On Quota Violations of OPEC Members

Abstract

Over the last decades quota violations have become a norm for OPEC countries. However, 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 with empirical evidence to explain OPEC members’ incentives for abiding 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 risking quota violation punishment. 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 a theoretical framing in which capacity constraints work as an enforcement mechanism in good times and OPEC’s quota system disciplining its members in bad times.

Keywords: Energy, OPEC, Cartels, Oil market, Compliance, Quota, Cournot Duopoly, Collusion

1. Introduction

OPEC has a fragile cartel structure. It does not have a formal enforcement mechanism in place (except for the occasional price war) to incentivize its members to comply with their quota allocations. (Alhajji and Huetterner (2000)). As a result, non compliance to the quota 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%. In this paper, we provide a framework to explain such a persistent non-compliance pattern based on strategic

---

1We thank Jason Lepre, Davood Taree, Franz Wirl, Philipp Schmidt-Dengler, Filippo Pavesi and seminar participants at California Polytechnic State University and Vienna University (Energy Economics Group) for their helpful comments. We also thank Robert Kaufmann for sharing data on OPEC members’ capacity. All remaining errors are ours.

Preprint submitted to IIOC January 22, 2018
interactions among members inside OPEC.

Quota violations vary significantly among different countries. For example, Saudi Arabia’s overproduction averaged around 3.2% from 1995 to 2007 but Qatar’s overproduction averaged around 18.5% during the same period. In extreme cases, some members overproduce their quotas by a large margin. For example, in the second quarter of 1998, OPEC requested that Qatar cut its production to reach a quota of 384 thousand bpd due to falling oil prices, but Qatar kept its production level at 670 thousand bpd—a 75% quota non-compliance. Another extreme example is Algeria who 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, the difficulty of supporting collusion varies with the economic state 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 counter-cyclical 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 quotas and violations. To name a few, Griffin (1985), 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. Gülén (1996) uses a similar intuition to test for parallel movements in members’ output levels. He finds evidence 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

\[2\text{See Feuerstein (2005) for a survey of collusion literature.}\]
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 field of industrial organization provides a 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 to repeat game models 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 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 prediction of this literature is that it’s harder to collude in booms because the incentive to cheat is higher. 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 demand,

\footnote{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.}
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 falling demand the value of cooperation 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 Markov 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.

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 are determined through maximizing the joint profit of the cartel members. In the third stage countries decide their production level and in the fourth 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 and leaves 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

---

4This 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).
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.

The key feature of our theoretical framework is to model the decision making process by the cartel members inside and outside of OPEC. We do this by highlighting the role of an important variable, under-emphasized in 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, as pointed out by Gault et al. (1999). Therefore, some of results are driven by the capacity limits. Third, larger capacity provides the option to produce in high-demand states for the larger member in the production game that follows the quota allocations.

Our empirical analysis supports these predictions. Using quarterly data for the period of 1995 to 2007, during when OPEC had a stable structure, we build a panel data model to test the statistical significance of the predictions of our theory model. We use Instrumental Variable approach to control for potential endogeneity of the size measure, Capacity. 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 that OPEC meets more frequently when oil prices are lower. 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.

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)), we provide a theoretical model that is tuned to resemble the OPEC decision-making process and delivers empirically testable predictions along with our empirical analysis. 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 contains the conclusion.

2. 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 between 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 account the installed capacities).
Countries make decisions regarding the 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 fourth 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, quota violations, and punishment are all endogenous.

2.1. Key Components and Assumptions of the Model

OPEC is a sophisticated and non-transparent organization. A lot of politics is involved in OPEC’s bargaining, agreements, and decision making. We recognize that a stylized model, such as the one presented in this paper, cannot 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 (apprx $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’s function and abstract from this alternative view.

\(^6\)It is questionable if OPEC indeed could achieve its price stabilization 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. Since the interest of the paper is in quota violations of non-police members, 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.

**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 of any other OPEC member to support a credible threat against deviations from the cartel allocation.\(^8\)

**Demand.** The demand for crude oil has a standard linear functional form: $P = \theta - \gamma Q$. Here $\theta$ is the demand shift parameter and $\gamma$ 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 $\theta_H$ (high demand) and $\theta_L$ (low demand), with probabilities of $p_H$ and $p_L$, respectively.

---

\(^8\)The optimal level of the excess capacity for the police can be found 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.
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 efficient fields or 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 have different levels and curvatures in their cost curves. This heterogeneity for 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.

\(^9\)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.
Under an efficient planner solution the country with a lower capacity building cost (i.e. cost curve to the right) will optimally build a larger capacity and receive a higher production quota.

**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) uses Venezuela as an example and shows that in the long-run, production capacity and actual productions of countries Granger-cause their OPEC quotas.

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

The problem with the standard definition of punishment in cartels (i.e the strategy of flooding the market with overproduction and thus reducing the price) is that it not only punishes the non-compliant member but also other “innocent” members (including the police
This causes the punishment strategy to be ex-post sub game 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. During important historical incidents (e.g. the oil price war of 1986) Saudi Arabia overproduced to punish the other members. In order to have a credible punishment strategy, cartel members need to have enough spare capacity, able to identify the deviant member (Green and Porter (1984a)), and not be jammed by demand uncertainty noises.

