On Quota Violations of OPEC Members *

Hamed Ghoddusi † Masoud Nili‡ Mahdi Rastad§

Abstract

Over the last decades quota violation has become a norm for the OPEC countries. Yet, the academic literature on OPEC focuses more on its production behavior than on analyzing the quota allocation process or characterizing quota violation patterns. This paper offers a theoretical model accompanied by empirical evidence to explain OPEC members' incentives for abiding with or violating quotas. We first offer a cartel model with a quota allocation rule and an *endogenous* capacity choice. The model highlights the trade-off between building spare capacity to bargain for a higher legitimate quota versus the risk of punishment for quota violation. Using the quarterly data from 1995 to 2007 we empirically support the main results and intuitions for the model. Our empirical evidence is consistent with capacity constraint working as an enforcement mechanism in good times and OPEC relying on quota system in bad times.

JEL classification: D40; L13; C32; Q4

Keywords: Energy; OPEC; Cartels; Oil market; Compliance; Quota; Cournot Duopoly

---

*We thank Robert Kaufmann for sharing data on OPEC members' capacity.

†School of Business, Stevens Institute of Technology, Hoboken, NJ, 07030, hghoddus@stevens.edu

‡Graduate School of Management and Economics, Sharif University of Technology, Azadi St., P.O. Box 11365/8639, Tehran, Iran. E-mail: m.nili@sharif.edu

§Department of Finance, Orfalea College of Business, California Polytechnic State University, San Luis Obispo, CA 93407. E-mail: mrastad@calpoly.edu

Introduction

OPEC has a fragile cartel structure. It does not have a formal enforcement mechanism in place (except for occasional switching to price wars) to incentivize its members to comply with their quota allocations. (Alhajji and Huettner (2000)). As a result, non compliance has become a norm among OPEC members over the past decades. In every single quarter from 1993 to 2005, total OPEC production exceeded the sum of its members' quotas. In this period, an OPEC member over-produced its quota by an average of 6.7%. A question that follows naturally is how to explain such a persistent non-compliance pattern. In this paper, we provide a framework to explain this pattern based on strategic interactions among members inside OPEC.

Quota violations vary significantly for larger members like Saudi Arabia, whose overproduction averaged 3.2% from 1995 to 2007, vs. the smaller ones like Qatar, which overproduced by an average of 18.5% over the same period. In extreme cases, some members overproduce their quotas by a large margin. For example, in the second quarter of 1998, in response to falling oil prices and despite OPEC's request that Qatar cut its production to reach a quota of 384 thousand bpd, the country kept its production level at 670 thousand bpd—a 75% quota non-compliance. Another extreme example is Algeria, which overproduced its quota by more than 50% in 13 out of 24 quarters between 2002 and 2006. Moreover, as suggested by a large body of theoretical literature[1], the difficulty of supporting collusion varies in good and bad times of the market. Based on these preliminary observations, we ask: is non-compliance related to oil market conditions? If so, is it a pro-cyclical or a countercyclical behavior? How does the degree of non-compliance differ for smaller vs. larger producers? To what extent does OPEC rely on the quota system?

There are a number of prominent empirical studies on OPEC behavior, but the majority of them focus on whether OPEC behaves as a cartel or not, rather than analyzing the nature of quota setting and violations. To name a few, Griffin (1985), for example, tests the market-sharing hypothesis by looking at the co-movement of each country's production with that of the rest of OPEC. He concludes that "OPEC is a looser cartel" given that market sharing considerations only partially affect production decisions. Gülen (1996), by using a

[1] See Feuerstein (2005) for a survey of collusion literature.

similar intuition, tests for parallel movements in members' output levels. He finds evidence in support of coordination among members, especially during the rationing period, 1982-1993. Dahl and Yücel (1991) find no evidence in support of several hypotheses for OPEC behavior including dynamic optimization, target revenue, cartel, competitive, and swing producers. Therefore, they conclude that "loose coordination or duopoly" is the closest description to OPEC behavior. However, as pointed out by Smith (2005), there are serious concerns regarding the low power of the statistical tests employed in these studies and the extent to which they are capable of distinguishing between collusive and competitive behavior within OPEC. He describes OPEC as "much more than a non-cooperative oligopoly, but less than a friction-less cartel." Instead of calling this strand of literature inconclusive, we prefer the conclusion that some of the studies reach—such as Geroski et al. (1987), Almoguera et al. (2011) and Kaufmann et al. (2008)—that OPEC behavior cannot be fitted into a single model, but rather it follows a varying-conduct model that switches between collusive and non-cooperative behavior over time. We use the same notion in our theory model, by formulating the proposed framework as a combination of a cooperative game and a non-cooperative behavior.

Although the empirical evidence for OPEC's collusion behavior is mixed, the economic theory in the field of industrial organization provides sharp predictions for one of the factors affecting collusion: the link between market conditions and the incentive to cheat in a cartel. The standard approach in this literature is a repeated game model in which cartel members interact via their choice of production levels or prices. In this approach, members choose to cooperate or cheat in each period by comparing the net value of cooperation vs. that of deviation. If they cooperate, they receive a moderate period payoff for cooperation with the promise of a continuation of this payoff in the future. However, if they decide to cheat, they will enjoy a high period payoff but will face punishment in the future. The common prediction of this literature is that it's harder to collude in booms because the incentive to cheat is higher[2]. For example, Rotemberg and Saloner (1986a) show that with an i.i.d. demand structure collusion is harder to support in booms, simply because the net gains of deviation are higher. Assuming a cyclical

[2]In contrast, Green and Porter (1984a) predict that price wars occur in periods of low demand. However, the key assumption in their model is that firms cannot observe the demand or the production of other firms, which is clearly not plausible for OPEC.

demand, Haltiwanger and Harrington Jr (1991) show that the incentive to cheat is stronger at the end of a boom when the demand is about to fall. This is because with a falling demand the value of cooperation in the future is at its lowest. Consistent with Rotemberg and Saloner (1986a), Kandori (1991) and Bagwell and Staiger (1995) generalize this result for the cases of serially correlated and Markove demand shocks respectively. Overall, the theory predicts a monotonic relationship between market conditions and the incentives to cheat: the more favorable the market conditions, the more difficult it is to support collusion.[3] Staiger and Wolak (1992b) introduce the capacity-constraints features to previous models and show that large excess capacities can result in severe price wars.

We use a less complex economy to characterize the non-compliance behavior of an OPEC-like cartel. Our model consists of a small- and a large-capacity country. We use a simple multistage model where in the first period countries choose their optimal level of capacity. In the second stage the optimal aggregate production of the cartel and the quota level for each member is determined through maximizing the joint profit of the cartel members. In the third stage countries decide about the actual production and in the forth stage possible deviations from the quota might be punished.

The model implies that OPEC members take into account the investment cost of building capacity with the possible benefit of obtaining a larger quota. We show that for the small member, the endogenously chosen capacity work as an implicit quota enforcement mechanism in good times, leaving very little room for non-compliance. Therefore, we predict that OPEC relies on its quota and punishment system more in bad times than in good times. We also allow the punishment reaction of the cartel's police (Saudi Arabia) to vary in a range (between very weak enforcement to a full enforcement) to study the reaction of investment and production decisions of small members to changes in the stringency of the punishment mechanism. We show that while more stringent punishment results in lower quota violations, it increases investment in capacity expansion.

Our empirical analysis supports these predictions. Using quarterly data for

[3]This monotonic relationship between demand/price and difficulty to support collusion is altered with adding the capacity constraint as in Brock and Scheinkman (1985), Staiger and Wolak (1992a), and Fabra (2003), or by assuming risk-averse members as suggested by Bernhardt and Rastad (2016).

the period of 1995 to 2007, over which OPEC had a stable structure, we build a panel data model to test the statistical significance of the predictions of our theory model. Consistent with the notion that capacity constraint works as an enforcement mechanism in good times, we find that unlike the police that always holds spare capacity to keep the potential punishment credible, other OPEC members become more capacity constrained when oil prices are higher. Next, consistent with OPEC relying on quota system in bad times, we find a negative relationship between the frequency of OPEC meeting and the oil price. Moreover, consistent with a size-dependent punishment mechanism, we provide empirical evidence that shows non-compliance relative to quota levels is more common among the smaller countries especially in bad times. In contrast, consistent with the swing producer hypothesis, we find evidence supporting more non-compliance for the police member in good times.

Our model is built on strong empirical and anecdotal evidences regarding OPEC's behavior. We assume that each country's quota is allocated based on its share of total cartel capacity. This allocation rule is supported by the conversations we have had with people participating in OPEC meetings. Thus, we believe this is very close to the process by which quotas are allocated inside OPEC. Once the quotas are set in a meeting, representatives return to their home countries where they decide how much actually produce, given the "recommended" quotas. As is well stated by OPEC's longstanding Secretary General, Abdalla Salem El-Badri, "It is not a quota as such, but rather a recommendation given to members whom we expect them to take"[4].

The key feature of our theoretical framework is to model the process by which the decisions are made by the cartel members inside and outside of OPEC. We do this by highlighting the role of an important variable, under-emphasized in the literature, as the key variable explaining the compliance behavior of cartel members. More specifically, we emphasize the role of *capacity* as the fundamental heterogeneity among cartel members in three ways. First, empirical and anecdotal evidence support the idea of a direct link between the production capacity of a country and its share of total quota. Second, we introduce convex capacity building costs, which pin down the optimal *maximum* production capacity for each member. We argue that the marginal cost of oil production is much smaller than the initial cost of building capacity,

[4]OPEC Quota A Mere 'Recommendation' To Members, Says Secretary General, Forbes, June 5, 2015

as pointed out by Gault et al. (1999). Therefore, some of results are driven by the capacity limits. Third, larger capacity provides option to produce in high-demand states for the larger member in the production game that follows the quota allocations. Moreover, our empirical analysis highlights the significance of capacity as the most prevalent determinant of cheating behavior.

