Market Power, Production (Mis)Allocation and OPEC*

John Asker
UCLA
NBER

Allan Collard-Wexler
Duke University
NBER

Jan De Loecker
KU Leuven and Princeton University
NBER and CEPR

September 6, 2017

Abstract

This paper estimates the extent to which market power is a source of production misallocation. Productive inefficiency occurs through more production being allocated to higher-cost units of production, and less production to lower-cost production units, conditional on a fixed aggregate quantity. We rely on rich micro-data covering the global market for crude oil, from 1970 to 2014, to quantify the extent of productive misallocation attributable to market power exerted by the OPEC. We find substantial productive inefficiency attributable to market power, ranging from 14.1 percent to 21.9 percent of the total productive inefficiency, or 105 to 163 billion USD.

Key words: *Market Power, Productive Inefficiency, Misallocation, Cartels, Oil, OPEC.*

JEL Code: *D2, L1, L4, L72*

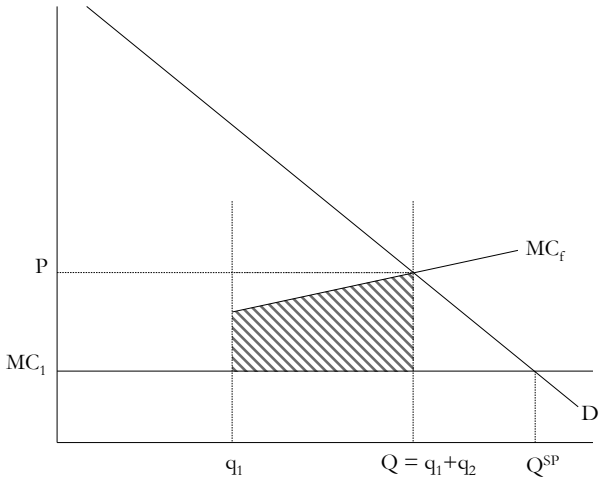
*We would like to thank Russell Cooper, Martin Hackmann, Volker Nocke, Michael Whinston, Rob Porter, and Jo Van Biesebroeck, and participants in various seminars and conferences for their comments. Excellent research assistance was provided for by El Hadi Caoui, Ardis Zhong, Sherry Wu and Yilin Jiang. Financial Assistance from Princeton University's Industrial Organization Group, NSF (SPS #220783) and FWO Odysseus Grant is greatly appreciated. Contact information: johnasker@econ.ucla.edu, collardwexler@gmail.com and jan.deloecker@kuleuven.be.

1 Introduction

This paper provides estimates of the extent to which market power is a source of production misallocation. This production misallocation, and the resulting welfare loss, occur through more production being allocated to higher-cost (less productive) units of production, and less production to the most efficient production units in the economy. That is, to borrow from Harberger (1959), we measure the extent to which the exercise of market power inhibits a society's ability to use "the resources at hand more effectively" to achieve observed output levels.

In quantifying the potential magnitude of this productive inefficiency, we examine the global market for crude oil, in which OPEC, a notable international cartel, is alleged to exert considerable market power at times. We conduct our analysis by leveraging detailed field-level data on production, costs and reserves for all significant oil fields across the globe from 1970 through 2014.

Figure 1: Production distortions due to market power



Notes: q_1 indicates total production from the cartel, while q_f indicates production from the competitive fringe. The cartel has marginal costs of MC_1, while the fringe has the marginal cost schedule of MC_f. Q^{SP} is the social planner's quantity.

Figure 1 presents a stylized, graphical, representation of the environment we examine. Consider a producer with market power, with constant marginal cost MC_1. Also present in the market is a competitive, price-taking, fringe that has an aggregated marginal cost curve given by MC_f. The market price is equal to P, and the quantity produced by the (low-cost) producer with market power, q_1, is less than total production $Q = q_1 + q_2$, where q_2 is the production of the fringe.[1] In this setting, the production done by the fringe, q_2, is done at a higher resource cost than is socially optimal: Indeed, the low-cost producer should do all the production. The welfare cost of this production misallocation is the shaded area.

[1]Strictly speaking, this means that MC_f measures the marginal cost of the $q_2 - q_1$'th unit of production of the fringe firms.

The objective of this paper to identify and quantify the size of (the analog of) this shaded trapezoid in the global oil market. In the case of the oil market, this calculation is more complex, due to the finite resource extraction problem embedded in oil production. This creates intertemporal linkages of supply and cost. By leveraging rich micro-data and a flexible dynamic framework, the productive inefficiency can be computed accounting for these dynamics.