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

A special case is the size-independent punishment technology, which will react to all positive deviations with a constant probability $m$. The random punishment can be interpreted as a mixed strategy played by the police\(^{11}\).

In the case of size-independent punishment, 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.

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 are several frictions that can result in a partially effective punishment technology. We summarize all of these forces using the single parameter $m$.

\(^{10}\)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.

\(^{11}\)A size-dependent punishment technology will make the probability of punishment a function of the magnitude of deviation. If a member deviates from the assign quota by $\Delta q$, a monitoring technology with strength $m()$ will trigger the punishment strategy with the probability of $m(\Delta q)$. In this case $m()$ is a function rather than a number.
2.2. Four Versions of the Model

A step-by-step development of cartel models is presented in the following order:

1. A planner’s solution for a cartel with a convex capacity-building costs under certainty
2. A planner’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, possibility of quota violation, and endogenous punishment for quota violation

2.3. 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 $\phi_i$ is the coefficient of resource efficiency and $K_i$ is the installed capacity. For countries with a larger resource base, $\phi$ 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. 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).

\footnote{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).}
The optimal cartel solution is:

\[
\begin{align*}
\max_{K_1,K_2} & \quad P(q_1 + q_2)(q_1 + q_2) - c_1(K_1) - c_2(K_2) \\
\text{s.t.} & \quad q_1 \leq K_1, \quad q_2 \leq K_2
\end{align*}
\]  

(1)

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 maximizes the joint profit of all plants. The FOCs of the cartel problem results in the following system of equation, in which the marginal cost of each member is equal to the marginal revenue of the cartel:

\[
\begin{align*}
c_1'(K_1) &= P'(K_1 + K_2)(K_1 + K_2) + P(K_1 + K_2) \\
c_2'(K_2) &= P'(K_1 + K_2)(K_1 + K_2) + P(K_1 + K_2)
\end{align*}
\]  

(2)

Given that the right hand sides are equal, we immediately observe that \(c_1'(K_1) = c_2'(K_2) \Rightarrow \frac{K_1}{K_2} = \frac{\phi_2}{\phi_1}\). Since capacity is set 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 \(\phi\)) will receive a lower production quota.

Plugging in the specific functional forms for \(c_i\) and \(P\):

\[
\begin{align*}
q_1 &= K_1 = \frac{X\phi_2}{\phi_1\phi_2 + 2\gamma(\phi_1 + \phi_2)} \\
q_2 &= K_2 = \frac{X\phi_1}{\phi_1\phi_2 + 2\gamma(\phi_1 + \phi_2)}
\end{align*}
\]  

(3)

The solution to the system of equation characterizes an efficient internal allocation while maximizing the total profit of the cartel.
2.4. 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 with the fixed capacity installed in the first stage, countries produce according to a profit-maximizing program.

The demand states are distributed according to $\theta_L, p_L$ and $\theta_H, p_H$, 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 $q^H_i$ and $q^L_i$.

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 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, $K_i = q^H_i$

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\[13\]

\[
\begin{align*}
  c'_1(K_1) &= p_H[P'(K_1 + K_2)(K_1 + K_2) + P(K_1 + K_2)] \\
  c'_2(K_2) &= p_H[P'(K_1 + K_2)(K_1 + K_2) + P(K_1 + K_2)]
\end{align*}
\]

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

\[
\begin{align*}
q_1^H = K_1 &= \frac{p_H \phi_2 \theta_H}{\phi_1 \phi_2 + 2p_H \gamma (\phi_1 + \phi_2)} \\
q_2^H = K_2 &= \frac{p_H \phi_1 \theta_H}{\phi_1 \phi_2 + 2p_H \gamma (\phi_1 + \phi_2)}
\end{align*}
\]  

(5)

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

\[
\max_Q P(Q, X)Q
\]  

(6)

The F.O.C of the monopoly problem:

\[
\begin{align*}
P'(Q, X)Q + P(Q, X) &= 0 \Rightarrow Q^*_L(X) = \frac{\theta_L}{2\gamma} \\
\Rightarrow P(Q^*_L) &= \theta_L - \gamma Q^*_L = \frac{X_L}{2}
\end{align*}
\]  

(7)

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

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

\[\text{Note that the capacity investment cost is sunk and does not enter the optimal production equation in the } L \text{ state.}\]
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. The operational deviation is more frequently observed in practice. Given the nature of our motivating empirical observations, our theoretical focus is on operational deviations. Consequently, we only model the punishment mechanism for the operational deviations and abstract from modeling the punishments for the investment deviations.

2.5. 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)).