Three things distinguish our paper from earlier studies on OPEC non-compliance. First, our paper is one of the few studies that look into strategic interactions and cheating behavior inside OPEC, instead of looking at the behavior of OPEC as a whole. Second, unlike the bulk of the literature on OPEC that are merely empirical (e.g., Molchanov (2003)), along with our empirical analysis we provide a theoretical model that is tuned to resemble the OPEC decision-making process and delivers empirically testable predictions. Third, we offer a more universal framework that integrates the effects of general market conditions (e.g., demand fluctuations) and country-specific characteristics (e.g., cost of capacity building) on quota violations into a single framework. This is in contrast to other studies such as Dibooglu and AlGudhea (2007) and Kaufmann et al. (2008) that look at cheating behavior on a country-by-country basis.

The rest of the paper is organized as follows. Section 1 presents the theory model in two stages: the cooperative quota-setting stage, and the non-cooperative production game. Section 2 describes the empirical design and the data. Section 3 illustrates empirical analysis and estimation results, and section 4 concludes.

Theoretical Model

This section introduces a series of stylized models of OPEC members' behavior. The objective of the modeling exercise is to identify major determinants of capacity building and production decisions, which can potentially include a deviation from the assigned quota (among other factors). The model highlights trade-offs in members investment and production decisions.

In its most complete form, our theoretical model consists of a four-stage game. In the first stage countries *endogenously* choose the optimal level of capacity investment. In the second stage production quotas are allocated (taking into accounted installed capacities). Countries make decisions regarding actual production (and possible deviations from quotas) in the third stage. We do not assume that the third stage's production plans fully follows the allocated quotas because *endogenous* deviations from quotas are allowed. The forth stage includes monitoring the actual performance of members by the large producer (aka the police) and potentially implementing a random punishment strategy. In our model capacity levels, quota allocation, actual production (hence, quota violations), and punishment are all endogenous.

Key Components and Assumptions of the Model

OPEC is a sophisticated and non-transparent organization. There are a lot of politics involved in its members bargaining, agreements, and decision making. We recognize that a stylized model, such as the one presented in this paper, can not fully capture all aspects of OPEC's dynamics. Moreover, given the size limitations of the current paper we have to make certain assumptions and take a pragmatic approach for building and solving the model in order to come up with empirically testable hypothesis.

focus on the key message related to the empirical section of the paper.

Objectives and Functions of OPEC OPEC acts as a cartel maximizing the total profit of its members[5]. OPEC's objective in reality is a mix of profit-maximizing and price-stabilizing goals. Especially in recent years, OPEC has tried to keep the crude oil prices around a reasonable bound (appx \$50-\$60) to reduce the incentives of the demand side to invest in alternative forms of energy[6]. We model OPEC only as a profit-maximizing cartel[7].

[5]There are alternative views that reject OPEC's cartel behavior and consider it as an information aggregation and a consulting institution. We focus on the cartel function and abstract from this alternative view.

[6]It is questionable if OPEC indeed could achieve its price stablization goal.

[7]The counter-cyclical price policing strategy is also consistent with a view that cartels try to endogenously reduce price in the high demand periods to mitigate the incentives of its own members to leave (Rotemberg and Saloner (1986b), Knittel and Lepore (2010)).

Cartel Members The cartel consists of a large producer (aka the police, who monitors other members production and their deviations from quotas) and a fringe of small producers. We assume the production capacity of the police is K_1 and the *aggregate* production capacity of all fringe members together is K_2. To keep the algebra simple we solve the game between one large producer and one *representative* small player. Thus, we abstract from strategic interactions between small producers. We only model the endogenous capacity investment, quota allocation, and production decisions of the representative small producer. The production capacity of the police is exogenous given.

Cartel's Police In order to preserve the stability of a cartel, there should be some credible punishment strategies vis-a-vis non-complying members. We assume Saudi Arabia is the *de facto* police of the OPEC cartel. No other member has the production or fiscal capacity to punish defying members. We assume the police holds an exogenously-specified level of *excess capacity* beyond the levels by any other OPEC member to support a credible threat against deviations from the cartel allocations[8].

Demand The demand for crude oil has a standard linear functional form: $P = \theta - \gamma Q$. Here θ is the demand shift parameter and γ is the fixed parameter representing the sensitivity of prices to quantities. We will first solve a deterministic model and then introduce a stochastic demand. The stochastic demand shift parameter includes two states which follow a discrete i.i.d distribution of θ_H (high demand) and θ_L (low demand), with probabilities of p_H and p_L, respectively.

Capacity Building Costs OPEC members face *convex* capacity building costs. This assumption is in line with the order of extraction hypothesis in the resource economics literature (Amigues et al. (1998)). According to the order of extraction theory resource holders (OPEC members here) first extract the most efficient resource base and then move to fields with higher costs (less

[8]The optimal level of the excess capacity for the police can be characterized by considering the option value of producing additional units in the high demand state along with the value of offering a credible threat. The full characterization is beyond the scope of the current paper. Thus, we abstract from this feature and leave it to the future research.

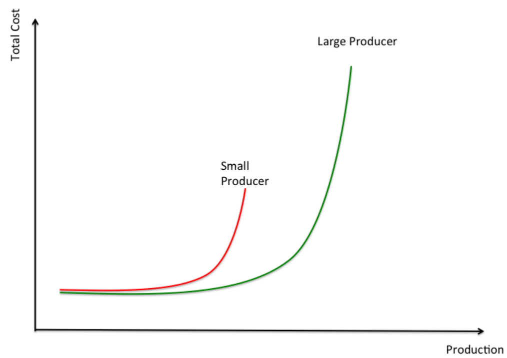

Figure 1: Convex investment costs of small and large producers. The large produce can build a larger level of capacity with the same level of capacity investment.

efficient fields, or even types of reserves such as offshore, deep water, etc) if the demand is high enough.

We assume that the marginal cost of production is zero[9]. The major component of the total cost function is the fixed cost of building capacity. The capacity cost is modeled as a stream of annualized payments (i.e., all equipment are perpetually leased). The limiting case of the convex capacity cost is a hard capacity constraint, which is equivalent to a locally infinite marginal cost (i.e. a vertical part of the cost curve). However, countries are different in the level and curvature of their cost curves. This heterogeneity in the size and quality of reserves distinguishes a large producer such as Saudi Arabia from a small producer like Gabon or Qatar. Figure 1 shows a stylized view of this model.

We will discuss that under an efficient planner solution, the country with a lower capacity building costs (i.e. cost curve to the right) will optimally build a larger capacity and receive a higher production quota.

[9]OPEC members are typically low cost producers. Their marginal production cost is much smaller than the initial cost of building capacity, as pointed out by Gault et al. (1999). Thus, the zero marginal cost assumption, relative to capacity investment cost, is not a major abstraction from the reality.

Allocation Rule of OPEC The theoretical foundations for the internal allocation of aggregate production to cartel members is a matter of debate (see Cave and Salant (1995)). We assume that the cartel uses a two-layer allocation rule. In the strategic layer the allocation rule minimizes the cartel's production cost (surplus maximization strategy) for capacity investment. In the production layer (when capacities are not binding due to a low demand) the rule suggests an allocation proportional to existing capacities. Anecdotal evidences suggest that the established capacity is a key factor in assigning initial quotas to OPEC members. In the strategic version of the model we will discuss how a proportional allocation rule incentivizes member countries to build excess capacity. Sandrea (2003) suggests that the following eight factors have been used in the initial stage: reserves, production capacity, historical production share, domestic oil consumption, production costs, population, dependence on oil exports, and external debt. Reynolds and Pippenger (2010) use the case of Venezuela and show that in the long-run production capacity and actual productions of countries Granger cause their OPEC quotas.

Punishment In reality one observes constant deviations from assigned quotas for multiple reasons. First, countries are typically not content with the officially assigned quota and prefer to producer more. Second, it is both difficult and costly for OPEC to constantly punish non-complying members.

The problem with the standard definition of punishment strategy in cartels (i.e the strategy of flooding the market with overproduction and reducing the price) is that it not only punishes the non-compliant member but also other "innocent" members (including the police itself)[10].

This causes the punishment strategy to be ex-post subgame perfect but not negotiation-proof. We are not aware of other credible punishment mechanisms (such as trade embargoes), which have been used by OPEC members. In all famous historical incidents (e.g. the oil price war of 1986) the punishment came through overproduction by Saudi Arabia.

In order to have a credible punishment strategy, cartel members need to have enough spare capacity, be able to identify the deviant member (Green and Porter (1984a)), and not be jammed by possible demand uncertainty noises.

[10]The relative cost to the police can potentially be less than to other members because the policy increases production and this partially offsets the impact of a lower price.

Punishment Technology Punishment in our model is random and the probability of punishment (conditional on cheating) is $m \in [0, 1]$.

If a member deviates from the assign quota by Δq, a monitoring technology with strength m will trigger the punishment strategy by the probability of $m(\Delta q)$. A size-dependent punishment technology will make the probability of punishment a function of the magnitude of deviation; whereas, a size-independent technology will react to all deviations with a constant probability. The random punishment can be interpreted as a mixed strategy played by the police.

When $m \to 1$ the punishment is very effective, and when $m \to 0$ the ability or the incentive to punish is very weak. When $m \to 0$ the non-police members are confident that the cartel will never punish their quota violations.

A partially effective punishment technology can be the result of several frictions including the inability to accurately observe members' actual production, the lack of a political will to punish the deviating member, or internal disagreements within the OPEC. We summarize all of these forces using the single parameter m.

Four Versions of the Model

One can consider a step-by-step development of cartel models in the following order:

1. A planer's[11] solution for a cartel with a convex capacity-building costs under certainty
2. A planer's solution for a cartel with a convex capacity-building costs under uncertainty
3. A Cartel with quota allocation and full enforcement of quotas
4. A Cartel with quota allocation and endogenous punishment for quota violation

[11] We do not call the planner in the model a *social-planner* because a social-planner maximizes the "total" social surplus (some of consumers' and producers' surpluses); whereas, the OPEC planner just maximizes the joint profit of cartel's members (i.e. the production side).