Economists have well understood the potential negative impact of market power on resource allocation, and the associated welfare cost at least since the work of Harberger (1954). Harberger famously concluded that rates of return on capital across US (manufacturing) industries during the 1920s were not sufficiently dispersed to generate any meaningful aggregate distortions attributable to market power. The intuition for this inference is that market power operates like a tax, where the implicit tax rate is reflected in the rates of return on capital (profits). If these rates are equal, then there can be no scope for misallocation of the incremental resource (production) unit. This analysis, along with its conclusion that market power scarcely impacts economy-wide outcomes, became the default view held by many economists and policy makers in the years that followed.[2]

Harberger's analysis is also notable for its focus on welfare triangles. This focus persists in much contemporary work on market power, which has tended to concentrate on impacts on consumer surplus and deadweight loss (for a survey of some recent work in I.O., see Ackerberg et al. (2007)). The welfare impact of market power on production allocation across heterogeneous production units has generally been less emphasized in the literature. An important exceptions is the work of Borenstein et al. (2002) on California electricity markets, which is closely related to the approach taken in this paper. There is also a small empirical literature that documents the effects of cartels on productivity, such as Bridgman et al. (2015) on the government-sponsored beet sugar cartel.[3] Also notable is the literature on industry performance and reallocation – pioneered by Olley and Pakes (1996) – in which industry-wide performance is connected to the allocation of producer-level market shares. That literature makes no distinction between cost and markups, and this is essential for any welfare analysis in the Olley-Pakes setting. Our study, like that of Borenstein et al. (2002), avoids this issue by leveraging detailed micro data on costs.

The relative lack of focus on productive inefficiency also seems a natural consequence for the onerous data requirements in quantifying productive inefficiency. Measurement requires the ability to observe reliable measures of marginal cost of production for each producer, ex ante knowledge of (potential) market power abuse, and the identity of the firms engaged therein. The arrival of producer-level data, recording the cost of production across quantities within, and across producers over time and production units, frees up the constraint faced by the analyses of Harberger (1954) and much of the work on the effects of market power that

[2]Harberger's conclusion is summarized as follows: "When we are interested in the big picture of our manufacturing economy, we need not apologize for treating it as competitive, for in fact it is awfully close to being so. On the other hand, when we are interested in the doings of particular industries, it may often be wise to take monopoly elements into account."

[3]There is a a broader empirical literature documenting the effects and functioning of cartels, such as Asker (2010); Bresnahan (1987); Byrne and de Roos (2016); Fink et al. (2015); Hyytinen et al. (2011); Genesove and Mullin (1998, 2001); Roller and Steen (2006); and some good overviews of the cartel literature can be found in Harrington (2006); Marshall and Marx (2012); Porter (2005) and Whinston (2006) .

has followed. This enables us to look inside the total cost of production of an industry, and to relate the potential dispersion in cost to potential market power abuses.

A predicate for production misallocation is inter-firm heterogeneity in the costs of production. Both production misallocation and firm cost heterogeneity have received considerable attention recently in many area of economics. On the cost side, the literature has established that firms that compete in narrowly defined product markets can have very different levels of productivity (see Syverson (2011, 2004); Foster et al. (2008); Collard-Wexler (2011)). This heterogeneity has proved to be key in predicting the impact of competition on many outcomes, such as the effect of trade or of a new technology on firm and industry performance.[4]

Production misallocation has also been extensively studied in the recent macro-misallocation literature, which finds evidence for substantial misallocation in plant-level data, based on very different input use across producers, and large differences in productivity. This literature has highlighted the potential aggregate returns from improving the apparent misallocation of resources within an economy (See, for a survey, Hopenhayn (2014)).[5] The ongoing macroeconomic literature on misallocation has considered a variety of distortions that affect the allocation of inputs across plants. However, this misallocation literature typically relies on a combination of a demand system (CES) and a market structure (monopolistic competition) that largely eliminate the effects of market power. In this paper, we focus exclusively on the role of market power, while allowing for other sources of dispersion in the cost of production in any given period.

To assess the impact of market power on production misallocation, we leverage the several specific features of the world oil market. First of all, oil is a relatively homogeneous product, and the presence of differences in the observed cost of production is informative about differences in welfare-relevant resource costs. Second, oil is extracted across a wide range of countries, using a variety of technologies, both of which give rise to natural cost differences. Third, this market has been affected over a long period of time by the presence of a well-known cartel (OPEC), whose internal organization is somewhat transparent. Fourth, we directly observe the cost of production by oil field, and, as such, we do not have to model the extent of market power, or how exactly the cartel is (dis)organized; rather, we only need to simulate the counterfactual aggregate supply curve in an environment absent market power (a market in which firms are price takers).

That is, in terms of Figure 1, the object to be estimated is the proportion of production that the low-cost firm would do if it were a price taker. This can then be compared to observed production patterns to back out the shaded rectangle. Note that this does not require any structural modeling of the manner in which market power is exercised – all that is required is knowledge of who has the market power (OPEC in our setting), their costs and a model of price-taking behavior. By holding the (observed) market-level output fixed, this approach avoids having to model the existence and workings of the cartel, which, in the context of the

[4]See Goldberg et al. (2010); De Loecker et al. (2016); Melitz (2003); Edmond et al. (2015); Holmes and Schmitz (2010); Atkeson and Burstein (2010); Syverson (2004); Collard-Wexler and De Loecker (2015); De Loecker (2011)' and Olley and Pakes (1996) .

[5]Some recent papers in this literature are Asker et al. (2014); Bartelsman et al. (2013); Hsieh and Klenow (2009) and Restuccia and Rogerson (2008).

world oil market, is a complicated matter. As a consequence, this paper is intentionally agnostic regarding the relative efficacy of the OPEC cartel, in terms of its methodological approach. Our analysis, however, sheds light on the (in)efficiency of the cartel by inspecting the cost differences, and the associated allocated output, across the members.

