases was not credible, which of course Russia understood. Russia kept increasing production and export levels, gradually becoming the number one producer of the world, while OPEC cut both quotas and production levels to defend prices.

It seems reasonable that a risk averse House of Saud, the royal family of Saudi Arabia, can account for this development. With domestic unrest due to declining standards of living for many of the country's residents, and with threats of a coup d'etat gaining strength, this past period has not given the Saudi regime any room of maneuver when developing its oil producing strategy. Its best response is to act passively in order to being able to provide public goods in realizations with relatively low crude oil demand. Hence, decreasing absolute risk aversion combined with increased debt levels might explain the passive strategy of Saudi Arabia.

How then can we explain that other OPEC countries have followed Saudi Arabia's desires to an extent never seen earlier in OPEC's history? A possible explanation might be that all OPEC members are dependent upon a stable Saudi Arabia, and aggressive cheating might trigger political events in Saudi Arabia that can cause major disruptions in the world oil market.

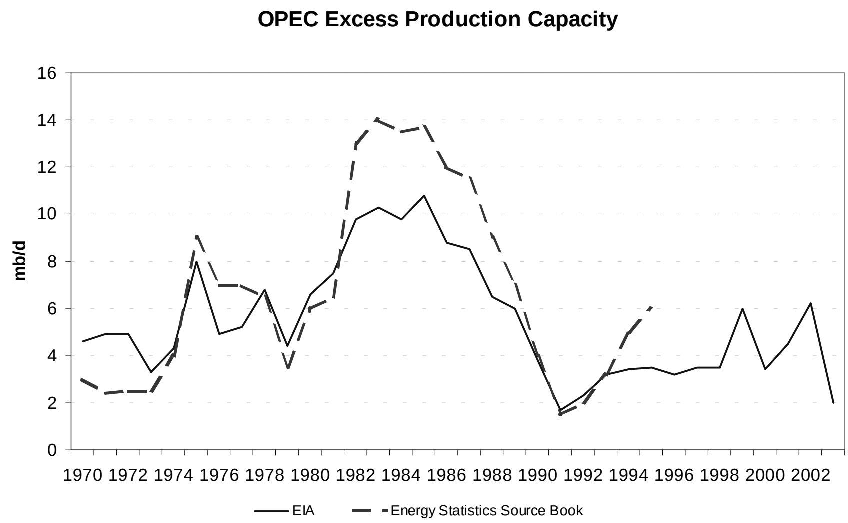
The distribution of market shares among oil producers also suggests that a country's economic system can have important consequences for the aggressiveness of a producer's behavior. Countries with nationalized resource ownership and production can easily restrain production in order to push up prices, while this is harder for countries with national resource ownership but private production. In countries where it is important to build (rebuild) long-term relations with private investors, a government is well served by committing not to interfere with private economic decisions. Fear of nationalization and rigid production restraints will make private capital shy away from you. It seems reasonable to assume that the long-term development of the political and economic system is more fundamentally important than short-term fluctuations in oil revenues, making this commitment credible. Hence, it is not surprising that Russia, with its fragmented collection of oil producers, and Venezuela during the 1990s, when the attraction of foreign private capital was given high priority, saw its market shares in the world oil market increase. The combined effects of reduced risk aversion with a move from public to (partially) private production and increased commitment to gain the confidence of private capital seem to be a basis of more aggressive best-response functions.

Part 7 - Risk Aversion Promoting Cartel Stability

Most economists have been skeptical with respect to the long-term stability of OPEC. We have already seen that OPEC lacks most of the characteristics of a traditional cartel, and it is also claimed that the organization as a whole produces at the inelastic part of the demand curve. This behavior counters what we would predict from economic theory. Cheating is a constant phenomenon within OPEC, but the extent of the cheating lacks the force that would lead to a complete breakdown of cooperation. We should then ask ourselves the following question: With high marginal profits from (aggressive) cheating, what factors prevents this from taking place?