Model 1: Efficient Allocation under Certainty

The benchmark case is a two-country cartel problem, with asymmetric convex capacity building cost curves $c(K_i) = \frac{1}{2}\phi_i K_i^2$, where ϕ_i is the coefficient of resource efficiency and K_i is the installed capacity. For countries with a larger resource base, ϕ is a smaller parameter.

We also assume a full enforcement of OPEC quota rules (through a planner). We first introduce this benchmark to provide the key intuitions from an efficient allocation and then will move to a more realistic case with capacity limits and incentives to cheat.

We assume that the two countries choose the optimal production capacity based on an efficient internal allocation mechanism. Since the capacity has not been built one can treat the annualized cost of capacity like *marginal* cost. Note that members of the cartel are bound by their maximum capacity; thus, even if the *operational marginal* cost is zero, they will not be able to deviate upward from the efficient solution. Moreover, no party has an incentive to build more capacity than what we recommend here because the marginal cost of building an additional unit of capacity exceeds the marginal benefits (taking the behavior of the other members as given).

The optimal cartel solution is:

$$(1) \qquad \begin{cases} \max_{K_1,K_2} P(q_1+q_2)(q_1+q_2) - c_1(K_1) - c_2(K_2) \\ \text{s.t.} \\ q_1 \leqslant K_1, q_2 \leqslant K_2 \end{cases}$$

where $c(K_1)$ and $c(K_2)$ are the total cost of building K_1 and K_2 units of capacity.

The cartel problem in this version of the model is the same as the standard multi-plant monopolist problem, which is maximizing the joint profit of all plants. The FOCs of the cartel problem result in the following system of equation, in which the marginal cost of each member is equal to the marginal revenue of the cartel:

$$
(2) \qquad \begin{cases} c_1'(K_1) = P'(K_1+K_2)(K_1+K_2) + P(K_1+K_2) \\ c_2'(K_2) = P'(K_1+K_2)(K_1+K_2) + P(K_1+K_2) \end{cases}
$$

Given that the right hand sides are equal, we immediately observe that $c'(K_1) = c'(K_2) \Rightarrow \frac{K_1}{K_2} = \frac{\phi_2}{\phi_1}$. Since capacity is binding in the model, we also note that $\phi_1 > \phi_2 \Rightarrow K_2 > K_1$, implying that the member with a less efficient cost curve (represented by a higher ϕ) will receive a lower production quota.

Plugging in the specific functional forms for c_i and P:

$$
(3) \qquad \begin{cases} q_1 = K_1 = \frac{X\phi_2}{\phi_1\phi_2 + 2\gamma(\phi_1+\phi_2)} \\ q_2 = K_2 = \frac{X\phi_1}{\phi_1\phi_2 + 2\gamma(\phi_1+\phi_2)} \end{cases}
$$

The solution to the system of equation characterizes an efficient internal allocation, while maximizing the total profit of the cartel.

Model 2: Efficient Allocation under Uncertainty

When the uncertainty is introduced, the optimization problem of the planner consists of two stages. In the first stage (before the resolution of uncertainty) member countries choose their optimal capacity levels (by considering the expected value of capacity over different states). In the second stage, and given the fixed capacity installed in the first stage, countries produce according to a profit-maximizing program.

The demand states are distributed according to θ_L, p_L and θ_H, p_H, where p_L, p_H are the probability of low and high states, respectively. We denote the production at the high and low states by q_i^H and q_i^L.

A key point to note is that the choice of capacity is *prior* to the resolution of uncertainty regarding the demand state. Therefore, the country maximizes based on the expectation of profits. However, production decisions are taken *after* observing the realized demand state (i.e after the uncertainty regarding the demand is resolved.)

Since the production in the low demand state will always be below the capacity, the marginal value (i.e. shadow price) of capacity at this state is zero. Therefore, the choice of capacity will be merely driven by the high state optimal production plan. In other words, there are four equations for $q_{L,1}, q_{H,1}, q_{L,2}, q_{H,2}$, out of which two are binding (in the H-state) and two are always slack (in the L-state). Therefore, $K_i = q_i^H$

The optimization problem of the planner looks very similar to the previous case, except that the marginal value of capacity is only relevant in the high state[12].

$$(4) \qquad \begin{cases} c_1^{'}(K_1) = p_H[P^{'}(K_1+K_2)(K_1+K_2) + P(K_1+K_2)] \\ c_2^{'}(K_2) = p_H[P^{'}(K_1+K_2)(K_1+K_2) + P(K_1+K_2)] \end{cases}$$

Solving the system of equations using the assumed functional forms one gets[13]:

$$(5) \qquad \begin{cases} q_1^H = K_1 = \frac{p_H \phi_2 \theta_H}{\phi_1\phi_2 + 2p_H\gamma(\phi_1+\phi_2)} \\ q_2^H = K_2 = \frac{p_H \phi_1 \theta_H}{\phi_1\phi_2 + 2p_H\gamma(\phi_1+\phi_2)} \end{cases}$$

The equations above pin down the binding capacity (and hence production plans) in the H state. Note that when $p_H = 1$ the solution is the same as the Model 1.

We also need to characterize the optimal production in the L state. In state L, after observing the realized state of demand and conditional on actual capacities OPEC decides on the aggregate supply and then allocates quotas to members. Economic theory has little to say about the internal allocation in the low demand state when members have excess capacities (especially when the marginal cost of production is zero). In practice, OPEC uses a

[12]In models of continuous state space the marginal value of capacity is calculated as a probability-weighted integral for all states with a binding capacity (e.g. Grimm and Zoettl (2013)). Since we only have two states the integral collapse to a single equation.

[13]If p_H is sufficiently small, the expected marginal benefit of capacity in the H-state might be too small. In this case, it is possible to get an optimal level of capacity, which is also binding in the L state. We assume this is not the case and the p_H is sufficiently large to result in a non-binding capacity in the L state.

combination of economic, social, and political factors to determine the quotas. We assume the allocation is based on installed capacities. The cartel first finds the monopoly production quantity. Since we assumed marginal cost is equal to zero the cartel just maximizes revenues[14].

$$\max_{Q} P(Q,X)Q \tag{6}$$

The F.O.C of the monopoly problem:

$$\begin{cases} P^{'}(Q,X)Q + P(Q,X) = 0 \Rightarrow Q_L^*(X) = \frac{\theta_L}{2\gamma} \\ \Rightarrow P(Q_L^*) = \theta_L - \gamma Q_L^* = \frac{X_L}{2} \end{cases} \tag{7}$$

After determining the total production $Q^*(X)$ OPEC allocates quota to each country.

Two types of deviation can be identified: an operational deviation by producing more than assigned quota (the focus of this paper) and an investment deviation by building too much initial capacity[15]. The operational deviation is more frequently observed in practice. Given the nature of our motivating empirical observations, our theoretical focus is also on this type of deviations. Consequently, we only model the punishment mechanism for the operational deviations and abstract from modeling the punishments in response to investment deviations.

Model 3: Strategic Incentives for Capacity Investment

We now consider the incentives of OPEC members to build excess capacity (ex-ante) in order to gain a larger quota in the low demand state. Our assumption of the strategic effect of capacity is in line with the intuition offered by some recent papers (e.g. Fagart (2016)).

[14]Note that the capacity investment cost is sunk and does not enter the optimal production equation in the L state.

[15]In 1997 Venezuela followed an ambitious plan to significantly increase its production capacity. In a reaction to this move Saudi Arabia increased its production and cause the oil price to drop to its historical low level of $10 per barrel.

We assume a reduced-form proportional quota allocation rule, which assigns quota based on the relative share of countries' *installed capacities* . If the theoretical share of country i is greater than its actual capacity, the quota will be capped at capacity.

Combining cartel total production and allocation rules:

$$\bar{q}_i(\theta) = \begin{cases} \frac{\theta}{2\gamma} \frac{K_i}{\sum K_j} & \text{if } \frac{\theta}{2\gamma} \frac{K_i}{\sum K_j} \leqslant K_i \\ K_i & \text{otherwise} \end{cases} \tag{8}$$

Note that OPEC members' individual *capacities* do not affect the optimal level of cartel's production (as long as the total capacity of the cartel is large enough). Moreover, the proportional rule does not guarantee that members will get quotas equal to their production capacities. As long as the aggregate capacity of OPEC is larger than the optimal cartel production, some members will receive quotas below their capacities[16].

The objective function of the nested optimization problem is

$$\max_{K_i} [\mathbb{E}(\max_{q_i \leqslant \min\{K_i, \bar{q}_i\}} \pi^*(K, q_i)) - c(K)] \tag{9}$$

where players choose the capacity in the first stage by considering its effect both in the low and the high demand states (and taking the capacity strategy of others as given.) The new feature of this version of the model is taking into account the strategic effect of capacity on the assigned quota for the low state (i.e. bargain for a higher quota)

Under this structure, and noting that $q_1^H = K_1$, the key equation to pin down the endogenous choice of capacity by a small cartel member is[17]:

[16] The outcome of this process will be like a prisoners' dilemma case, in which every country builds excess capacity but does not benefit from it. However, given the strategy of other players it is always optimal to build the excess capacity.

[17] As a first-order approximation we ignore the effect of increased capacity on the production of the H-state and assume the production constraints is still binding in that state.

(10)

$$\underbrace{c_1^{'}(K_1)}_{\text{Marginal Cost of Capacity}} = \underbrace{p_H[P^{'}(q_1^H + q_2^H)q_1^H + P(q_1^H + q_2^H)]}_{\text{Marginal production value in the H state}} + \underbrace{p_L[\frac{\partial \pi_L}{\partial q_1^L}\frac{\partial q_1^L}{\partial K_1}]}_{\text{Marginal quota value in the L state}}$$

We derive the value of the expression in the second bracket in the next subsection.