In the data, our estimates of productive inefficiency are founded on the location of high- and low-cost fields in the world. The preponderance of the world's low-cost oil reserves are in OPEC countries, while most of the world's high-cost deposits are outside OPEC. For instance, the world's largest oil field, the Ghawar field in Saudi Arabia, has costs of approximately $3 per barrel. By contrast, offshore fields in Norway and fracking shale deposits in the Bakken in North Dakota, have costs of $12 and $24, respectively, per barrel. Moreover, the fields in OPEC countries have very large reserves and are depleted relatively slowly: in 2014 Saudi Arabia's active fields had 17 percent of global recoverable reserves and were being depleted at close to half the speed of the mean non-OPEC field.[6] This implies that production was being diverted toward high-cost productive units, while low-cost productive units were being utilized at comparatively lower rates – with discounting, this results in a welfare loss. Different specifications of our model put the productive inefficiency due to OPEC's market power at, in NPV terms, between 103 and 163 billion 2014 USD.

Providing evidence on the impact of market power, and the extent to which it may have aggregate implications, contributes to a range of conversations across economics.[7] Of these, likely the most far-reaching discussion relates to why some countries are poorer than others and how poorer countries can catch up to the richer countries. The existing misallocation literature addresses this question using a 'top-down' approach, discussing the macro potential of various mechanisms to improve productive efficiency.

By comparison, this paper takes a more 'bottom-up' approach by examining a specific market and looking for sources of misallocation that may have aggregate significance - the implication being that if the effects can be large in one (large) market, then similar conduct across many other sectors in an economy can have similarly significant aggregate consequences. By focusing on a globally significant industry, and measuring the production inefficiencies resulting from its cartelized state, this project takes a preliminary step in understanding the potential for endemic market power abuses to retard growth in both more and less developed countries.

This paper is organized as follows. In Section 2 presents a short description of the oil market, and introduces the unit of observation used throughout the analysis. Section 3 introduces the theoretical structure common to the entire paper. The preliminary evidence of the role of market power is presented in Section 4 by means of reporting details of the cost distribution, production and reserves across units within countries and regions. Section 5 presents the main results, and presents various robustness checks. Alternative modeling choices are discussed and evaluated in Section 6, and Section 7 concludes.

[6] OPEC had 50% of reserves and were being depleted at a slower rate than in the rest of the world. See Table 3.

[7] De Loecker and Eeckhout (2017) discuss the impact of market power on various macroeconomic outcomes, such as productivity growth and labor markets.

2 The Oil Market: Production and Institutions

This section introduces some of the important institutional details of the global oil market. In particular, these features of the upstream oil industry are important for understanding the measurement issues that arise in handling the data. As a consequence, in what follows, we introduce production units and the market-level institutions.[8]

2.1 Unit of analysis

The analysis in this paper focuses on the upstream oil industry (that part of the industry concerned with extraction), as opposed to activity further downstream (such as refining). Data on the upstream oil industry were obtained from Rystad Energy (Rystad hereafter), an energy consultancy based in Norway that covers the global oil industry.

The data record all significant oil fields across the globe from 1970 through 2014, and as such, constitute an unusually rich dataset compared with most studies of the oil market, which either use detailed micro data on a small subset of oil fields (see Covert (2015)'s or Kellogg (2014)'s study of recent activity in North Dakota Shale and Texas or Hendricks and Porter (1988)'s earlier work on offshore oil in the Gulf of Mexico), or examine the global oil market with data aggregated to the country level (see, for example, Kilian (2009)). For each field, the data include annual production, reserves and a breakdown of operating and capital costs, as well as the characteristics of the field, such as the location, geology and climate zone. The distinction between a production unit (field) and its smaller components (wells) is important since, in our data, we observe cost and production information at the field level. A field, in the data, is defined as a geologically homogeneous area with the same management and ownership structure. Fields vary considerably in the number of wells and the associated infrastructure.[9]

The fact that the data cover all oil fields in the world implies that there is some heterogeneity across oil crudes produced at various locations. This leads to a series of measurement issues. The first is how to measure the quantity associated with a deposit in terms comparable across deposits. The data measure output in energy equivalent barrels, where the benchmark is one barrel of Brent Crude. Hence, the measure of quantity accounts for the compositional heterogeneity of crudes. The second issue is that different crudes trade at different premia and discounts related to their composition. Thus, the choice of a price index needs to be consistent with the measure of quantity. The price of Brent Crude is the price measure used here to be consistent with the production measure.[10]

Production units (oil fields) can have very different costs for exogenous (geological) reasons. That is, a Norwegian deposit that exists in deep water far offshore or a Canadian tar

[8]The Online Appendix provides the reader with a more detailed discussion of the data sources, measurement and on the specifics of oil production.

[9]For instance, in the data, the Gullfaks offshore field in Norway is decomposed into two separate oil fields; Gullfaks, which has three oil platforms, and Gullfaks South, which has a single platform. On the other hand, the Ghawar Uthmamiyah onshore field, which is one of the largest fields in the world, is composed of many hundred wells. Different fields can, of course, be owned by a single owner.