From our discussion above we have seen that production capacity is one of the factors determining the allocation of quotas within OPEC. The first thought that comes to mind when one hears this, is that one would expect that this allocation mechanism would lead to increased investment in capacity by all members. Like in an arms race, no player would like to see his position threatened by aggressive action by the others, and he then takes preemptive action building up capacity himself. It may be possible to find a stable equilibrium, but this equilibrium is unlikely to be optimal collectively (that is, for the group as a whole). An equilibrium is likely to be characterized by massive over-capacity in order to ensure that one gets high quotas oneself and in order to reduce the attractiveness for others to invest. Of course, in the presence of asymmetric information, no escalation of real production capacity would be necessary if one had the means to convince the other cartel members that ones capacity is higher than it in fact is.

The data, however, does not show that OPEC has been suffering from this escalation of production capacity. Of course, at times the degree of capacity utilization has been relatively low, but over the past decade it seems to have been increasing.



Sources: EIA and The Energy Statistics Sourcebook

The figure above shows the development of excess production capacity within OPEC over the past three decades. The true amount of excess capacity is uncertain, and the estimates vary somewhat. I have only found data on the full time series from the US Department of Energy's Energy Information Agency. The current consensus is that the excess capacity within OPEC is quite low, around 2 mbd, with Saudi Arabia accounting for most of this excess capacity. Hence, most countries within OPEC are producing at close to full capacity, and the problem of deterring (substantial) cheating seems entirely absent.

What is striking about the figure above, is how low the level of excess capacity has been over the past decade relative to the mid-1980s. If we take into account that Saudi Arabia always has kept an excess capacity of about 2-4 mbd, we can immediately see that the other producers cannot have large amounts of excess production capacity.

Data on how relative quotas have developed over the past decade show that only Venezuela has seen its relative quota improve, while those of the other countries have been relatively stable. It is also the case that even though the reserve levels of most OPEC countries are high, there has hardly been any increases in capacity levels.

This reluctance to make investments in production capacity is somewhat surprising, especially since we have seen that Venezuela has strengthened its position within OPEC after it decided to increase capacity and production in the mid 1990s. One possible solution is to leave the assumption that OPEC producers are risk neutral. A standard assumption in economic theory is exactly that producers are risk neutral. The justification must be that firms are owned by well-diversified investors who have spread their investments in a manner that has eliminated all exposure to unsystematic risk. When this is done, it is in the interest of the owners that the firm maximizes expected profits, not expected utility. To be fair, the assumption of risk neutrality does not seem to fit the members of OPEC. One possibility for owners of large petroleum resources is to diversify risk by selling stakes in their oil wealth and then invest the proceeds in projects with a return that, to as large an extent as possible, is negatively correlated to the return in the market for crude oil production. The OPEC countries have done almost exactly the opposite. The oil wealth has been nationalized, with the consequence that the exposure to risk increases. Further, oil revenues are to a large extent paying for the provision of public goods, and with politicians and regimes that are interested in staying in power, this quite certainly influences their risk attitude as oil producers.

The description of the economic development of Saudi Arabia above showed us that the vulnerability with respects to negative shocks in the oil market has grown over time. Above, we also saw that a loss in wealth combined with decreasing absolute risk aversion implies increased aversion to risk. Further, one might also argue that the drop in oil prices in the mid-1980s exposed the OPEC members to the inherent volatility in oil prices and oil revenues. With increased uncertainty with respect to oil revenues, the optimal level of investment in excess capacity also gets uncertain, and we would expect a drop in the aggressiveness in capacity building.

With this in mind, I will show how risk aversion among the cartel members may contribute to the recent lack of investment in capacity. The basic intuition is that risk aversion makes the producers give relatively greater weight to the outcomes with low profits, or here, low consumption of public goods. The best response strategies for the players will therefore be directed towards relatively high consumption in the states that affect consumption negatively.

I will make the model simple[30]. The model is one of semicollusion, where the producers are perfectly colluding in the output market, but each producer is competing with the others in production capacity in order to increase his the share of the cartel output. First, we assume that the cartel is composed of identical players. This means that all of them will be faced with the same optimization problem, and also have the same utility functions with identical risk aversion parameters. I will consider a static model with only one period. The producers' problem is to maximize consumption from its initial endowment and profits from oil production. The decision variable is the amount to invest in capacity. Each producer gets his share of the cartel (monopoly) profits, as we assume that the cartel operates alone in the market, and that it manages to allocate quotas such that production takes place on the cartel

[30] Professor Lars Sørgard has told me that this model has many parallels to those of an article by Fershtman and Gandal (1994).

contract surface. Quotas are allocated based on relative production capacity. After looking at the result under risk neutrality, we will look at the problem in a world of uncertainty and risk aversion. The uncertainty arrives as we assume that the cartel profits are uncertain due to shifts in demand. To simplify the exposition, we will assume that there are only two players involved in the cartel. These players have identical oil production costs, but the cost of increasing capacity may vary between the players.