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:

$$q_i(\theta) = \begin{cases} \frac{\theta_i K_i}{\sum_j K_j} & \text{if } \frac{\theta_i K_i}{\sum_j K_j} \leq K_i \\ K_i & \text{otherwise} \end{cases}$$

(8)

Note that OPEC members’ individual capacities do not affect the optimal level of the 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

\[\text{In 1997 Venezuela followed an ambitious plan to significantly increase its production capacity. In a reaction to this move Saudi Arabia increased its production and caused the oil price to drop to its historical low level of $10 per barrel.}\]
production, some members will receive quotas below their capacities\[17\]

The objective function of the nested optimization problem is

\[
\max_{K, q} \left[ \mathbb{E} \left( \max_{q \leq \min \{ K, \pi \}} \pi^* (K, q) \right) - c(K) \right]
\]

(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 a given.) The new feature of this version of the model takes 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 with \( q^H_1 = K_1 \), the key equation to pin down the endogenous choice of capacity by a small cartel member is\[18\]

\[
\frac{d}{dK_1} \pi^* (K, q) = \begin{cases} \frac{\partial \pi_L}{\partial q_L} \frac{\partial q_L}{\partial K_1} & \text{Marginal quota value in the L state} \\ \frac{\partial P_H}{\partial q_H} (q_H^1 + q_H^2) q_H^1 + \frac{\partial P_L}{\partial q_L} (q_L^1 + q_L^2) & \text{Marginal production value in the H state} \end{cases}
\]

(10)

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:

\[
q_i(X) = \begin{cases} Q^*_L \frac{K_i}{\sum K_i} & \text{L state} \\ K_i & \text{H state} \end{cases}
\]

(11)

Denoting the assigned quota in the low demand state by \( q^L_i \) and assuming \( K_i \ll \sum K_i \Rightarrow \)

\[17\]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.

\[18\]As a first-order approximation we ignore the effect of increased capacity of production during the H-state and assume the production constraints are still binding in that state.
\[
\sum_i K_i \approx 0, \text{ one gets:}
\]

\[
\frac{\partial q^L_i}{\partial K_i} = Q^* \frac{\sum_i K_i - K_i}{(\sum_i K_i)^2} \approx \frac{Q^*}{\sum K_i}
\]

We see that with a larger total capacity of OPEC members, the quota allocation effect of additional unit of capacity for an individual member becomes weaker. 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, closer production values in the two states results in a higher incentive for bargaining.

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

\[
\frac{\partial \pi^L_i}{\partial K_i} = \frac{\partial \pi^L_i}{\partial q^L_i} \frac{\partial q^L_i}{\partial K_i} \approx \pi^L_i \frac{Q^*}{\sum K_i} \approx \frac{X^2 L}{4\gamma \sum K_i}
\]

Equation 13 suggests that when the total installed capacity of the cartel is large, the demand in the low state is weak, and 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 bargaining purposes; however, if capacity investment costs are high, the country may give up building excess capacity.

The optimality condition to pin-down the equilibrium capacity for the representative small member is:

\[
c'_1(K_1) = p_H [P'(q^H_1 + q^H_2)q^H_1 + P(q^H_1 + q^H_2)] + p_L \frac{X^2 L}{4\gamma \sum K_i}
\]
2.6. 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 police. 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 of their current decisions. The first stages of the time-line occur only once; however, the production and punishment stages are a summarized form a repeated game.

2.6.1. 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.

2.6.2. 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:

\[
\text{Pay-off} = \begin{cases} 
\text{Compliance} : & \pi(q, s) + \frac{1}{1+r} \mathbb{E}[V(s)] \\
\text{Violation} : & \pi(q^*, s) + \frac{1}{1+r} \left\{ (1-m) \mathbb{E}[V(s)] + m \mathbb{E}[V^C(s)] \right\}
\end{cases}
\]  

(15)

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 compliance. 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 shifts the value function to a Cournot value function. The details of the Cournot value function is presented in the Appendix.

The violation branch 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:

\[
\pi(q^*) - \pi(q) > m \times (\mathbb{E}[V^C] - \mathbb{E}[V])
\]

(16)

Any punishment strategy that the benefit of a one-shot deviation is smaller than the lost

\(^{19}\text{The demand state switches between } L \text{ and } H, \text{ thus, the country considers the expected value function given the current state of the demand.}\)
value of cooperation will induce members to comply.\footnote{We know from the folks theorem of repeated games that there are many strategies in the game that support a Nash equilibrium.}

From Equation \ref{eq:incentive} 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.

\textbf{Proposition 2.1.} \textit{The incentive to cheat is higher in the H-state compared to the L-state.}

\textit{Proof.} The R.H.S of Equation \ref{eq:incentive} 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)). \hfill \square

\textit{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.

\textbf{Proposition 2.2.} \textit{If the size of the strategic excess capacity is small, punishment is only required in the low state.}

\textit{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 deviation can be significant\footnote{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 \textit{marginal value} of an extra dollar of export revenue could be higher in bad times. This results in a stronger temptation for deviation or a myopic behavior.}. In summary, in the high state the capacity constraints function as the enforcement mechanism and in the low state quotas need 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 adjustments should be negatively associated with the level of crude oil price.

2.6.3. *Occasional Punishment*

We make a final assumption to create a model similar 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 occurs as 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. The 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 can 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. A plausible scenario is when a few periods of quota violations (until the cheating is detected) is followed by a severe punishment and reversion to the cooperation.

---

22The folks theorems of repeated games suggest that there might be many choices of several periods, all supporting an equilibrium
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.