[10]The unit cost of production of a field is strongly negatively correlated with the price of the oil it produces. That is, low-quality oils tend to come from high-cost fields. See the Online Appendix for an extended discussion of oil quality.

sands deposit will have very different average (equivalently, marginal) costs of production as compared to the larger onshore deposits in Saudi Arabia, for purely geological reasons. This means that the vast proportion of the cost differences observed across time and field are predetermined by geology. This is the fundamental starting point for the analysis in this paper, in which the costs of a field are taken to be exogenous. Section 4.1 contains further details regarding measurement of the cost of production.

We now turn to a brief overview of market-wide conditions in the oil market, with a particular focus on the role of OPEC.

2.2 The global oil market and OPEC

The global upstream market for oil is characterized by a range of actors. The buyers are refineries. The producers are oil companies, which are state-run enterprises, substantially-state-run, or independent enterprises. The state-run (nationalized) oil companies, can be split into those that are run by OPEC states and those that are from non-OPEC states. Every OPEC country has its own nationalized company, which controls production, albeit at times contracting with independents to run specific facilities. For instance, Saudi Arabia operates Saudi Aramco; Kuwait operates the Kuwait Petroleum Company; and Ecuador operates Petroecuador.

Outside OPEC, nationalized (or substantially-state-run) companies exist in Mexico, Brazil, Russia, China, Malaysia, Norway and India, and in several other smaller producing nations. In other major producing countries (such as the USA, the UK or Canada), production is conducted by private (independent) companies. These private companies can be divided into the five (as of 2014) oil majors (ExxonMobil, Chevron, BP, Royal Dutch Shell and Total) – all having revenues in excess of 100 billion US dollars – and other independent companies.

Table 1 shows the production shares, for the period 1970-2014, of the seven largest OPEC and non-OPEC countries. The US has the largest production, followed closely by Russia and Saudi Arabia. While these three countries have the largest production, it is important to bear in mind that production occurs in different ways within each country. The US is very decentralized, having many private firms, while Saudi Arabia has a nationalized oil company (Saudi Aramco).

In 2014 (the limit of the data) OPEC comprised the countries of Algeria, Angola, Ecuador, Indonesia, Iran, Iraq, Kuwait, Libya, Nigeria, Qatar, Saudi Arabia, UAE, and Venezuela. The membership has varied slightly over time, with the core Middle East membership being unchanged from OPEC's inception in 1960.[11] Due to the relative stability of the OPEC membership and the likely close affiliations that may persist during periods of a country's non-membership, in handling the data we treat a country as an OPEC country if it had active membership between 1970 and 2014.

OPEC characterized its objective, in 2017, as being to "co-ordinate and unify petroleum

[11]The original membership in 1960 was Iran, Iraq, Kuwait, Saudi Arabia, and Venezuela. Other members are listed together with the year they first joined OPEC, and (if appropriate) years in which membership was suspended or terminated: Qatar (1961), Indonesia (1962, suspended 1/09), Libya (1962), the United Arab Emirates (1967), Algeria (1969), Nigeria (1971), Ecuador (1973, suspended 12/92-8/07), Gabon (1975, terminated 1/95) and Angola (2007). See `www.opec.org/opec_web/en/about_us/25.htm` accessed 29 August 2016. The first 30 years of OPEC are well documented in Yergin (1991) and Crémer and Salehi-Isfahani (1991).

policies among Member Countries, in order to secure fair and stable prices for petroleum producers; an efficient, economic and regular supply of petroleum to consuming nations; and a fair return on capital to those investing in the industry".[12] Given this description, and its well documented history of coordinated price and production policies, this paper views OPEC as a cartel, albeit one that has varied in its effectiveness.

Figure 2 shows OPEC's market share and the price of crude from 1970 to 2014. Before OPEC started coordinating extensively on price reductions, it had a global production share fluctuating around 48 percent. This fell to a low point of 29.2 percent in 1985 after reductions in production during the late 70s and early 80s. Following that, OPEC's share of production rose to 40.6 percent in 1993 and has stayed relatively constant since then.

For its first ten years, OPEC was ineffective in moving the oil price. This changed in 1973 when, following the outbreak of war in the Middle East, the OPEC member countries were successful in raising the oil price four-fold, albeit with (according to secondary sources) little coordination among themselves. Until 1982, the cartel did little more than try to set price guidance for its members, with little hope that they would comply.

In March 1982, OPEC took its first step in evolving toward its modern form: the cartel introduced country-specific production quotas for the first time. Despite the initial quotas being ineffective, quotas have continued to be a defining feature of OPEC's operation. Notably, in 1985, after becoming frustrated with defections by other OPEC members, Saudi Arabia expanded production in an effort to discipline production. This might have worked had it not been for a slowdown in oil demand in the 1980s and the first Gulf War in the early 1990s.

During the 1990s expansion in oil demand outstripped the ability of non-OPEC producers to expand supply and the power of the OPEC cartel appears to have increased.[13] By November 1997, the cartel was in a position to exert substantial influence on the market price, but appears to have either broken down or misjudged global demand, expanding production to coincide with the Asian crisis. This led oil prices to decline from $35 per barrel to $10 per barrel.[14] Finally, in March 1999, with the cooperation of Russia, Norway and Oman, OPEC countries were able to cut back production and move the price back into the $30-$40 per barrel range.