Marginal Quota Effect of Capacity The proposed allocation rule allows us to derive the marginal "allocation value" of a unit of capacity. Assuming that the marginal effect on the total capacity of cartel is negligible, the marginal allocation value can be defined as follows:

$$q_i(X) = \begin{cases} Q_L^* \frac{K_i}{\sum K_j} & \text{L state} \\ K_i & \text{H state} \end{cases} \tag{11}$$

Denoting the assigned quota in the low demand state by q_i^L and assuming $K_i << \sum K_i \Rightarrow \frac{K_i}{\sum_i K_i} \approx= 0$, one gets:

$$\frac{\partial q_i^L}{\partial K_i} = Q^* \frac{\sum_i K_i - K_i}{(\sum_i K_i)^2} \approx \frac{Q^*}{\sum K_i} \tag{12}$$

We see that the larger the total capacity of OPEC members, the weaker the quota allocation effect of additional unit of capacity for an individual member will be. Note that Q^* is the total production of cartel in the low state and $\sum K_i$ is the total production in the high state. Thus, the closer the production in the two states are, the stronger the incentive for bargaining would be.

The marginal value of capacity in the low state then can be calculated by:

$$\frac{\partial \pi_i^L}{\partial K_i} = \frac{\partial \pi_i}{\partial q_i^L}\frac{\partial q_i^L}{\partial K_i} \approx \underbrace{\pi^{'}(q_i^L)}_{\text{Marginal Profit}} * \underbrace{[\frac{Q_L^*}{\sum K_i}]}_{\text{Changes in the Production}} = \frac{X_L^2}{4\gamma}\frac{1}{\sum K_i} \tag{13}$$

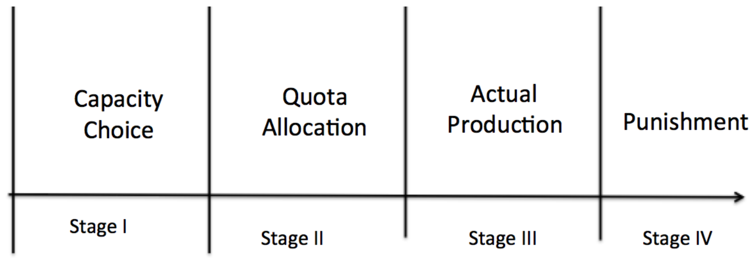

Figure 2: Stages of the Game

Equation 13 suggests that when the total installed capacity of the cartel is large, when the demand in the low state is too weak, and when the sensitivity of price to quantity is high, the incentives for building excess capacity (for quota bargaining purposes) would become smaller. The small member of the cartel compares the marginal value of capacity with the marginal cost of building (an almost unused) unit of excess capacity. If capacity costs are relatively low, the country may build and keep spare capacity for the bargaining purposes; however, if capacity investment costs are high, the country may give up building excess capacity.

Model 4: Strategic Interaction

In the final version of the model the full quota compliance assumption is relaxed. OPEC members can potentially deviate from the allocated quota; however, they are subject to possible punishment by the policy. The game consists of four major stages as described in Figure 2. The small producers use backward induction and consider the effect of equilibrium outcomes in later stages on their current decisions. The first stages of the time-line occur only once; however, the production and punishment stages summarize a repeated game.

Capacity Stage

As before, small producers take into account the marginal cost of building one additional unit of capacity versus its marginal benefits. The incentives to build capacity are twofold: first, the option value of producing higher in good time, and second, the strategic effect of capacity on the allocated quota. We assume that the first effect is small and abstract from it to focus on the second effect, which is the key focus of this paper.

Actual Production Stage

Given K_i, q_i the individual small (i.e. non-police) country i solves a constrained optimization problem under uncertainty. Note that since the capacity has been already built and there are no operational costs (as we assumed before), the *marginal* cost of production is zero.

The member country considers the following two cases:[18]:

(14)

$$\text{Pay-off} = \begin{cases} \text{Compliance} : \pi(\overline{q}, s) + \frac{1}{1+r}\mathbb{E}[V(s)] \\ \text{Violation} : \pi(q^*, s) + \frac{1}{1+r}\{\underbrace{(1-m)\mathbb{E}[V(s)]}_{\text{No Punishment}} + \underbrace{m\mathbb{E}[V^C(s)]}_{\text{Punishment}}\} \end{cases}$$

The first line refers to the compliance strategy. If the country complies, it earns the profit associated with the assigned quota and the repeated game continues. The game in this case is basically the repeated version of the one-shot game of comppliance. However, if the country violates the quota (second line), it may trigger punishment. The country enjoys a higher profit at the current period by accepting the risk of punishment, which materializes in the form of switching to a Cournot value function. The details of the Cournot value function is presented in the Appendix.

The violation branch again creates two sub-scenarios ahead of the cartel member: with probability $(1 - m)$ no punishment will occur and the member continues with the same value function. However, with the probability m the police will punish the non-compliant behavior and the game continues in a non-cooperative (Cournot) regime.

After simplifying the algebra, it turns out that the small member has incentive to deviate if the following condition holds:

[18]The demand state switches between L and H, thus, the country considers the expected value function given the current state of the demand.

$$\underbrace{\pi(q^*) - \pi(\overline{q})}_{\text{One-Shot Benefit}} > \underbrace{m * (\mathbb{E}[V^C] - \mathbb{E}[V])}_{\text{Loss of Value Function}} \tag{15}$$

Any punishment strategy such that the benefit of a one-shot deviation smaller than the lost value of cooperation will induce members to comply[19].

From Equation 15 we can see that the cartel is more stable when the immediate benefit of deviation is smaller or when the expected cost of punishment is higher.

Proposition 0.1. *The incentive to cheat is higher in the H-state compared to the L-state.*

Proof. The R.H.S of Equation 15 is independent of the current state (note that the demand follows an i.i.d regime.). However, the LHS is a function of the current state. In the high-price state the marginal value of an additional unit of production is higher than a low-price state (which is the standard result of Rotemberg and Saloner (1986a)). □

Corollary 1 Small producers face binding constraints in the high demand state. Thus, the potential magnitude of deviation for a small producer is larger in the L-state compared to the H state. Capacity constraints work as an implicit enforcement mechanism.

In the H state, small producers typically will have a small excess capacity to unilaterally expand their production. Thus, the immediate value of deviation is limited. However, in the low demand state there is plenty of excess capacity to be used for quota violation.

Proposition 0.2. *If the size of the strategic excess capacity is small, punishment is only required in the low state.*

Proof. Small members do not have major excess capacity in the H state. Therefore, the total benefit of quota violation is small. However, they have significant excess capacity in the L-state and the potential value of a one-shot

[19]We know from the folks theorem of repeated games that there are many strategies in the game that support a Nash equilibrium.

deviation can be significant[20]. In summary, in the high state the capacity constraints function as the enforcement mechanism; whereas, in the low state quota needs to be enforced by the cartel.

Even if members have some excess capacity to deviate in the H state, such a deviation is not too costly for the cartel. On contrary, a deviation during the L time can be very costly to everybody. Thus, one can expect that during the low demand regime more quota-related activities (e.g., quota-setting, political bargains, meetings, etc) should be observed. □

Corollary 2 The frequency of OPEC meetings and quota adjustment should be negatively associated with the level of crude oil price.

Occasional Punishment

Here we make an assumption to be able to finalize the model in a fashion close to the real world. In a classic cartel model with a credible punishment strategy, no shock or noise, and fully rational players no quota violation or punishment will be observed on the *equilibrium path*. Punishment is only an off-equilibrium phenomena. However, in reality a few episodes of punishment within OPEC have been observed. Green and Porter (1984b) introduce a model of collusion under imperfect information, in which some punishment is observed. Authors show that when the "observed" market price is low, cartel members may follow a non-cooperative Cournot behavior and then revert back to a collusion behavior after a few periods. This model of behavior fits well with the observed behavior of OPEC cartel members.

We justify occasional (but infrequent) punishment and some episodes of quota violations by shocks to the ability/willingness of the police to punish deviations. The following trigger strategy by the police constitutes a Nash equilibrium in states when the police is able to punish: *if a deviating member is detected the game will switch to Cournot for a few periods and then will revert back to cooperative strategy.*.

[20] In this paper we do not impose any behavioral assumption on the marginal value of extra dollar in the high and low states. In reality, OPEC governments may be under fiscal pressure during low demand states; thus, the *marginal value* of an extra dollar of export revenue could be higher in bad times resulting in a stronger temptation for deviation or even a myopic behavior.

A plausible scenario in this case is a few periods of quota violations (until the cheating is detected) followed by a severe punishment and then reversion to the cooperation.

When the punishment technology is very weak (i.e the police is incapable in the majority of periods), $m \to 0$ and as a result the game switches to a traditional Cournot competition. On the other hand, if the punishment is extremely effective (i.e. a tiny deviation will be caught and punished in every single period), we will have $m \to 1$ and the game converges to a textbook cartel game in which no deviation will be observed. The more interesting case is the intermediate value of $0 < m < 1$, meaning that while the OPEC member is worried about possible punishment, they still may take the gamble and cheat because punishment is not certain.

Proposition 0.3. *When the punishment becomes more effective, violations decreases but capacity investment increase. Average spare capacity increases with the effectiveness of the monitoring technology.*

The intuition behind Proposition 0.3 is that a higher likelihood of punishment for non-compliance will reduce the incentive to cheat in the second stage; on the other hand, it shifts the strategic incentives to the first stage to gain a higher allocated share. This, however, will not necessarily result in very large capacities because each members compares the marginal benefit of capacity (i.e. the probability of a marginal unit of quota) to the investment cost of capacity.

Size-Dependent Punishment So far we assume that the punishment function $m()$ does not depend on the aggregate size of deviations. The police may tolerate small deviations (in absolute terms) but react to large deviations.

The intuition over the optimization problem suggests that deviations are smaller when the punishment is size-dependent.

One can also easily observe that if the punishment probability depends on the aggregate deviation of all members, small members will have a larger proportional deviations. The same percentage of deviation by a small and a larger producer will have different absolute effects on the market. When the the likelihood of punishment depends on the size of deviation, each member considers the marginal effect of its deviation on the probability of being punished. This acts as a cost and reduces the optimal level of production for the one-shot deviation.