Let us first however look at the model with no uncertainty. This means that utility maximization will give the same result as profit maximization. We are now interested only in the process of opportunistic investment in capacity, in order to get a larger share of the allocated quotas. Therefore we assume that the producers already have enough capacity for the production of the monopoly/cartel output, and we assume (perhaps unrealistically) that the current perfect stability of the cartel is unaffected by escalation of capacity. The players are then faced with a static game situation, where they can invest in additional capacity in order to steal quotas. The objective function of producer *i* takes the following form:

(7.1) $$V^i = w^i + \left(\pi^M \frac{q_+^i + (q^M/2)}{q_+^i + q_+^j + q^M} \right) - k^i q_+^i - f^i .$$

This function gives us the value of the producer's wealth as a function of w^i, the initial endowment, the profits from oil production, where the share of the monopoly profit is given by the share of aggregate production capacity. The amount of surplus capacity is given by q^i_+, and the amount of capacity prior to the increase in capacity is equal for both players, and given by ($q^M/2$), as we assume that the players are equally positioned prior to this static game. The variable cost of building up capacity is given by k^i, whereas there is also a fixed cost, f^i, involved in the cost of increasing capacity.

The first order condition of this problem is

(7.2) $$\frac{\partial V^i}{\partial q_+^i} = \pi^M \frac{q_+^j + (q^M/2)}{(q_+^i + q_+^j + q^M)^2} - k^i = 0 .$$

This condition tells us that the marginal benefit from building up excess capacity must in equilibrium equal the marginal cost of doing this. After some manipulation we get the following best-response functions:

(7.3)

$$b^i(q_+^j) = q_+^i = \sqrt{\frac{\pi^M}{k^i}[q_+^j + (q^M/2)]} - q_+^j - q^M .$$

$$b^j(q_+^i) = q_+^j = \sqrt{\frac{\pi^M}{k^j}[q_+^i + (q^M/2)]} - q_+^i - q^M$$

By substituting the best response function for player *j* into player *i*'s best-response function, we can solve for $q^i{}_+$. Due to the symmetric problems for the two players, we have then found the following Nash equilibrium strategies:

(7.4)

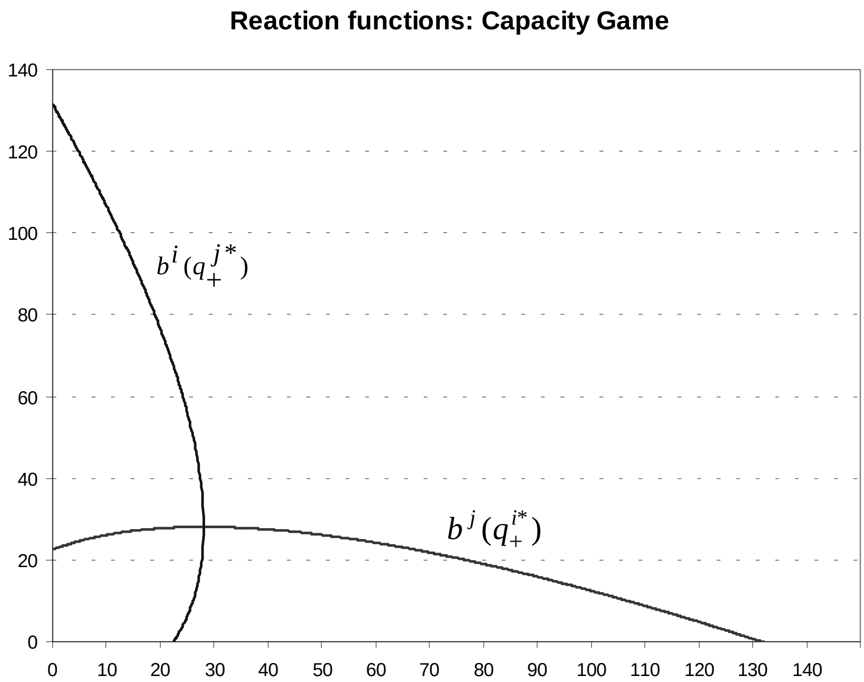
$$b^i(q_+^{j*}) = q_+^{i*} = \frac{\pi^M}{k^j}\frac{1}{\left(k^i/k^j + 1\right)^2} - \frac{q^M}{2}$$