In 2000, the cartel explicitly announced a target price band of $22-$29 per barrel. The quota system was adjusted to include automatic adjustments should the reference price fall outside this interval. This heralded the modern era of OPEC and the most sophisticated coordinating mechanism seen to date. Unfortunately for the cartel, the second Gulf War and an expansion in demand from developing countries, in particular, made the price band unsustainable. In January 2005, it was suspended. The quota system, however, lived on.[15] Although the quotas are not immediately transparent, it is clear that they are highly asymmetric.[16] By July 2008, the Brent spot price had increased to an intra-day high of over $140 per barrel, a level in stark contrast to the price band in (ineffective) operation a mere four years earlier. Through to the end of 2014, the oil price remained high by historical standards.

[12] www.opec.org accessed 4/10/17.

[13] See Kohl (2005) for a description of this period.

[14] All prices in the text of this subsection are nominal (not deflated).

[15] See, for instance, Kohl (2005) and Fattouh (2007).

[16] See various monthly publications of the Monthly Oil Report (Organization of Petroleum Exporting Countries (2008)).

The efficacy of the OPEC cartel over time has been, at best, variable. Especially in the early 2000s, its policies mirrored those of collusive structures seen in other industries. In other periods, the ability of OPEC to coordinate its members' production seems less clear. In this paper, the question of whether OPEC is best characterized as a political vehicle for Saudi Arabia and other gulf countries, or as a long-running industrial cartel, does not require an answer. To the extent that an OPEC member has market power and distorts production, the measurement approach adopted in Section 5 will account for it. This is a consequence of working with cost data directly. By contrast, if the more standard IO approach were adopted, in which demand estimates were combined with a pricing model (embedding specific behavioral assumptions about OPEC conduct) to back out costs, then this issue would likely be a significant hurdle, if not fatal, to the credibility of any estimates.

3 Analytical framework

In a static environment, the definition of a productive inefficiency is intuitive: as Figure 1 illustrates, it is the difference between the realized cost of production and the cost of producing the same quantity, had all firms been price takers. In the empirical setting confronted here, a purely static approach is inappropriate due to the finite nature of oil extraction.[17] Thus, we need to adopt a definition of productive inefficiency appropriate for a dynamic context.

Definition 1 *Productive inefficiency is the net present value of the difference between the realized costs of production, and the cost of production had the realized production path been produced by firms taking prices as exogenous.*

That is, the competitive benchmark is derived by holding the production in each year fixed and shifting demand for that year inward until a competitive industry would have produced, in equilibrium, the observed production. The path of costs of production thus generated is the counterfactual benchmark against which realized costs are compared to measure the extent of any production inefficiency due to market power.

Given the finiteness of the resource, it is clear that, at some finite end date, all resources will have been extracted. Hence, the source of inefficiency, in an industry such as this, is via sub-optimal inter-temporal substitution of production among production units. Given this, the central economic object of interest is the order of extraction of assets that a competitive industry would have undertaken. The central result of this section is to provide a characterization of that order.

In addition to characterizing the extraction policy of the counterfactual competitive industry, this section builds the underlying cost function that is used to guide measurement and modeling. It also provides the algorithm used to compute the competitive solution to the social planner's solution. As usual, the data and empirical setting impose some additional measurement issues that are discussed in the sections directly related to empirical analysis. This section focuses on the details of the theoretical structure common to the entire paper.

[17] There is a long literature in natural resource economics on non-renewable resources, starting with Hotelling (1931), and some of the empirical tests of this model for oil are documented in Slade and Thille (2009) and Anderson et al. (2014).

3.1 Modeling Preliminaries: Costs

In modeling costs, the production unit is the field, denoted f, that is the unit of observation in the most disaggregate data to which we have access to. For some of these fields, such as some offshore oil platforms, a field is an oil well. However, for most of the onshore oil fields, a field is composed of many different oil wells. Fields make input choices in order to minimize costs, conditional on a given level of production.

Let the production function for a field f in year t, be given by:

$$\begin{aligned} q_{ft} \quad &= \quad \min\left\{\alpha_{ft}K_{ft}, \gamma_{ft}L_{ft}\right\} \\ &\text{s.t.} \quad q_{ft} \leq R_{ft},\ R_{ft} = R_{ft-1} - q_{ft-1} \\ &\qquad R_{f0} > 0\ R_{ft} \geq 0, \end{aligned} \tag{1}$$

where K and L are fixed and variable inputs, respectively, and R are reserves. We write down the model with capital and labor inputs (K and L), but of course these are meant to stand in for the different inputs in the production process for oil, such as drilling equipment, production workers, and energy. These coefficients are field-specific, and as such, subsume the differences across technology (onshore, offshore, shale, etc.). The fact that the coefficients are allowed to vary across fields also implies that they capture any Hicks-neutral productivity shocks ω_{ft}.[18]

Assume that the price of capital inputs is given by r_{ft} and the price of variable inputs is given by w_{ft}. These input prices are assumed to be exogenous. This means that the total cost of production, assuming cost minimization at the field level, is given simply by:

$$C\left(q_{ft}\right) = \left(\frac{w_{ft}}{\gamma_{ft}} + \frac{r_{ft}}{\alpha_{ft}}\right) q_{ft}. \tag{2}$$

Additional structure is put on the process governing the evolution of the ratio of input prices to the technology parameters such that

$$\frac{w_{ft}}{\gamma_{ft}} = \frac{w_f}{\gamma_f}\mu_{ft} \quad \text{and} \quad \frac{r_{ft}}{\alpha_{ft}} = \frac{r_f}{\alpha_f}\mu_{ft}. \tag{3}$$

This allows for variation in either field (Hicks-neutral) productivity, or common variation across the ratio of input prices to technology, or a combination of both.