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 observe that if the punishment probability depends on the aggregate deviation of all members, small members will have a larger proportional deviation. The same percentage of deviation by a small and a larger producer will have different absolute effects on the market. When 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.

**Proposition 2.3.** When the punishment becomes more effective, violations decreases but capacity investment also increases. Average spare capacity increases with the effectiveness of the monitoring technology.

**Proof.** The small member will compared expected value of a unit of additional production in the \( L \) state versus the cost of reaching that excess unit of production.

\[
\frac{p_L \pi'(q_L^i)}{r} = \frac{c'(K_1)}{X^L_i} \frac{1}{\gamma} \sum K_i + [m \ast (\mathbb{E}[V^C] - \mathbb{E}[V])] \tag{17}
\]
In the above equation, small members consider the value of a marginal legitimate quota through excess capacity building (given by Equation 14) versus the expected benefit of deviating from the legal quota (Equation 16). If the punishment is size-dependent, small members will choose a combination of building excess and some deviation.

The optimality condition for an interior solution is that the marginal cost of attaining one extra unit of production is the same through capacity building and cheating. There is a substitution effect between the two ways of increasing production. If \( m \) becomes bigger, then the member state will increase the excess capacity.

The intuition behind Proposition 2.3 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 member compares the marginal benefit of capacity (i.e. the probability of a marginal unit of quota) to the investment cost of capacity.

### 2.7. 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.
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 stringent punishment strategy increases incentives for building excess capacity to gain a larger legitimate quota.
3. 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.

3.1. 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 tend to under-report their production levels to hide quota violations. A comparison between OPEC vs. EIA confirms this conjecture.\textsuperscript{23} 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).\textsuperscript{24}

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.

\textsuperscript{23}Total OPEC production reported by OPEC was consistently lower than what is reported by EIA between 1983 and 2003, except in 1992.

\textsuperscript{24}Kaufmann et al (2008) estimate quarterly capacities by interpolating annual capacity data provided to them by Erik Kriel of the US Department of Energy.
Table 1: **Descriptive Statistics.** This table presents the summary statistics for the variables used in the empirical analysis. 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 (mbpd) and Reserves are in billion barrels (bb). 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>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserves</td>
<td>519</td>
<td>74.66</td>
<td>73.75</td>
<td>3.55</td>
<td>16.85</td>
<td>43.70</td>
<td>97.80</td>
<td>264.30</td>
</tr>
<tr>
<td>Capacity</td>
<td>519</td>
<td>2.96</td>
<td>2.71</td>
<td>0.53</td>
<td>1.44</td>
<td>2.31</td>
<td>2.83</td>
<td>11.40</td>
</tr>
<tr>
<td>Quota</td>
<td>519</td>
<td>2.47</td>
<td>2.08</td>
<td>0.38</td>
<td>1.28</td>
<td>2.00</td>
<td>2.72</td>
<td>9.10</td>
</tr>
<tr>
<td>Production</td>
<td>519</td>
<td>2.59</td>
<td>2.15</td>
<td>0.40</td>
<td>1.36</td>
<td>2.05</td>
<td>2.71</td>
<td>9.60</td>
</tr>
<tr>
<td>Spare Capacity</td>
<td>519</td>
<td>0.50</td>
<td>0.77</td>
<td>-1.30</td>
<td>0.14</td>
<td>0.31</td>
<td>0.58</td>
<td>3.89</td>
</tr>
<tr>
<td>Non-Compliance</td>
<td>519</td>
<td>7.7%</td>
<td>13.6%</td>
<td>-46.8%</td>
<td>1.4%</td>
<td>5.0%</td>
<td>10.9%</td>
<td>74.5%</td>
</tr>
<tr>
<td>Price (nominal $)</td>
<td>519</td>
<td>31.20</td>
<td>17.81</td>
<td>10.98</td>
<td>18.27</td>
<td>25.58</td>
<td>40.01</td>
<td>85.07</td>
</tr>
<tr>
<td>Price (real 2001$)</td>
<td>519</td>
<td>25.60</td>
<td>11.51</td>
<td>10.16</td>
<td>16.68</td>
<td>23.48</td>
<td>30.66</td>
<td>58.69</td>
</tr>
<tr>
<td>Meeting Frequency</td>
<td>135</td>
<td>0.36</td>
<td>0.58</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

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% for Venezuela to 74.5% for Qatar. It has an average of 7.7% and a median of 5%. Note that a negative number means producing below the quota or complying with the quota allocations, and a positive number signifies a quota violation. We also observe that 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 non-compliance sample because non-compliance is virtually impossible for those countries.