Summary of Theoretical Insights

We summarize the lessons learned through the modeling exercise.

1. Deviation is more likely in the high price state. However, hard capacity constraints might function as a limit to this temptation.
2. OPEC needs to work harder to sustain the cartel quota system during low demand state.
3. Small producers have higher incentives to deviate. They can make larger deviations (proportional to their capacity) before triggering a punishment strategy.
4. A more stringer punishment strategy increases incentives for building excess capacity and gain a larger quota.

Empirical Analysis

In this section, we review our data collection, summary statistics of the sample and the empirical design we use to test model predictions and additional empirical insights.

Data

We use country level quarterly data for production, capacity and quota allocations for OPEC members excluding Iraq between 1995 and 2007. We obtain the data from three sources: quota allocations and crude oil prices come from the OPEC statistical bulletin, production and capacity data are obtained from the Energy Information Agency (EIA) website, and reserves data comes from BP Statistical Review of World Energy website. We intentionally do not use OPEC data on production given that OPEC members have a tendency to under-report their production levels to hide quota violations. A comparison between OPEC vs. EIA confirms this conjecture[21]. Data availability dictates the time span of our sample. 1995 is the first year that EIA provided the quarterly production data for OPEC members, and 2007 is the last year that EIA published members' production capacity data on their website. Given that EIA does not provide capacity data for years before 2001, we complement this sample with the data used in Kaufmann et al. (2008)[22].

Moreover, OPEC as an organization had a very stable structure with no entry or exit during this time window (Ecuador exited OPEC before 1995, and Indonesia exited after 2007.) We follow the common practice of dropping Iraq from our sample due to the Persian Gulf War and trade sanctions that limited its interactions with OPEC during this period. This leaves us with a sample of 10 OPEC members: Algeria, Indonesia, Iran, Kuwait, Libya, Nigeria, Qatar, Saudi Arabia, United Arab Emirates (U.A.E.), and Venezuela.

Variables	**N**	**Mean**	**SD**	**Min**	**Q1**	**Median**	**Q3**	**Max**
Capacity	519	2.96	2.71	0.53	1.44	2.31	2.83	11.40
Quota	519	2.47	2.08	0.38	1.28	2.00	2.72	9.10
Production	519	2.59	2.15	0.40	1.36	2.05	2.71	9.60
Spare Capacity	519	0.50	0.77	-1.30	0.14	0.31	0.58	3.89
Non-Compliance	519	7.7%	13.6%	-46.8%	1.4%	5.0%	10.9%	74.5%
Price (nominal $)	519	31.20	17.81	10.98	18.27	25.58	40.01	85.07
Price (real 2001$)	519	25.60	11.51	10.16	16.68	23.48	30.66	58.69
Meeting Frequency	135	0.36	0.58	0	0	0	1	3

Table 1: **Summary Statistic of Variables.** This table presents the summary statistics of the variables for OPEC members except Iraq. Non-compliance is calculated relative to the quota level. *Spare Capacity* is the difference between capacity and quota. Capacity, Quota, Production and Spare Capacity are in million barrel per day (bpd). Meeting Frequency is the number of OPEC meetings in a quarter that resulted in a quota change. The data is reported quarterly between 1995 and 2007, except for Meeting Frequency which is between 1983 and 2016.

Table 1 presents summary statistics for this sample between 1995 and 2007. As can been seen from the Table, quota non-compliance, ranges from -46.8%

[21]Total OPEC production reported by OPEC was consistently lower than what is reported by EIA between 1983 and 2003, except in 1992.

[22]Kaufmann et al (2008) estimate quarterly capacities by interpolating annual capacity data provided to them by Erik Kriel of the US Department of Energy.

Countries	**N**	**Mean**	**SD**	**Min**	**Q1**	**Median**	**Q3**	**Max**
Algeria	52	0.21	0.23	-0.05	0.03	0.11	0.50	0.61
Indonesia	52	0.11	0.17	-0.25	-0.03	0.18	0.23	0.34
Iran	51	0.02	0.05	-0.09	-0.01	0.01	0.05	0.13
Kuwait	52	0.06	0.06	-0.01	0.02	0.04	0.11	0.26
Libya	52	0.07	0.05	-0.04	0.03	0.07	0.11	0.17
Nigeria	52	0.04	0.05	-0.05	0.00	0.04	0.08	0.15
Qatar	52	0.15	0.15	-0.01	0.06	0.10	0.16	0.75
Saudi Arabia	52	0.05	0.04	-0.01	0.02	0.04	0.06	0.15
UAE	52	0.03	0.03	-0.03	0.01	0.03	0.05	0.10
Venezuela	52	0.03	0.19	-0.47	-0.10	0.05	0.13	0.43
Total	519	0.08	0.14	-0.47	0.01	0.05	0.11	0.75

Table 2: **Summary Statistic of Non-compliance by each OPEC Member.** This table presents the summary statistics of non-compliance relative to quota for OPEC countries. Data is on quarterly basis between 1995 and 2007.

for Venezuela to 74.5% for Qatar, with an average of 7.7% and a median of 5%. Note that here a negative number basically means producing below the quota or complying with the quota allocations, and a positive number means a quota violation. We also observe that (ex-ante) spare capacity for some members occasionally is a negative number. This means that their production quota is set above their capacities. We drop those observations from our sample because non-compliance is virtually impossible for those countries. This results in losing 54 observations that are mainly related to Indonesia, Iran and Venezuela. Negative spare capacity for those countries is usually the result of a shrinking capacity while quota is fixed.

Table 2 presents summary statistics for non-compliance by country between 1995 and 2007. Looking at the median or mean column of Table 2, we can see that Algeria, Qatar and Indonesia are on average the most non-compliant members. In contrast, Iran, UAE and Venezuela are on average the most compliant members. This immediately suggests an inverse pattern between non-compliance and producer size, measured by capacity or production. To further investigate this conjecture we depicted the average non-compliance vs. average production capacity in a logarithm scale over the sample period

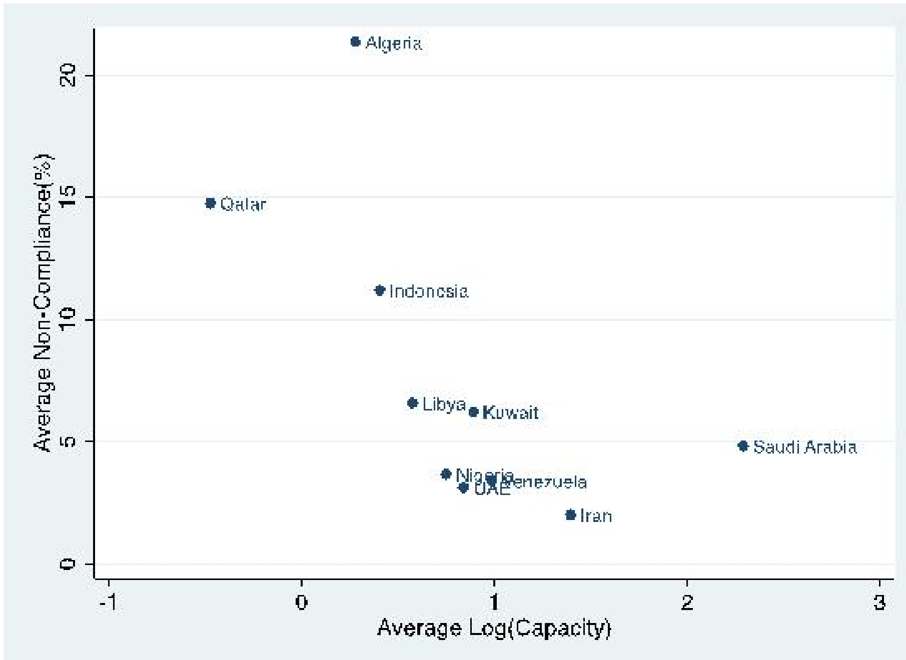

Figure 3: **Non-compliance vs. Capacity Size.** This figure presents average relative non-compliance for OPEC members vs. average logarithm of production capacity in million bpd between 1995 and 2007.

in Figure 3. It is clear from this figure that smaller producers like Algeria and Qatar have higher levels of non-compliance than larger producers such as Iran, Venezuela and UAE. We revisit this conjecture more formally later to see if it is a robust pattern in a multiple regression framework.

Empirical Model and Estimation Results

To test empirically the predictions by our theoretical model, we investigate the production or compliance behavior of the individual OPEC members using a Logit, a time-series and a panel data model in this section. To investigate if capacity constraint work as an enforcement mechanism in H state, we first study the relationship between the likelihood of being capacity constrained and the level of oil prices. Next, we examine if OPEC relies on quota and punishment system as an enforcement mechanism in L-state by evaluating the relationship between the frequency of OPEC meetings and the oil prices. Finally, we characterize the pattern of non-compliance (in both absolute and relative terms) as a function of country size and market conditions to see if we can find supportive evidence for size-dependent punishment strategy or the swing producer hypothesis.

Capacity Constraint

Does OPEC rely on quota system in both L and H states? Does capacity constraint work an implicit enforcement mechanism in either state? Proposition 0.2 and Corollary 1 predict that in H-state the capacity constraint works as an enforcement mechanism for the small members to ensure that they do not overproduce their quota levels[23]. This is in opposite to the behavior of the police country who keeps a large spare capacity strategically in both L and H states to make sure that punishment is credible in L-state and it can overproduce its quota if needed in H-state when everyone else is more likely to be capacity constrained.

To see if small members are in fact more capacity constrained in H-state, we illustrate the frequency of producing at capacity for each member as a function of quartiles of the real oil price[24] in Figure 4. Consistent with Corollary 1, we find that while for almost all OPEC members the likelihood of being capacity constrained rises with oil price, for the police/large country this likelihood does not vary with oil price but is always zero.