This yields the following cost function:

$$C\left(q_{ft}\right) = \left(\frac{w_f}{\gamma_f} + \frac{r_f}{\alpha_f}\right) \mu_{ft} q_{ft}. \tag{4}$$

Unit cost is then given by:

$$c_{ft} = MC\left(q_{ft}\right) = AC\left(q_{ft}\right) = \begin{cases} c_f \mu_{ft} & \text{if } q_{ft} \leq R_{ft} \\ +\infty, & \text{otherwise,} \end{cases} \tag{5}$$

[18] That is, the production function could have been written as $q_{ft} = \min(\{\alpha_{ft}K_{ft}, \gamma_{ft}L_{ft}\}, \omega_{ft})$.

where $c_f \equiv \left(\frac{w_f}{\gamma_f} + \frac{r_f}{\alpha_f}\right)$. That is, costs have a hockey stick shape: constant marginal costs up to a capacity constraint given by reserves. From a measurement point of view, the constant returns to scale assumption on the components of the Leontief production function provides economic assumptions under which average cost and marginal cost are equal, and, thus, costs are hereby invariant to changes in demand conditions.

We further assume that μ_{ft} is governed by a martingale process such that $E\left(\mu_{ft+k}\left|\mu_{ft}\right.\right) = \mu_{ft}$ for $k \geq 1$.[19] This μ_{ft} term captures the convolution of long-run trends in technological change, and changes in the absolute or relative cost of inputs or technology parameters (γ and α). The process determining μ_{ft} is assumed to be exogenous, which is an assumption with economic content and underscores the partial equilibrium nature of the exercise being conducted here. In an alternate, broader, context, μ_{ft} is an equilibrium object. In particular, if lower-cost fields get extracted first in the competitive counterfactual (as is the case), and these lower-cost fields have lower input intensity, and the inputs are specialized, such that they are not readily deployable in some other sector, then this reallocation of production could change the equilibrium value of μ_{ft}.

3.2 Production paths in competitive equilibrium

In competitive equilibrium, all producers take prices as given. Let δ be the common discount factor. Thus, for a given price path (or expectation thereof), a price-taking producer solves the following problem:[20]

$$\begin{aligned} & \max_{\{q_{ft}\}} \sum_{t=1}^{T} \delta^{t-1} \left(p_t - c_f\right) q_{ft} \\ \text{s.t.} \qquad & R_{f0} \geq \sum_{t=1}^{T} q_{ft}, \quad \text{and} \quad q_{ft} \geq 0 \quad \forall t \in \{1, \cdots, T\}. \end{aligned} \tag{6}$$

Proposition 1 and corollary 1 together establish that the lowest-cost fields are extracted first in any competitive equilibrium.[21]

Proposition 1 *Let marginal costs be described by equation 5. Consider two fields, $\underline{F}$ and $\overline{F}$, with c_f equal to $\underline{c}$ and $\overline{c}$, respectively. In any competitive equilibrium, if $\underline{c} < \overline{c}$, then if $\underline{R}_t > 0$, $\overline{q}_t > 0$ implies that $\underline{q}_t > 0$.*

Proof. Toward a contradiction, assume not. Consider two periods such that, w.l.o.g., $t = t_2 > t_1 = 1$. Consider a single unit of production for both $\underline{F}$ and $\overline{F}$, such that $\underline{q} = \overline{q} = 1$ (since marginal costs at the field level are constant, this is w.l.o.g.). Employ the normalization $\mu_{f1} = 1$. Hence, $E\left(\mu_{ft}\left|\mu_{f1}\right.\right) = 1$. Thus $E\left(c_f\mu_{ft}\left|c_f\mu_{f1}\right.\right) = c_f$. We assume by contradiction that $\overline{q}_1 = 1$ and $\underline{q}_1 = 0$. Then, there must exist periods such that

[19] This implies that we assume the same process for both Hicks-neutral productivity shocks and input prices.

[20] If prices are not known, it is assumed that all producers have the same expectations. In the maximand, the price process is assumed to be known merely to keep notation simple. There are no fixed or setup costs. A treatment thereof is in Section 6.2.

[21] The proposition and corollary are stated so as to align with the empirical model taken to the data. In fact, these results can be stated somewhat more generally. The key feature is to note that the proof can be applied at the barrel level. An implication is that, provided that the cost function is multiplicatively or additively separable in μ_{ft}, non-constant (increasing) marginal costs can be accommodated at the field level – i.e., the functional form assumptions on the production function can be relaxed.

$$\delta^{t-1}(p_t - \underline{c}) \geq (p_1 - \underline{c}) \tag{7}$$

and

$$\delta^{t-1}(p_t - \overline{c}) \leq (p_1 - \overline{c}), \tag{8}$$

where at least one inequality is strict. Assume, for exposition, that the inequality in equation 8 is strict.