$$b^j(q_+^{i*}) = q_+^{j*} = \frac{\pi^M}{k^i}\frac{1}{\left(k^j/k^i + 1\right)^2} - \frac{q^M}{2}$$

The slope of the reaction function is given by the derivative of the reaction function with respect to the opponent's strategic variable, here q_+. This gives us the following relationship:

(7.5) $$\frac{\partial b^i(q_+^{j*})}{\partial q_+^j} = \frac{1}{2}\left(\frac{(\pi^M/k^i)}{q_+^j + (q^M/2)}\right)^{\frac{1}{2}} - 1 .$$

From this we can see that the sign of the slope of the reaction function varies over $q^j{}_+$. When your opponent does not invest heavily in excess capacity you respond aggressively if he were to increase his capacity. However, as the opponent builds up a greater arsenal of excess capacity, you become softer, and reduce your own investments. This means that excess capacity in some ranges works as strategic substitutes (when you reduce your investments for increases of the opponent's), and in other ranges as strategic complements (when you respond by increasing your own investments). This relationship can be seen from the figure below[31].



Let us now introduce uncertainty with respect to demand and cartel profits. For simplicity, we will assume that the cartel precommits to produce the expected optimal cartel output, which is given by the monopoly output $q^M=(a-c)/2$ with no uncertainty. After the output is produced demand is realized, with the inverse demand given by $p(q^M)=a+\varepsilon-q^M$. The objective of the producer is now to maximize his expected utility, which is given by the following expression:

(7.6) $$V^i = E(U^i(W^i)) = E\left[U^i\left(w^i + \left(\pi^M \frac{q_+^i + (q^M/2)}{q_+^i + q_+^j + q^M}\right) - k^i q_+^i - f^i\right)\right]$$

Here we assume that the utility function is twice differentiable and concave, with $U^i_{W^i} > 0$ and $U^i_{W^iW^i} \leq 0$. We have assumed that profits and production is unaffected by the investment

[31] The figure has been made based on a model with linear demand (p=a-Q), and with the following values of the parameters: a=100, c1=c2=10, k1=k2=10.

in capacity, and since extraction costs are the same for both players, the size of the profits is unaffected by how production is split between the two cartel members. Differentiating the objective function with respect to $q^i{}_+$, yields the first order condition

(7.7) $$V^i_{q^i_+} = E(U^i_{W^i} W^i_{q^i_+}) = 0$$

To get a sense of how risk aversion influences the producers' strategy choice, we can rewrite (7.7) by making use of the property $E(U^i_{W^i} W^i_{q^1_+}) = E(U^i_{W^i})E(W^i_{q^i_+}) + Cov(U^i_{W^i}, W^i_{q^i_+})$. We then get that

(7.8) $$E(W^i_{q^i_+}) + \frac{Cov(U^i_{W^i}, W^i_{q^i_+})}{E(U^i_{W^i})} = 0 .$$

We know that the denominator in (7.8) is positive. Further, when the agent is risk neutral, the covariance term is zero, and the problem is just like the one we solved above. With risk aversion, we know that profits and marginal utility is negatively related, as marginal utility declines when profits increase. To assess how risk aversion affects the strategy choice, we therefore have to establish the relationship between marginal profits and profits, which will then make us able to establish the sign of the covariance term. We have assumed that demand uncertainty takes the form of $p(q) = a + \varepsilon - q$, where ε is drawn from some given distribution and with some upper and lower bound ($\varepsilon \in (\varepsilon_L, \varepsilon^H)$), the latter defined so that a cannot be negative. The derivative of final wealth with respect to the disturbance term, $\partial E(W^i)/\partial \varepsilon$, is now positive, and also the derivative of marginal final wealth with respect to the disturbance term, $\partial^2 E(W^i)/\partial q^i_+ \partial \varepsilon$, is positive. This means that profits and marginal profits are positively correlated, and further that marginal utility and marginal profits are negatively correlated. For (7.8) to hold, we must then have that $E(W^i_{q^i_+}) > 0$. By using this result on the problem solved with no uncertainty, we can see the implications for the producers' reaction functions. Condition (7.2) is now