Table 2 presents summary statistics for non-compliance of each 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. We revisit this conjecture more formally later to see if it is a robust pattern in a multiple regression framework.
<table>
<thead>
<tr>
<th>Countries</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>52</td>
<td>0.21</td>
<td>0.23</td>
<td>-0.05</td>
<td>0.03</td>
<td>0.11</td>
<td>0.50</td>
<td>0.61</td>
</tr>
<tr>
<td>Indonesia</td>
<td>52</td>
<td>0.11</td>
<td>0.17</td>
<td>-0.25</td>
<td>-0.03</td>
<td>0.18</td>
<td>0.23</td>
<td>0.34</td>
</tr>
<tr>
<td>Iran</td>
<td>51</td>
<td>0.02</td>
<td>0.05</td>
<td>-0.09</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.05</td>
<td>0.13</td>
</tr>
<tr>
<td>Kuwait</td>
<td>52</td>
<td>0.06</td>
<td>0.06</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.04</td>
<td>0.11</td>
<td>0.26</td>
</tr>
<tr>
<td>Libya</td>
<td>52</td>
<td>0.07</td>
<td>0.05</td>
<td>-0.04</td>
<td>0.03</td>
<td>0.07</td>
<td>0.11</td>
<td>0.17</td>
</tr>
<tr>
<td>Nigeria</td>
<td>52</td>
<td>0.04</td>
<td>0.05</td>
<td>-0.05</td>
<td>0.00</td>
<td>0.04</td>
<td>0.08</td>
<td>0.15</td>
</tr>
<tr>
<td>Qatar</td>
<td>52</td>
<td>0.15</td>
<td>0.15</td>
<td>-0.01</td>
<td>0.06</td>
<td>0.10</td>
<td>0.16</td>
<td>0.75</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>52</td>
<td>0.05</td>
<td>0.04</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.04</td>
<td>0.06</td>
<td>0.15</td>
</tr>
<tr>
<td>UAE</td>
<td>52</td>
<td>0.03</td>
<td>0.03</td>
<td>-0.03</td>
<td>0.01</td>
<td>0.03</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>Venezuela</td>
<td>52</td>
<td>0.03</td>
<td>0.19</td>
<td>-0.47</td>
<td>-0.10</td>
<td>0.05</td>
<td>0.13</td>
<td>0.43</td>
</tr>
<tr>
<td>Total</td>
<td>519</td>
<td>0.08</td>
<td>0.14</td>
<td>-0.47</td>
<td>0.01</td>
<td>0.05</td>
<td>0.11</td>
<td>0.75</td>
</tr>
</tbody>
</table>

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.

### 3.2. Empirical Model and Estimation Results

In this section, to test the predictions of our theoretical model empirically, we investigate the compliance behavior of the individual OPEC members using a Logit, a time-series, and a panel data model. To investigate if the capacity constraint works as an enforcement mechanism in the 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.

#### 3.2.1. 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 [2.2] and Corollary 1 predict
that in the H-state the capacity constraint works as an enforcement mechanism for the small members to ensure that they do not overproduce their quota levels. 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.

![Figure 3: 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.](image)

To see if small members are in fact more capacity constrained in the H-state, we illustrate the frequency of producing at capacity for each member as a function of quartiles of the real oil price in Figure 3. 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

---

25 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.

26 Here we assume that oil price signals the demand state. The potential concern regarding the endogeneity of oil price due to reverse causality from production to price is alleviated by the fact that we find a positive relationship in Figure 3 that works in our favor as reverse causality would suggest a non-positive relationship.
country this likelihood is always zero.

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

\[
Constrained_{it} = \beta_0 + \beta_1 Price_t + \beta_2 Capacity_{it} \\
+ \beta_3 Price_t \times Capacity_{it} + \varepsilon_{it}
\]

(18)

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 mbpd.

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, \(\beta_3\), 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 the H-state.

3.2.2. OPEC Meetings Frequency

In previous section we observed that smaller members are more likely to be capacity constrained in H-state. Next, we investigate 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 the L-state than in the H-state? Proposition 2.2 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. To test if OPEC behaves as a cartel more in L-state than in H-state, we investigate if OPEC is more responsive to an oil price shock in L-states than in H-states.

To do this, we depict the relationship between OPEC meeting frequency and the real oil
Table 3: **Who and When Is Capacity Constrained?** This table presents regression results that describe which countries are more likely to produce at their capacity and when. The model specification is given in equation (18). 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 mbpd. 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.

```
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price(_t)</td>
<td>0.056***</td>
<td>0.061***</td>
<td>0.053***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.51)</td>
<td>(8.66)</td>
<td>(4.65)</td>
<td></td>
</tr>
<tr>
<td>Capacity(_t)</td>
<td>-0.321***</td>
<td>-0.366***</td>
<td>-0.513**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-4.38)</td>
<td>(-4.64)</td>
<td>(-2.57)</td>
<td></td>
</tr>
<tr>
<td>Price(_t) × Capacity(_t)</td>
<td>0.003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.83)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-2.984***</td>
<td>-0.358**</td>
<td>-2.237***</td>
<td>-1.924***</td>
</tr>
<tr>
<td></td>
<td>(-11.92)</td>
<td>(-1.97)</td>
<td>(-7.68)</td>
<td>(-4.01)</td>
</tr>
<tr>
<td>PseudoR(^2)</td>
<td>0.120</td>
<td>0.052</td>
<td>0.179</td>
<td>0.180</td>
</tr>
<tr>
<td>Observations</td>
<td>628</td>
<td>628</td>
<td>628</td>
<td>628</td>
</tr>
</tbody>
</table>
```

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 4 confirms a negative relationship between OPEC meetings frequency and the oil price.