From equation 7,

$$\delta^{t-1}(p_t - \overline{c}) + \delta^{t-1}(\overline{c} - \underline{c}) \geq (p_1 - \overline{c}) + (\overline{c} - \underline{c}), \tag{9}$$

Since $\delta^{t-1}(\overline{c} - \underline{c}) < (\overline{c} - \underline{c})$, this implies that $\delta^{t-1}(p_t - \overline{c}) \geq (p_1 - \overline{c})$, which is a violation of equation 8. ■

Corollary 1 *In any competitive equilibrium, if $\underline{c} < \overline{c}$, then if $\overline{R}_t > 0$, $\underline{q}_t > 0$ does not imply $\overline{q}_t > 0$.*

Proof. This follows the line of argument used above, noting that $\delta^{t-1}\left(\underline{C} - \overline{C}\right) > \left(\underline{C} - \overline{C}\right)$. ■

An immediate implication is that, in the absence of other distortions, when low-cost fields are not being exploited prior to higher-cost fields coming on line, then this is an indication of the exercise of market power. As is usual, firms with market power have an incentive to delay extraction to push prices higher. Any residual demand that results will be absorbed by fringe producers, with higher unit costs.[22]

Unsurprisingly, given the first welfare theorem, the production plan resulting from the competitive equilibrium coincides with that of the social planner that seeks to minimize the social cost of producing that production plan.

Lemma 1 *The social planner's production plan, which minimizes the net present value of costs subject to satisfying a aggregate production path, coincides with that of the competitive equilibrium.*

Proof. The proof is straightforward, and proceeds via contradiction. ■

Following proposition 1, the production path resulting from a competitive equilibrium,which generates known aggregate production (equivalently consumption) levels in each year (Q_t), can be computed using the following algorithm (which we refer to in later sections as the *Sorting Algorithm*): *1)* Start in year $t = 1$; *2)* set the field index i to order fields from lowest to highest marginal cost given costs, $c_f \mu_{ft}$, such that a lower i corresponds to a lower cost; *3)* start with $i = 1$; *4)* drain field i until remaining reserves equal zero ($R_{it} - q_{it} = 0$) or the aggregate production target is met ($\sum_{j=1}^{i} q_{jt} = Q_t$). Update remaining reserves for this field (set $R_{i,t+1} = R_{it} - q_{it}$); *5)* if $\sum_{j=1}^{i} q_{jt} < Q_t$, set $i = i + 1$, and go back to step *4*; *6* set $t = t + 1$ and *7)* if remaining reserves are positive for any field and $t < T$, go to step *2*, or else, stop.

This algorithm is used to generate the counterfactual production path, against which the observed production path is compared to measure the extent of production misallocation.

[22] See Sweeney (1992) for an extended discussion of the comparison of competitive equilibrium and equilibrium with market power in these settings.

4 Descriptive evidence of production inefficiency

Central to the existence of a productive inefficiency is the existence of cost dispersion between productive units, as well as the capacity of low-cost units to expand production to displace the production of high-cost units. This section documents these features in the data. It also provides reduced-form evidence consistent with the existence of market power by OPEC, and by Saudi Arabia in particular. We begin by introducing the dataset and providing summary statistics of the main variables used throughout the analysis.

4.1 Data

Table 2 presents summary statistics for the 13,248 active fields in the data across the entire sample. The average field produces 3.4 million barrels per year and has reserves of 99 million barrels (the medians, are 0.2 and 3.7, respectively). There is wide variation in field size, with the 5th percentile field producing fewer than 1,000 barrels, and the 95th percentile field producing 11 million barrels. The largest annual production for a field observed in the data was that of the Samotlor field in Siberia in 1980 with almost 1.2 billion barrels produced that year. Almost 19 percent of fields are offshore. The analysis presented in this paper is restricted to fields that were active at some point between 1970 and 2014.[23]

Given the Leontief production function, yielding the cost function given by equation (5), the average and marginal cost of oil production are the same. Hence, the marginal cost of production is recovered by dividing the total cost of production by the reported production, q_{ft}, (in million bbl/day), and the total cost of production is obtained by summing over the cost categories as listed in table A.1, in the Appendix. In particular, our baseline measure of marginal (and average) cost is computed as follows:

$$c_{ft} = \frac{\sum_h \text{Expenditure}_{hft}}{q_{ft}}, \tag{10}$$

where the various expenditure categories are h={ Well Capital, Facility Capital, Abandonment cost, Production Operating,Transportation Operating, and SGA}, and all expenditures are deflated by the US GDP deflator with 2009 as the base year. This specification rules out curvature in the cost schedule as an oil well gets depleted. Given this, careful consideration of the nature of the Leontief assumption is warranted. As in every production process, some fixed costs and scale effects undoubtedly exist in this industry. It is helpful to keep in mind the level of aggregation at which the analysis is being done. The analysis is industry-wide, aggregating the equivalent of an industry supply curve over all fields.[24] The Leontief technology assumption makes this supply curve a step function. Modeling each well and aggregating up would, at best, put a small amount of curvature in each step, which, given the level of aggregation,

[23] Of the 66,920 fields in the Rystad data, 45,687 of these did not produce between 1970 and 2014. The cost and reserves data on these fields are based solely based on engineering and geological modeling, and these fields are not used in the paper. More detail on the sample frame can be found in Table OA.1 in the Online Appendix.