(7.2') $$\frac{\partial W^i}{\partial q^i_+} = \pi^M \frac{q^j_+ + (q^M/2)}{(q^i_+ + q^j_+ + q^M)^2} - k^i > 0 .$$

Again we solve for q^i_+ in order to find *i*'s best response function, which in this case gives us

$$(7.3') \quad b^i(q^j_+) = q^i_+ < \sqrt{\frac{\pi^M}{k^i}[q^j_+ + (q^M / 2)]} - q^j_+ - q^M .$$

We can see that demand uncertainty and risk aversion will make the best response for each producer less aggressive than in the case of risk neutrality/no uncertainty. The implication for the reaction functions is that they will shift towards the origin, and in equilibrium the amount of excess capacity will be reduced. This theory of risk aversion can help us understand why the countries are acting more passively within the OPEC than what one might suspect would be the case, and that the equivalent of an arms race has not developed.

It might be commented that there still exists some excess capacity within OPEC. However, this excess capacity does not seem to be developed with the purpose of gaining larger production quotas. Rather, the excess capacity is almost single-handedly held by Saudi Arabia, and should be regarded as a stabilizing investment rather than as an opportunistic investment. It seems more reasonable to consider the Saudi excess capacity as the factor that makes collusion possible by giving Saudi Arabia the means to punish deviations. Further, since Saudi Arabia is the producer with the longest horizon in the oil market, it is rational for her to invest in excess capacity, especially in order to avoid large price increases due to market disruptions that will cause further substitution away from oil.

As a final comment we might speculate how the troubles of the Saudi economy is affecting cartel cohesion. A possible hypothesis is that the other OPEC members gets more committed to follow OPEC quotas and strategies, as the weakness is the organization's watchdog transfers responsibility to the other members. One would suspect that nobody would like to pressure Saudi Arabia so far that political unrest and increased opposition towards the House of Saud causes a shift in the Saudi leadership, with uncertain and possibly very negative consequences. A mutual understanding within OPEC that this must be avoided might explain the increased cohesion of the cartel over the past years.

Part 8 – Concluding Remarks

In this paper I have tried to give a picture of the dynamics determining the supply of crude oil. It is now time to give some concluding remarks, including some speculations on the future development of this market and its structure.

Resource exhaustibility is a complicating factor when we want to get an understanding of how the oil market works. The predictions from theory are clear when supply of reserves is fixed, but I have shown how the world's remaining oil reserves thus far have grown faster than production. This leaves us with a price path that is dependent upon expectations of future discoveries of oil and the true amount of recoverable reserves. To this form of uncertainty comes the uncertainty with respect to future demand, the potential of substitutes, and the short term supply decisions among OPEC and non-OPEC members. All this contributes to the volatility in crude oil prices.

We have also seen how exhaustibility affects the possibility for collusion. The difficulty of supporting collusion increases as the divergence between the players reserve level increases. This may account for the fact that the core of OPEC consists of the world's largest reserve holders, and that most countries with low levels of reserves are staying outside the cartel. Still, several members of OPEC are both small in terms of reserves (and production) and they do not seem to benefit from OPEC members in terms of a concave sharing rule. In fact, it seems hard to rationalize their membership purely on economic grounds.

I have also argued that the increased cohesion of OPEC over the past years might stem from the economic weakness of the Saudi Arabian economy. Fear of what will happen if the political regime in Saudi Arabia is toppled have created a common understanding of the need for collective action in order to sustain high oil prices and relatively high government revenues. The cost of this is that OPEC is accommodating increased production among non-OPEC producers, especially Russia.

With respect to the future, I think there are two intertwined major developments that might affect the structure of the market. First, we have the potentially large resources in the Caspian region. Second, we have the scenario that OPEC's minor members will break out of the organization.