Next, we use the following time-series model to go beyond the link to the price level and ask if the direction of the price change and its interaction with price level matters:

\[
Meeting Frequency\(_t\) = \beta_0 + \beta_1 Price\(_t-1\) + \beta_2 Price Jump\(_t-1\) + \beta_3 Price\(_t-1\) \times Price Jump\(_t-1\) + \varepsilon\(_{it}\) \tag{19}
\]

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
Figure 4: **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 prices 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 to change the quotas.

An indicator variable that captures oil price jumps. A positive jump is an increase in real oil price that exceeds 30%, a negative jump is a drop in real oil price that exceeds 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 $\beta_2$ when there is a negative price jumps. Moreover, if the drop in oil price matters more in L-state, we expect a negative and significant sign for $\beta_3$. 

31
Table 4: OPEC Meeting Frequency and Oil Market Conditions. This table presents the 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 (19). 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 the real oil 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>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Price_{t-1}$</td>
<td>-0.005*</td>
<td>-0.005*</td>
<td>-0.007</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(-1.76)</td>
<td>(-1.73)</td>
<td>(-0.83)</td>
<td>(-1.24)</td>
</tr>
<tr>
<td>$PriceJump_{t-1}$</td>
<td>-0.047</td>
<td>-0.313</td>
<td>1.989***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.11)</td>
<td>(-0.98)</td>
<td>(2.69)</td>
<td></td>
</tr>
<tr>
<td>$Price_{t-1} \times PriceJump_{t-1}$</td>
<td>0.006</td>
<td>0.003</td>
<td>-0.072**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.57)</td>
<td>(0.35)</td>
<td>(-2.46)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.514***</td>
<td>0.516***</td>
<td>0.776**</td>
<td>0.460***</td>
</tr>
<tr>
<td></td>
<td>(5.00)</td>
<td>(4.77)</td>
<td>(2.60)</td>
<td>(4.39)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.023</td>
<td>0.027</td>
<td>0.036</td>
<td>0.073</td>
</tr>
<tr>
<td>Observations</td>
<td>135</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
</tbody>
</table>

Table 4 presents the estimation results for equation (19). Consistent with Corollary 2 and similar to Figure 4, the first column of this table shows a significant and negative sign for $\beta_1$ indicating a negative relationship between OPEC meetings frequency and the oil price. Moreover, when we investigate the impact of the price dynamics on meeting frequency we find strong evidence for this effect. Columns (2), (3) and (4) show the estimation results for positive, negative and no jump scenarios. While the overall relationship between meeting frequency and the real oil price remains negative in all three columns, an important distinction between the results in column (4) and other two columns is made. The likelihood of OPEC meetings significantly increase in response to a negative price jump during low oil prices. This can be seen by the fact that the only time $\beta_2$ and $\beta_3$ are significant is in column (4) where there is a negative price jump. Moreover, the size of the estimated coefficient for $\beta_2$ 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 act more as a quota-driven cartel in L-states than in H-state, as stated by Corollary 2, especially when there is a price drop.

3.2.3. Non-compliance Pattern

In this section we analyze the non-compliance pattern for non-police (small) members. A size-dependent punishment strategy implies that the tolerance for non-compliance is higher for the smaller members. Thus, in equilibrium, we expect to observe a negative relationship between a producer’s size and non-compliance when it is measured in relative terms (i.e., overproduction as a fraction of quota). However, if the non-compliance is measured in absolute terms (i.e., the difference between quota and production), we either may observe a smaller pattern of non-compliance, or we may see a larger degree of non-compliance for the larger members simply because of their larger absolute production scale. To investigate this conjecture, we depicted the average non-compliance in relative terms vs. the average production capacity in a logarithm scale over the sample period in Figure 5. 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.

To examine this more formally, we use the following panel data model:

\[
NC_{it} = \beta_0 + \beta_1 Price_{it} + \beta_2 Price\ Increase_{it} + \beta_3 Capacity_{it}
\]

\[
+ \beta_4 Price\ Increase_{it} \times Capacity_{it} + \varepsilon_{it}
\]

where \( NC_{it} \) is the non-compliance by country \( i \) in period \( t \) in either relative or absolute terms; \( Price \) is the quarterly real oil price in 2001 dollars; \( Price Increase \) is the unexpected shift in oil price measured by an indicator variable that switches on when the real oil price exceeds its three-quarter moving average. As illustrated in Figure 5 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 decisions in response to a new change in market conditions. Finally, Capacity is a member’s production capacity in mbpd. A size-dependent punishment strategy that predicts higher non-compliance for the smaller countries implies a negative coefficient for Capacity, $\beta_3$, when non-compliance is measured in relative terms. Instead, if we measure non-compliance in absolute terms we may expect a negative sign for $\beta_3$ just as before or a positive sign simply because of the scale effect. The $\beta_4$ coefficient captures the asymmetry in the link between size and non-compliance in good (high demand) vs. bad (low demand) times of the oil market.