To formally test this conjecture, we use the following Logit model:

$$\begin{aligned} Constrained_{it} &= \beta_0 + \beta_1 Price_t + \beta_2 Capacity_{it} \\ &\quad + \beta_3 Price_t \times Capacity_{it} + \varepsilon_{it} \end{aligned} \tag{16}$$

[23]Of course, this is true as long as small members do not have large spare capacities and therefore their quota levels are very close to their capacity levels.

[24]Here we assume that oil price signals the demand state. One may have potential concerns regarding the endogeneity of oil price because of reverse causality. However, the fact that we find a positive relationship in Figure 4 works in our favor as reverse causality would suggest a non-positive relationship.

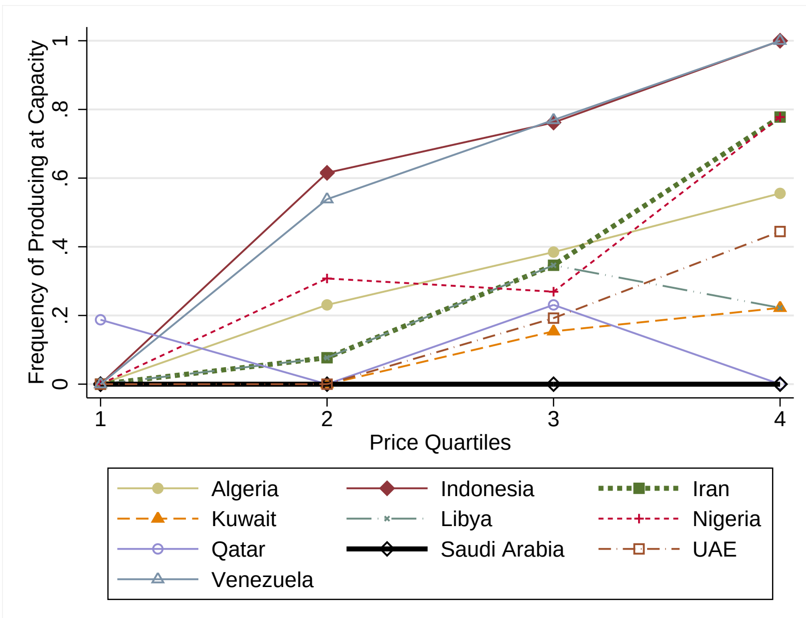

Figure 4: **Likelihood of Producing at Capacity vs. Oil Price.**This figure presents average frequency of producing at capacity for each OPEC member as a function of oil price between 1995 and 2007. Horizontal axis shows 4 bins (quarterlies) of the real oil price in 2001 dollars. The vertical axis shows the average number of times an OPEC member's production level equals or exceeds its capacity level.

where $Constrained_{it}$ is an indicator variable for country i in period t that switches on when it produces at its capacity level, and a member's size is measured by $Capacity$, which is a member's production capacity in million bpd.

Estimation results are presented in Table 3. Two immediate observations are clear from Table 3: the likelihood of being capacity constrained is higher (i) for smaller countries, and (ii) when the oil prices are high. However the results in column (4) suggest that the coefficient for the interaction term, β_3, is not statistically significant. Overall, consistent with Corollary 1, this result confirms that the capacity constraint works as an enforcement mechanism especially for the smaller members in H-state.

Dependent Variable: Indicator for Producing at Capacity

	(1)	(2)	(3)	(4)
$Price_t$	0.056***		0.061***	0.053***
	(8.51)		(8.66)	(4.65)
$Capacity_t$		-0.321***	-0.366***	-0.513**
		(-4.38)	(-4.64)	(-2.57)
$Price_t \times Capacity_t$				0.003
				(0.83)
$Constant$	-2.984***	-0.358**	-2.237***	-1.924***
	(-11.92)	(-1.97)	(-7.68)	(-4.01)
$PseudoR^2$	0.120	0.052	0.179	0.180
$Observations$	628	628	628	628

Table 3: **Who and When Is Capacity Constrained?** This table presents regression results that describes which countries are more likely to produce at their capacity and when. The model specification is given in equation (16). The dependent variable is an indicator variable for producing at capacity . *Price* is the real oil price in 2001 dollars; *Capacity* is a member's production capacity in million bpd. The data is on a quarterly basis between 1995 and 2007. Numbers in parenthesis are t-statistics, and significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

OPEC Meetings Frequency

In previous section we observed that smaller members are more likely to capacity constrained in H-state. Next, we ask if this observation implies that OPEC relies on capacity constraint as an implicit enforcement mechanism in H-state? If so, does this imply that OPEC relies on quota system more in L-state than H-state? Proposition 0.2 and Corollary 2 predict that given that small members are capacity constrained in H-state, the actual role OPEC plays is in L-state. This is because the L-state is exactly when small members produce under their capacity limits and therefore can potentially overproduce their quota levels. We break down this prediction into two parts. First, we test if small members are actually more capacity constrained in H-state, as opposed to the police/large country who strategically keeps an excess capacity in both states and therefore is never capacity constrained. Second, to test if OPEC behaves as a cartel more in L-state than in L-state, we investigate if OPEC is more responsive to an oil price shock in L-states than in H-states.

To investigate the second part, we depict the relationship between OPEC meeting frequency and the real oil price in Figure 5. In this figure the horizontal axis presents 10 bins of real oil price and the vertical axis shows the average number of OPEC meetings in a quarter for each price bin. If OPEC is more responsive to oil price shocks in L-states, we expect to see more meeting in L-state than in H-state; a negative relationship. Consistent with Corollary 2, Figure 5 confirms a negative relationship between OPEC meetings frequency and the oil price.

Next, we use the following time-series model to test for the statistical and economic significance of this relationship:

$$\begin{aligned} \text{(17)} \qquad Meeting\ Frequency_t &= \beta_0 + \beta_1 Price_{t-1} + \beta_2 Price\ Jump_{t-1} \\ &\quad + \beta_3 Price_{t-1} \times Price\ Jump_{t-1} + \varepsilon_{it} \end{aligned}$$

where *Meeting Frequency* is the number of OPEC meetings in a quarter that resulted in a quota change, *Price* is the quarterly real oil price in 2001 dollars, and *Price Jump* is an indicator variable that captures oil price jumps. A positive jump is as an increase in real oil price that exceeds 30%, a

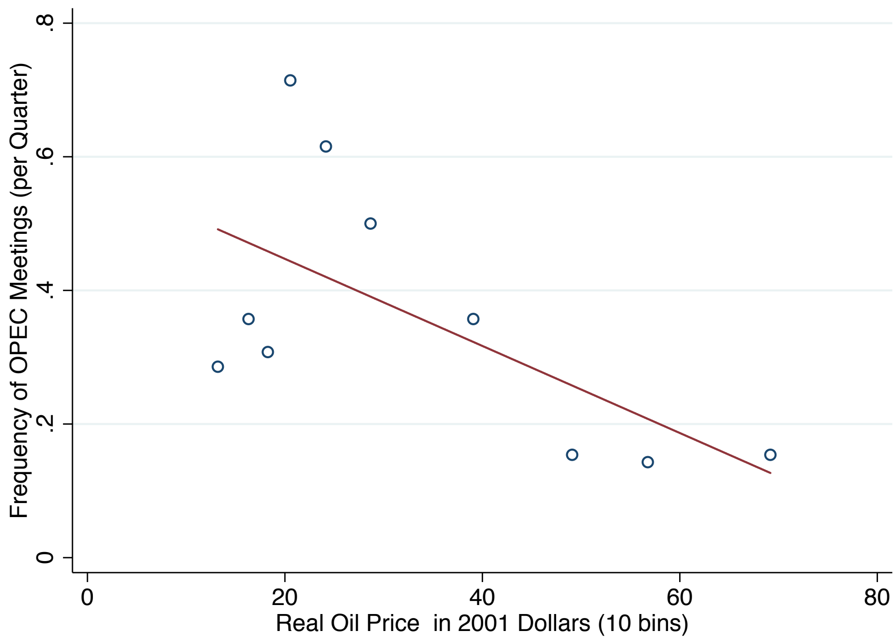

Figure 5: **Meeting Frequency vs. Oil Price.**This figure presents average frequency of OPEC meetings in each quarter as a function of oil price between 1983 and 2016. Horizontal axis shows 10 bins of the real oil price in 2001 dollars. The vertical axis shows the average number of OPEC meetings in each quarter for each price bin. OPEC meetings is restricted only to the meetings that members decided a change in quotas.

negative jump is a drop in real oil price by more than 30%, and no jump refers to quarters in which the oil price changes stays within the -30% and +30% bands. If OPEC is more responsive to negative jumps in oil price we expect to observe a positive and significant sign for β_2 when there is a negative price jumps only. Moreover, if the drop in oil price matters more in L-state, we expect a negative and significant sign for β_3.

Dependent Variable: Number of OPEC Meetings in a Quarter

		Positive Price Jump	No Price Jump	Negative Price Jump
	(1)	(2)	(3)	(4)
$Price_{t-1}$	-0.005*	-0.005*	-0.007	-0.003
	(-1.76)	(-1.73)	(-0.83)	(-1.24)
$PriceJump_{t-1}$		-0.047	-0.313	1.989***
		(-0.11)	(-0.98)	(2.69)
$Price_{t-1} \times PriceJump_{t-1}$		0.006	0.003	-0.072**
		(0.57)	(0.35)	(-2.46)
$Constant$	0.514***	0.516***	0.776**	0.460***
	(5.00)	(4.77)	(2.60)	(4.39)
R^2	0.023	0.027	0.036	0.073
Observations	135	132	132	132

Table 4: **OPEC Meeting Frequency and Oil Market Conditions.** This table presents regression results of frequency of OPEC meetings as a function of oil market conditions, i.e., oil price movements. The model specification is given in equation (17). The dependent variable is the number of OPEC meetings in a quarter. *Price* is the real oil price in 2001 dollars; *Price Jump* is a dummy variable that captures positive, negative or no jump in oil price. A jump is defined as quarterly change in real pil price that exceeds 30%. The data is on a quarterly basis between 1983 and 2016. Numbers in parenthesis are t-statistics, and significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

Table 4 presents the estimation results for equation (17). Consistent with Corollary 2 and similar to Figure 5, the first column of this table shows a significant and negative sign for β_1 indicating a negative relationship between OPEC meetings frequency and the oil price. Columns (2), (3) and (4) show the estimation results for positive, negative and no jump scenarios. While the overall relationship between meeting frequency and real oil price remains negative in all three columns, an important distinction between the results in column (3) and other two columns is clear. That is the likelihood of OPEC meetings significantly increase in response to a negative price jump especially when the oil price is lower. This can be seen by the fact that the only time β_2 and β_3 are significant is in column (3) when there is a negative price jump. Moreover, the size of the estimated coefficient for β_2 shows the impact is economically significant: a negative price jump increases the number of meetings in a quarter by almost 2. This result is consistent with the notion that the OPEC is more of a cartel when there is drop in oil prices especially in L-states than in H-state as stated by Corollary 2.