With support from US interests and capital, one would expect production in the Caspian region to increase relatively rapidly to substantial levels when (and if) large oil fields are discovered. With continued weakness in the Saudi Arabian economy I think that it will be impossible for OPEC to counter this production expansion with an aggressive response. This will gradually lead to a fall in OPEC's market shares, and, with moderate increases in demand, perhaps also stagnation in the overall OPEC production level. This might trigger the small OPEC members to reconsider their membership.

One might question whether the departure of OPEC's minor members actually will have dramatic consequences for the organization's continued existence. Looking at the distribution of reserves among the OPEC members, we see that the world's six largest owners of reserves are Saudi Arabia, Iraq, United Arab Emirates, Kuwait, Iran, and Venezuela, all OPEC members. The mutual interests of this group of countries would not be negatively affected by departures from comparatively minor countries like Algeria, Indonesia, Libya, Nigeria, and Qatar.

What would be important for the first group of countries if the countries of the Caspian region were to find massive amounts of reserves is that they invite these countries to collude. In theory, the possibility of collusion among the core of OPEC and the countries of the Caspian region countries will increase in the reserves of the latter. However, due to the political support from the US and the interests of US oil companies, the Caspian region countries may find this impossible. In my opinion, nationalized production is of great importance in order to sustain the discipline that is needed to support collusion. Decentralized production will adversely hinder this. Hence, the potential for collusion might be reduced if production in both Iraq and the Caspian region is decentralized. One should not stretch things too far, but the American involvement in both the Caspian region and Iraq could be based, at least partially, on these arguments. In the long run collusive behavior can probably only be sustained if the great majority, if not everyone, of the major reserve holders are members of the cartel. At some critical value, the amount of free riding by non-members will no longer be tolerable.

Part 9 - References

Abreu, D. (1986) Extremal Equilibria of Oligopolistic Supergames, *Journal of Economic Theory*, 39: 191-225

Adelman, M.A. (1993a) *The Economics of Petroleum Supply*, Cambridge, MA, MIT Press

Adelman, M.A. (1993b) Modelling World Oil Supply, *Energy Journal*, Vol. 14(1)

Adelman, M.A. (1995) *The Genie Out of the Bottle – World Oil Since 1970*, Cambridge, MA, MIT Press

Alhajji, A.F and Huettner, D. (2000a) OPEC and Other Commodity Cartels: A Comparison, *Energy Policy*, Volume 28, Issue 15

Alhajji, A.F and Huettner, D. (2000b) The Target Revenue Model and the World Oil Market: Evidence From 1971 to 1994, *The Energy Journal*, Volume 21, No. 15

Asplund, A. (2002) Risk-averse Firms in Oligopoly, *International Journal of Industrial Organization*, 20: 995-1012

Baron, D.P. (1971) Demand Uncertainty, Utility, and Industry Equilibrium in Pure Competition, *International Economic Review*, 12: 196-208

Cairns, R.D. and Davis, G.A. (2001) Adelman's Rule and the Petroleum Firm, *The Energy Journal*, 22:31-54

Claes, D.H. (2001) *The Politics of Oil Producer Cooperation*, Westview Press

Conrad, J.M. (1999) *Resource Economics*, Cambridge University Press

Crémer, J. and D. Salehi-Isfahani, (1991) *Models of the Oil Market*, Chur: Harwood Academic Publishers

Dasgupta, P.S. and Heal, G.M. (1979) *Economic Theory and Exhaustible Resources*, Cambridge University Press

Davis, G.A. and Cairns, R.D. (1998) Simple Analytics of Valuing Producing Petroleum Reserves, *The Energy Journal*, 19:133-142

Fershtman, C., and Gandal, N. (1994)Disadvantageous Semicollusion, *International Journal of Industrial Organization* 12:141-154

Fishelson, G. (1989) Imperfect Competition Under Uncertainty, *Journal of Economics and Business*, 41:253-263

Fudenberg, D. and Maskin, E. (1986) The Folk Theorem for Repeated Games with Discounting and Incomplete Information, *Econometrica*, 54: 533-54

Gause III, F.G. (2000) Saudi Arabia Over a Barrel, *Foreign Affairs*, 79:80-94

Gibbons, R. (1992) *A Primer in Game Theory*, UK, Harvester Wheatsheaf

Griffin, J. and Teece, J. (1982) *OPEC Behavior and World Oil Prices*, George Allen & Unwin

Griffin, J. M. (1985) OPEC Behavior: A Test of Alternative Hypotheses, *American Economic Review*, 75: 954-963.