OLS estimation results for equation (20) are reported in Table 5. Columns (1) and (2) show the estimation results for non-compliance in absolute terms, whereas the results for
## Table 5: Non-compliance Pattern-OLS Model

This table presents regression results that describe non-compliance as a function of country size and market conditions. The model specification is given in equation (20). The dependent variable is $\log(\text{Production}_t - \text{Quota}_t + 1)$ (non-compliance is in absolute terms) in columns (1) and (2), and $(\text{Production}_t - \text{Quota}_t)/\text{Quota}_t$ (non-compliance is in relative terms) in columns (3) and (4). $\text{Production}$ and $\text{Quota}$ are in mbpd. $\text{Price}$ is the real oil price in 2001 dollars. $\text{Price Increase}_t$ is an indicator variable that switches on when the real oil price in 2001 dollars exceeds its three-quarter moving average. $\text{Capacity}_{it}$ is a member’s production capacity in mbpd. The sample includes all OPEC countries between 1995 and 2007 except Saudi Arabia. Numbers in parenthesis are t-statistics, and significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Absolute NC</th>
<th>Relative NC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Price</strong></td>
<td>-0.003***</td>
<td>-0.003***</td>
</tr>
<tr>
<td></td>
<td>(-3.18)</td>
<td>(-3.18)</td>
</tr>
<tr>
<td><strong>Price Increase_t</strong></td>
<td>0.005</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(1.28)</td>
</tr>
<tr>
<td><strong>Capacity_{it}</strong></td>
<td>0.043**</td>
<td>0.064***</td>
</tr>
<tr>
<td></td>
<td>(2.18)</td>
<td>(2.63)</td>
</tr>
<tr>
<td><strong>Price Increase_t \times Capacity_{it}</strong></td>
<td>-0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(-1.32)</td>
<td>(-0.72)</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>0.057</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(1.06)</td>
<td>(0.20)</td>
</tr>
<tr>
<td><strong>R^2</strong></td>
<td>0.0085</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>(2.14)</td>
<td>(1.26)</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>467</td>
<td>467</td>
</tr>
</tbody>
</table>

### Notes

- Absolute non-compliance in relative terms are reported in columns (3) and (4).
- Focusing on the third row, we observe a positive and significant coefficient of 0.064 for $\beta_3$ in column (2) shows that among non-police members, on average larger producers with an additional 1 mbpd capacity overproduce their quotas in absolute terms by 6.6% more than smaller members.
- In contrast, the results in columns (3) and (4) show no significant relationship between size and non-compliance in relative terms. Moreover, consistent with Corollary 1, a negative and significant value of -0.003 for $\beta_1$ in the first row indicates that absolute non-compliance is counter-cyclical among small producers. Similarly, a negative value of -0.001 in columns (3) and (4) for $\beta_1$ suggests that relative non-compliance is also counter-cyclical although not
One potential caveat with the OLS model specification in equation \((20)\) is the endogeneity between size and non-compliance. It’s not at all clear if the sign of the Capacity coefficient is driven by a mechanical relationship between non-compliance and a 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. Therefore, it might be the case that the result is just a reflection of the way quotas are assigned and does not necessarily reflect any behavioral decision regarding compliance. We address this issue by using an instrumental variable (IV) approach. This eliminates the direct link between the capacity on the left and right-hand sides of the equation \((20)\).

<table>
<thead>
<tr>
<th>Dependent Variable: Degree of Non-Compliance (NC)</th>
<th>Absolute NC</th>
<th>Relative NC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Price</td>
<td>-0.003***</td>
<td>-0.003***</td>
</tr>
<tr>
<td></td>
<td>(-3.44)</td>
<td>(-3.35)</td>
</tr>
<tr>
<td>Price Increase(_t)</td>
<td>0.006</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(0.66)</td>
</tr>
<tr>
<td>Capacity(_it)</td>
<td>-0.021</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(-0.82)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Price Increase(_t) \times Capacity(_it)</td>
<td>-0.016</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(-0.61)</td>
<td>(-0.46)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.200***</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>(3.11)</td>
<td>(1.24)</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.0251</td>
<td>0.0304</td>
</tr>
<tr>
<td>Observations</td>
<td>467</td>
<td>467</td>
</tr>
</tbody>
</table>

Table 6: **Non-compliance Pattern-IV Model.** This table presents regression results that describe non-compliance as a function of country size and market conditions. The model specification is given in equation \((20)\). The dependent variable is \(Log(Production\(_t\) - Quota\(_t\) + 1)\) (non-compliance is in absolute terms) in columns (1) and (2), and \((Production\(_t\) - Quota\(_t\))/Quota\(_t\)\) (non-compliance is in relative terms) in columns (3) and (4). Production and Quota are in mbpd. Price is the real oil price in 2001 dollars. Price Increase is an indicator variable that switches on when the real oil price in 2001 dollars exceeds its three-quarter moving average. Capacity is a member’s production capacity in mbpd. The sample includes all OPEC countries between 1995 and 2007 except Saudi Arabia. Numbers in parenthesis are t-statistics, and significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.
We use variations in oil reserves as an instrument for changes in size (measured by capacity). The unreported first stage estimation results suggest a strong relationship between reserves and capacity with a $p$-value $< 0.001$. The estimation results for the IV model are reported in Table 6. Consistent with the notion of a size-dependent punishment strategy, we observe a negative and significant relationship between size and relative non-compliance. The results in columns (3) and (4) show that on average for a larger member with an additional capacity of 1 mbpd the non-compliance is 5% lower. This confirms our earlier conjecture and is consistent with the findings in Figure 5. In contrast, we observe no size impact on non-compliance in absolute terms. The results in columns (1) and (2) in Table 6 suggest that the coefficient of $Capacity$ is not statistically significant. Both of these results are contrary to the OLS model which predicted a positive size impact on absolute non-compliance, but no impact for relative non-compliance. This suggests that the concerns regarding the endogeneity between capacity and non-compliance are valid, and therefore one should trust the results based on the IV model more than the ones from the OLS model.