Non-compliance Pattern

A size-dependent punishment strategy implies that the tolerance for non-compliance is higher for the smaller members. Thus, we expect to see a negative relationship between producer size and non-compliance when it is measured in relative terms (i.e., overproduction as a fraction of quota). In contrast, if we measure the non-compliance in absolute terms (i.e., the difference between quota and production), we may observe a larger degree of non-compliance for the larger country simply because of its larger production scale. To investigate this, we use the following panel data model:

$$
\begin{aligned}
NC_{it} = \beta_0 + \beta_1 Price\ Increase_{it} + \beta_2 Large_{it} \\
+ \beta_3 Price\ Increase_{it} \times Large_{it} + \varepsilon_{it}
\end{aligned}
\tag{18}
$$

where NC_{it} is the non-compliance by country i in period t in either relative or absolute terms, *Price Increase* is an indicator variable that switches on when real oil price exceeds its three-quarter moving average. As illustrated in the appendix this indicator well captures the boom periods of the oil market. We use this measure instead of the level of oil price here because non-compliance is a short-term phenomenon. When countries adjust to new market conditions it becomes their new reference point. This reference point plays an important role in countries compliance decision in response to a new change in market conditions. *Large* is also an indicator variable for the police/large country. A size-dependent punishment strategy that predicts higher non-compliance for the smaller countries, implies a negative coefficient for *Large*, β_2, when non-compliance is measured in relative terms. Instead, if we measure non-compliance in absolute terms we may expect a positive sign for β_2 simply because of its scale. β_3 captures the asymmetry in non-compliance pattern in good (high demand) vs. bad (low demand) times of the oil market. For example, if the police country plays a "swing producer" role (Jones (1990)) by over-producing more in good times than bad times, we expect a positive sign for β_3.

Estimation results for equation (18) are reported in Table 5. Columns (1) and (2) show the estimation results for non-compliance in absolute and relative terms respectively using an OLS model. Focusing on the second row, a positive and significant coefficient of 0.121 for β_2 in column (1) shows that on average the large producer overproduces its quota in absolute terms by 12% more than smaller members. Similarly, a negative and significant value of -0.047 for β_2 in column (2) indicates that on average the non-compliance of the large country relative to its quota is 4.7% less than that of the smaller countries. Moving to the third row, an insignificant β_3 in column (2) suggests that relative overproduction of smaller members is not state dependent. However, consistent with the "swing producer" hypothesis a positive and significant β_3 in column (1) is indicative of an average of 20% more overproduction by the large member in H-state as opposed to L-state in absolute terms.

One potential caveat with the OLS model specification in equation (18) is that it's not clear if the sign of the *Large* coefficient is driven by a mechanical relationship between non-compliance and member's size. Given that quotas are allocated proportional to a member's capacity, for a given production level, larger capacity translates into larger quota and therefore lower non-compliance. We address this issue in two ways. First, using instrumental variable (IV) approach, we proxy countries' capacity by their oil reserves. This eliminates the direct link between the capacity on the left and right hand sides of the equation (18). Second, we re-arrange equation (18) in such a way that quotas or capacities do not directly impact the left hand side of the equation; non-compliance.

Dependent Variable: Degree of Non-Compliance

	OLS Model		IV Model	
	Absolute	Relative	Absolute	Relative
	(1)	(2)	(3)	(4)
$Price\ Increase_t$	-0.029	0.003	-0.026	0.004
	(-0.52)	(0.10)	(-1.24)	(0.33)
$Large_{it}$	0.121***	-0.047**	0.08	-0.165*
	(5.81)	(-2.25)	(0.91)	(-1.73)
$Price\ Increase_t \times Large_{it}$	0.187***	0.025	0.159**	0.013
	(3.37)	(0.84)	(2.06)	(0.30)
$Constant$	0.089***	0.079***	0.093***	0.090***
	(4.27)	(3.76)	(3.81)	(3.64)
R^2	0.104	0.006	0.104	0.005
$Observations$	519	519	519	519

Table 5: **Non-compliance Pattern-Conventional Model Specification.** This table presents regression results that describes non-compliance as a function of country size and market conditions. The model specification is given in equation (18). The dependent variable is $Log(Production_t - Quota_t + 1)$ (non-compliance in absolute terms) in columns (1) and (3), and $(Production_t - Quota_t)/Quota_t$ (non-compliance in relative terms) in columns (2) and (3). *Production* and *Quota* are in million bpd. *Price Increase* is an indicator variable that switches on when real oil price in 2001 dollars exceeds its three-quarter moving average. *Large* is an indicator variable for Saudi Arabia. The data is on a quarterly basis between 1995 and 2007. Numbers in parenthesis are t-statistics, and significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

As for the first one, we use variations in oil reserves as an instrument for changes in size (measured by capacity). The first stage regression estimation results are reported in the appendix suggesting a strong relationship between reserves and capacity. Overall, the estimation results for the IV model, reported in columns (3) and (4) of Table 5, confirm the findings of the OLS model, albeit with different magnitudes for the coefficients. Even though β_2 in column (3) is still positive, but it is no longer significant. This observation, plus a positive and significant coefficient for β_3 in the same column suggest that the large country overproduces its quota by 17% more than smaller countries only in booms the oil market. Similarly, the negative and significant coefficient of -16.5 for β_2 in column (4) suggest that the non-compliance of the large country relative to its quota is 16.5% less than that of the smaller countries.

For the second one, we modify equation (18) where NC_{it} is non-compliance in relative terms. We multiplying both sides of the equation by *Quota* and solve for *Production*, resulting in the following specification:

(19)

$$\begin{aligned} Production_{it} &= \beta_0 Quota_{it} + \beta_1 Quota_{it} \times Price\ Increase_{it} \\ &+ \beta_2 Quota_{it} \times Large_{it} + \beta_3 Quota_{it} \times Price\ Increase_{it} \times Large_{it} + \varepsilon_{it} \end{aligned}$$

where *Production* and *Quota* are the same as before and X_{it} is a vector of controls including measures of size and market conditions. If OPEC members fully comply with allocated quotas, then we expect a significant and positive coefficient for *Quota*, β_0, that is equal to 1. Moreover, we expect all other X variables and their interaction with *Quota* to be insignificant. However, if there is a systematic pattern for non-compliance among OPEC countries, we no longer expect $\beta_0 = 1$. For example, if larger country overproduces its quota more than smaller countries we expect a positive and significant β_2.

Estimation results reported in Table 6 are qualitatively consistent with our findings in Table 5. Focusing on the last column, a Chi-square test rejects the null hypothesis that $\beta_0 = 1$ with $p - value < 0.0001$. A $partialF - test$ rejects the null hypothesis that $\beta_1 = \beta_2 = \beta_3 = 0$ with a $p - value = 0.0002$.

Dependent Variable: Production

	(1)	(2)	(3)	(4)	(5)
$Quota_{it}$	1.042***	1.036***	1.032***	1.027***	1.046***
	(328.06)	(200.65)	(197.04)	(154.74)	(123.76)
$Quota_{it} \times Price\ Increase_t$		0.009		0.009	-0.022**
		(1.30)		(1.32)	(-2.11)
$Quota_{it} \times Large_{it}$			0.015**	0.015**	-0.015
			(2.31)	(2.32)	(-1.44)
$Quota_{it} \times Price\ Increase_t$					0.049***
$\times Large_{it}$					(3.65)
R^2	0.995	0.995	0.995	0.995	0.995
Observations	519	519	519	519	519

Table 6: **Non-compliance Pattern-New Model Specification.** This table presents regression results that describes non-compliance as a function of country size and market conditions. The model specification is given in equation (19). The dependent variable is *Production* in million bpd. *Price Increase* is an indicator variable that switches on when real oil price in 2001 dollars exceeds its three-quarter moving average. $Large_{it}$ is an indicator variable for Saudi Arabia. *Quota* are in million bpd. The data is on a quarterly basis between 1995 and 2007. Numbers in parenthesis are t-statistics, and significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

This suggest quotas are not the only determinants of the production, and OPEC members compliance varies with their size and market conditions. An estimated value of -0.022 for β_1 suggests that non-compliance in relative terms for smaller countries ($Large = 0$) is 0.02 less in good times of the oil market than its bad times. This is consistent with our earlier findings in Table 3 that small producers are more capacity constrained when oil prices are higher and therefore are less likely to deviate from their quota levels. Consistent with our findings in Table 5, the estimated value of -0.015 for β_2 suggest that non-compliance in relative terms is 1.5% lower for the large or equivalently higher for the smaller countries with a $p - value = 0.15$. Finally, the estimated coefficient of 0.049 for the triple interaction term, β_3, suggests that large country indeed does play a swing producer role by overproducing relative to its quota by 4.9% more in good times compared to smaller countries. This means in net the larger country overproduces its quota by 2.7% (=4.9% - 2.2%) in booms of the oil market.