Griffin, J. and Neilson, W. (1994) The 1985-86 Oil Price Collapse and Afterwards: What Does Game Theory Add? *Economic Inquiry*, Volume 32

Harrington, J.E. (1989) Collusion Among Asymmetric Firms – The Case of Different Discount Factors, *International Journal of Industrial Organization*, 7:289-307

Harrington, J.E. (1991) The Determination of Price and Output Quotas in a Heterogeneous Cartel, *International Economic Review*, 32:767-792

Hartwick, J.M. (1989) *Non-renewable Resources Extraction Programs and Markets*, Harwood Academic Publishers

Hogan, C. and Holland S. (2002) *Collusion With Exhaustible Resources*, Working Paper

Hotelling, H. (1931) The Economics of Exhaustible Resources, *The Journal of Political Economy*, Volume 39, Issue 2

Kneese, A. and Sweeney, J. (1993) *Handbook of Natural Resource Economics, Volume III*, Elsevier Science Publishers B.V.

Kreps, D. and Scheinkman, J. (1983) Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes, *Bell Journal of Economics*, 14: 326-337.

Lewis, T. and Schmalensee, R. (1980) On Oligopolistic Markets For Nonrenewable Natural Resources, *The Quarterly Journal of Economics*, Volume 95

Léonard, D. and Van Long, N. (1992) *Optimal Control Theory and Static Optimization in Economics*, Cambridge University Press

Loderer, C. (1985) A Test of the OPEC Cartel Hypothesis: 1974-1983, *The Journal of Finance*, Volume 40, Issue 3

Loury, G. (1986) A Theory of 'Oil'igopoly: Cournot Equilibrium in Exhaustible Resource Markets With Fixed Supplies, *International Economic Review*, Volume 27, Issue 2

Mason, C. and Polasky, S. (1998) *Non-Renewable Resource Cartels: Who's in the Club?*, Working Paper

Morse, E.L. and James, R. (2002) The Battle for Energy Dominance, *Foreign Affairs*, 81:2

Polasky, S. (1991) Do Oil Producers Act as 'Oil'igopolists?, *Journal of Environmental Economics and Management*, Volume 23, 216-247

Reynolds, D.B. (1999) Modelling OPEC behavior: theories of risk aversion for oil producer decisions, *Energy Policy*, Issue 27,

Rouleau, E. (2002) Trouble in the Kingdom, *Foreign Affairs*, 81:75-89

Salant, S.W. (1976) Exhaustible Resources and Industrial Structure: A Nash-Cournot Approach to the World Oil Market, *Journal of Political Economy*, 84: 1079-1093

Sandmo, A. (1971) On the Theory of the Competitive Firm Under Price Uncertainty, *American Economic Review*, 61: 65-73

Schelling, T.C. (1960) *The Strategy of Conflict*, Cambridge, Mass.: Harvard University Press

Schmidt, R. (1988) Hotelling's Rule Repealed? An Examination of Exhaustible Resource Pricing, *Federal Reserve Bank of San Francisco, Economic Review*

Simon,J. Hotelling's Law: Is It Useful or Wasteful?, Manuscript, *www.juliansimon.org*

Smith, J. (2002) Inscrutable OPEC? Behavioral Tests of the Cartel Hypothesis, Working Paper

Stiglitz, J. (1976) Monopoly and rate of extraction of exhaustible resources, American Economic Review, 66:655-661

Sydsæter, K. (1990) *Matematisk Analyse - Bind II*, Universitetsforlaget

Teece, J. (1982) A Behavioral Analysis of OPEC: An Economic and Political Synthesis, *Journal of Business Administration*, 13: 127-159.

Telhami, S. and Hill, F. (2002) Does Saudi Arabia Still Matter, *Foreign Affairs*, 81:167-185