As for the impact of market conditions on non-compliance, the results of the IV and OLS models are comparable. Consistent with Corollary 1, the negative and significant value of $-0.003$ for $\beta_1$ in columns (1) and (2) and $-0.001$ in columns (3) and (4) in Table 6 suggest that non-compliance, in both absolute and relative terms, is counter-cyclical among small members. Moreover, although not significant, the negative sign of $\beta_4$ in columns (2) and (4) is also consistent with Corollary 1. In response to a positive oil price shock, smaller countries have larger degrees of non-compliance.

Overall, consistent with the implications of a size-dependent punishment strategy, the results in Tables 5 and 6 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.
4. 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 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 a 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 active during low price episodes. We also show that the likelihood of producing at capacity will be much higher in the high demand states. 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) 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 the OPEC cartel in low and high demand states.

Second, the model can be extended by choosing the severity of the punishment mechanism (i.e. the probability of punishment) endogenously. The police understand the trade-off that small producers face between building additional capacity and deviating from their allocated quota. Given the parameters of the model, there could be an optimal stringency level of the punishment mechanism that would balance the cost of spare capacity (built 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 could examine the optimal choice of punishment mechanism as a function of background variables.

Finally, the optimal level of 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. A full characterization is beyond the scope of the current paper. Future research would provide additional insights regarding the optimal choice of capacity for a cartel police.
Appendix A. Details of Calculations

Appendix A.1. One-Short Deviation

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

$$\pi(q^*, X_s) = q^*(X_s - b(\bar{Q} + q^*)) \quad (A.1)$$

The FOC is:

$$q^* = \frac{X_s - b\bar{Q}}{2b} \quad (A.2)$$

The equilibrium price will be:

$$P = X_s - b(\bar{Q} + \frac{X_s - b\bar{Q}}{2b}) = \frac{X_s - b\bar{Q}}{2} \quad (A.3)$$

The profit is

$$\pi(q^*, s) = \left[\frac{X_s - b\bar{Q}}{2b}\right]^2 \quad (A.4)$$

Appendix A.2. Value Function of Punishment State

The value function in the punished regime is:

$$V_{N,i} = \mathbb{E}[\pi_{i}\text{Cournot}]/r \quad (A.5)$$

$$\max_{q_i} \pi_i = [X - \gamma(q_i + q_j)]q_i - C \quad (A.6)$$

$$q_i^H = K_i = \frac{p_H\phi_j X_H}{\phi_i\phi_j + 2p_H\gamma(\phi_i + \phi_j)} \quad (A.7)$$
\[
q^L_i = \frac{X_L}{3\gamma} \rightarrow \pi_L = \frac{X^2_L}{9\gamma} - C \tag{A.8}
\]

\[
\begin{cases}
V_H = \pi_H + \left(\frac{1}{1+r}\right)(p_{H,H}V_H + p_{H,L}V_L) \\
V_L = \pi_L + \left(\frac{1}{1+r}\right)(p_{L,L}V_L + p_{L,H}V_H)
\end{cases} \tag{A.9}
\]

\[
\begin{cases}
V_H = \frac{\pi_H \left(1 - \frac{p_{L,L}}{1+r}\right) + \left(\frac{X^2_L}{9\gamma} - C\right) \left(\frac{p_{H,L}}{1+r}\right)}{\left(1 - \frac{p_{L,L}}{1+r}\right) \left(1 - \frac{p_{H,H}}{1+r}\right) - \left(\frac{p_{H,L}}{1+r}\right) \left(\frac{p_{L,H}}{1+r}\right)} \\
V_L = \frac{\left(\frac{X^2_L}{9\gamma} - C\right) \left(1 - \frac{p_{H,H}}{1+r}\right) + \pi_H \left(\frac{p_{L,H}}{1+r}\right)}{\left(1 - \frac{p_{L,L}}{1+r}\right) \left(1 - \frac{p_{H,H}}{1+r}\right) - \left(\frac{p_{H,L}}{1+r}\right) \left(\frac{p_{L,L}}{1+r}\right)}
\end{cases} \tag{A.10}
\]
Appendix B. Figures

Figure B.6: **Price Increase Indicator vs. Historical Oil Price.** This figure presents the historical oil price in 2001 dollars between 1995 and 2007 and an indicator variable (Price Increase) that switches on in quarters when oil price exceeds its three-quarter moving average.