Overall, consistent with the implications of a size-dependent punishment strategy, the results in Tables 6 and 5 confirm that smaller members have higher degrees of non-compliance. Moreover, consistent with Corollary 1, we find that smaller members are more likely to deviate from their quota levels in bad times of the oil market. Finally, we find evidence consistent with swing producer behavior for the police/large member.

Conclusion

In this paper we have formulated the optimal decision-making process of OPEC members to determine the equilibrium capacity investment, quota allocations and the corresponding production choices that countries make outside of OPEC. Our model includes strategic choices of production capacity and also a punishment mechanism to (at least partially) enforce the cartel's allocated quotas.

The theoretical model highlights the role of resource endowment, demand state (price level), excess capacity, and the strength of the punishment mechanism in shaping members incentives to deviate or comply. Our empirical results provide support to theoretical insights. We show that capacity constraints in the high demand state function as an enforcement mechanism. However, in the low demand state, OPEC needs to work hard to ensure that the large spare capacity of members will not cause them to deviate from their quotas.

Our empirical analysis confirms that OPEC is indeed more during low price episodes. We also show that the likelihood of producing at the capacity will be much higher in the high demand state. Finally, we show that small members make larger proportional deviations, compared to larger producers.

The current research can be extended in multiple directions. First, one can further extend the theoretical model to include a Markovian demand structure (rather than i.i.d) together with heterogeneous production costs to study the incentive for deviation in the L and H states. The extended model can provide additional insights regarding the stability of OPEC cartel in low and high demand states.

Second, the model can be extended by making the choice of the severity of the punishment mechanism (i.e. the probability of punishment) endogenous. The police understands the trade-off considered by small producers between building additional capacity and deviating from allocated quota. Given the parameters of the model there could be an optimal choice of the stringency of the punishment mechanism, which will balance the cost of spare capacity (build to obtain larger quotas) and the cost of tolerating deviations from the optimal production. Given the debates around the rationale for OPEC's quota system one can examine the optimal choice of the punishment mechanism as a function of background variables.

Finally, The optimal level of the excess capacity for the police can be characterized and modeled by considering the option value of producing additional units in the high demand state along with the value of offering a credible threat. The full characterization is beyond the scope of the current paper. Future research can provide additional singhts regarding the optimal choice of capacity for a cartel police.

References

A Fayçal Alhajji and David Huettner. Opec and other commodity cartels: a comparison. *Energy Policy*, 28(15):1151–1164, 2000.

Switgard Feuerstein. Collusion in industrial economics—a survey. *Journal of Industry, Competition and Trade*, 5(3-4):163–198, 2005.

James M Griffin. Opec behavior: a test of alternative hypotheses. *The American Economic Review*, 75(5):954–963, 1985.

S Gürcan Gülen. Is opec a cartel? evidence from cointegration and causality tests. *The Energy Journal*, pages 43–57, 1996.

Carol Dahl and Mine Yücel. Testing alternative hypotheses of oil producer behavior. *The Energy Journal*, pages 117–138, 1991.

James L Smith. Vinscrutable opec. *Behavioral Tests of the Cartel Hypothesis. V Energy Journal*, 26(1):51, 2005.

Paul A Geroski, Alistair M Ulph, and David T Ulph. A model of the crude oil market in which market conduct varies. *The Economic Journal*, 97:77–86, 1987.

Pedro A Almoguera, Christopher C Douglas, and Ana María Herrera. Testing for the cartel in opec: non-cooperative collusion or just non-cooperative? *Oxford Review of Economic Policy*, 27(1):144–168, 2011.

Robert K Kaufmann, Andrew Bradford, Laura H Belanger, John P Mclaughlin, and Yosuke Miki. Determinants of opec production: Implications for opec behavior. *Energy Economics*, 30(2):333–351, 2008.

Edward J Green and Robert H Porter. Noncooperative collusion under imperfect price information. *Econometrica: Journal of the Econometric Society*, pages 87–100, 1984a.

Julio J Rotemberg and Garth Saloner. A supergame-theoretic model of price wars during booms. *The American Economic Review*, 76(3):390–407, 1986a.

John Haltiwanger and Joseph E Harrington Jr. The impact of cyclical demand movements on collusive behavior. *The RAND Journal of Economics*, pages 89–106, 1991.

Michihiro Kandori. Correlated demand shocks and price wars during booms. *The Review of Economic Studies*, 58(1):171–180, 1991.

Kyle Bagwell and Robert W Staiger. Collusion over the business cycle. Technical report, National Bureau of Economic Research, 1995.

William A Brock and Jose A Scheinkman. Price setting supergames with capacity constraints. *The Review of Economic Studies*, 52(3):371–382, 1985.

Robert W Staiger and Frank A Wolak. Collusive pricing with capacity constraints in the presence of demand uncertainty. *The RAND Journal of Economics*, pages 203–220, 1992a.

Natalia Fabra. Tacit collusion in repeated auctions: uniform versus discriminatory. *The Journal of Industrial Economics*, 51(3):271–293, 2003.

Dan Bernhardt and Mahdi Rastad. Collusion under risk aversion and fixed costs. *The Journal of Industrial Economics*, 2016.

Robert W Staiger and Frank A Wolak. Collusive pricing with capacity constraints in the presence of demand uncertainty. *The RAND Journal of Economics*, pages 203–220, 1992b.

John Gault, Charles Spierer, Jean-Luc Bertholet, and Bahman Karbassioun. How does opec allocate quotas? *Journal of Energy Finance & Development*, 4(2):137–148, 1999.

Pavel Molchanov. A statistical analysis of opec quota violations. *Economics. Duke University, Durham, NC*, pages 1–31, 2003.

Sel Dibooglu and Salim N AlGudhea. All time cheaters versus cheaters in distress: An examination of cheating and oil prices in opec. *Economic Systems*, 31(3):292–310, 2007.

Julio J Rotemberg and Garth Saloner. A supergame-theoretic model of price wars during booms. *The American Economic Review*, 76(3):390–407, 1986b.

Christopher R Knittel and Jason J Lepore. Tacit collusion in the presence of cyclical demand and endogenous capacity levels. *International Journal of Industrial Organization*, 28(2):131–144, 2010.

Jean-Pierre Amigues, Pascal Favard, Gérard Gaudet, and Michel Moreaux. On the optimal order of natural resource use when the capacity of the inexhaustible substitute is limited. *Journal of Economic Theory*, 80(1): 153–170, 1998.

Jonathan Cave and Stephen W Salant. Cartel quotas under majority rule. *The American Economic Review*, pages 82–102, 1995.

Rafael Sandrea. Opec's next challenge–rethinking their quota system. *Oil and Gas Journal*, 101:29, 2003.

Douglas B Reynolds and Michael K Pippenger. Opec and venezuelan oil production: Evidence against a cartel hypothesis. *Energy Policy*, 38(10): 6045–6055, 2010.

Veronika Grimm and Gregor Zoettl. Investment incentives and electricity spot market competition. *Journal of Economics & Management Strategy*, 22(4):832–851, 2013.

Thomas Fagart. Collusion in capacity under irreversible investment. *Working Paper*, 2016.

Edward J Green and Robert H Porter. Noncooperative collusion under imperfect price information. *Econometrica: Journal of the Econometric Society*, pages 87–100, 1984b.

Clifton T. Jones. Opec behaviour under falling prices: Implications for cartel stability. *The Energy Journal*, 11(3):117–129, 1990.

Details of Calculations

One-Short Deviation

Given the aggregate production of the rest of OPEC (denoted by $\overline{Q}$) the deviating member finds the optimal production q^* to maximize the following function:

$$\pi(q^*, X_s) = q^*(X_s - b(\overline{Q} + q^*) \tag{20}$$

The FOC is:

$$q^* = \frac{X_s - b\overline{Q}}{2b} \tag{21}$$

The equilibrium price will be:

$$P = X_s - b(\overline{Q} + \frac{X_s - b\overline{Q}}{2b}) = \frac{X_s - b\overline{Q}}{2} \tag{22}$$

The profit is

$$\pi(q^*, s) = [\frac{X_s - b\overline{Q}}{2b}]^2 \tag{23}$$

Value Function of Punishment State

The value function in the punished regime is:

$$V_{N,i} = \frac{\mathbb{E}[\pi^i_{\text{Cournot}}]}{r} \tag{24}$$

$$\max_{q_i} \pi_i = [X - \gamma(q_i + q_j)]q_i - C \tag{25}$$

$$q_i^H = K_i = \frac{p_H \phi_j X_H}{\phi_i \phi_j + 2p_H \gamma(\phi_i + \phi_j)} \tag{26}$$

$$q_i^L = \frac{X_L}{3\gamma} \rightarrow \pi_L = \frac{X_L^2}{9\gamma} - C \tag{27}$$

$$\begin{cases} V_H = \pi_H + (\frac{1}{1+r})(p_{H,H}V_H + p_{H,L}V_L) \\ V_L = \pi_L + (\frac{1}{1+r})(p_{L,L}V_L + p_{L,H}V_H) \end{cases} \tag{28}$$

$$\begin{cases} V_H = \dfrac{\pi_H\left(1-\frac{p_{L,L}}{1+r}\right)+\left(\frac{X_L^2}{9\gamma}-C\right)\left(\frac{p_{H,L}}{1+r}\right)}{\left(1-\frac{p_{L,L}}{1+r}\right)\left(1-\frac{p_{H,H}}{1+r}\right)-\left(\frac{p_{H,L}}{1+r}\right)\left(\frac{p_{L,H}}{1+r}\right)} \\ \\ V_L = \dfrac{\left(\frac{X_L^2}{9\gamma}-C\right)\left(1-\frac{p_{H,H}}{1+r}\right)+\pi_H\left(\frac{p_{L,H}}{1+r}\right)}{\left(1-\frac{p_{L,L}}{1+r}\right)\left(1-\frac{p_{H,H}}{1+r}\right)-\left(\frac{p_{H,L}}{1+r}\right)\left(\frac{p_{L,H}}{1+r}\right)} \end{cases} \tag{29}$$