[24] In Section 5 we use 11,455 fields, rather than the reported number of fields (13, 248) reported in the summary statistics, since we drop fields with no reported discovery year. This leaves us with 99.985 percent of global reserves.

would be difficult to notice for the typical field. When quantifying the production misallocation, in Section 5, we employ a measure of marginal cost admitting curvature in the cost of production coming from aggregate shocks in input markets and technology, and we verify the robustness of our results to the presence of within-field curvature (Section 6.3).

Central to much of the discussion in this paper is the notion of reserves. The reserve is the unextracted, but recoverable, quantity of oil remaining in the ground in a field. The most reliable way to measure the reserve at a point in time is to see the entire production life of a field. The total extracted oil is the maximal reserve. Most fields are not fully exploited in the data. Hence, industry reserve estimates need to be used. The oil industry reports reserves at different levels of extraction probability. There are three levels. P90 (or P1) is the quantity able to be recovered with a 90 percent probability, given current technical and economic conditions. The P90 reserve is the asset value that can be reported on company balance sheets under U.S. GAAP. Clearly, this definition means that reserves will fluctuate with the oil price. In the data used here, reserves are measured and reported assuming an oil price of $70 (in 2014 dollars), which is closest to the historical average price for oil. P50 (or P1 + P2) are the reserves recoverable with a 50 percent probability. Finally, P10 or (P1 + P2 + P3) are total reserves recoverable with a 10 percent chance. The level of P90, P50 and P10 can vary significantly within a field. For instance, in the North Ward Estes field discussed above, P90, P50 and P10 in 1975 were estimated at 26.6, 56.4 and 66.4 million barrels, respectively.

In this paper, in descriptive discussions (prior to Section 5,) P50 values at an oil price of $70 a barrel are used to report reserves. In section 5, a field's reserves in 1970 are computed as the sum of: i) the actual production history from 1970 to 2014; and ii) the P50 value at an oil price of $70 a barrel in 2014.

4.2 Preliminary Evidence

We begin by focusing on a small number of major oil-producing nations. By focusing on a small number of countries, we can illustrate more features of the underlying data. Attention is then shifted to the entire global market in which the aggregate data is shown to mirror the patterns observed in the more detailed country-level analysis.

Figures 3 and 4 show moments of the distribution of production costs for each year from 1970-2014, for each of Saudi Arabia, Kuwait, Venezuela, and Nigeria (OPEC Countries, Figure 3); and the United States, Russia, Canada, and Norway (non-OPEC Countries, Figure 4).

Panel (a) of Figure 3 examines Saudi Arabia. The solid black line is the oil price. Below that, for each year, is a black bar that shows the range of costs lying between the 5th and 95th percentiles, where the unit of observation is the barrel. That is, 90 percent of barrels produced by Saudi Arabia in a year have a unit cost lying in the range indicated by the black bar. The grey bar combined with the black bar indicates the range of costs between the 1st and 99th percentiles. Where circles are shown, these indicate the maximum unit cost for the country.

An examination of Figure 3 illustrates the tight range of costs for Saudi Arabian and Kuwaiti production. For both countries, costs per barrel rarely exceed $10. Further, costs are stable relative to the oil price. By contrast, costs in Venezuela and Nigeria are much higher and exhibit much greater dispersion. This is an important feature of the data, suggesting that, even within

OPEC, scope exists for efficiency gains due to reallocation of production. If OPEC were run as an efficient cartel, this feature would not exist, as allocations would be determined by a constrained social planner, with the production path having the same features as in Proposition 1. Given the many internal and external political challenges faced by OPEC– which mirror those faced by any real-world cartel – it is unsurprising that it fails to act as a theoretically efficient cartel might.[25]

Figure 4 allows us to compare the within-OPEC patterns in Figure 3 with those in non-OPEC countries. Panels (a) and (b) of Figure 4 show the US and Russia, the two biggest oil producers between 1970 and 2014. Both the US and Russia have more dispersion in costs than that observed in Saudi Arabia or Kuwait, although a significant proportion of production, particularly prior to 2000, has equivalent costs. Importantly, the more expensive production in both countries occurs at cost levels more than twice the levels than that characterize production in Saudi Arabia or Kuwait. This is particularly pronounced in the years following 2000, and particularly in the US, where the ramp up in high-cost production follows the rise in the oil price and is largely driven by unconventional onshore production (mostly shale). In 2014, 2039 million barrels were produced by shale (out of 4173 million barrels produced in the United States). These shale deposits had units cost of 32.6 dollars per barrel, while onshore fields had unit costs of 7.4 dollars per barrel (production weighted averages). By contrast, in 2005, shale accounted for only 24 million out of 2480 million barrels produced in the United States, and costs for onshore fields were 7.3 dollars per barrel. Thus, much of the large increase in costs in the United States is driven by the increased production from shale. Canada mirrors the US, with a similar ramp up in costs following 2000.

Norway is distinct from the three other countries in Figure 4, by virtue of having the vast majority of its production offshore. This accounts for the late start in production. Deepwater offshore drilling technology became commercially viable only in the late 1970s. The spikes in the ranges of unit costs reflect the starting years of oil rigs, the low production levels that the first year of production often brings, and the large scale of the infrastructure involved.[26] Interestingly, the rise in the oil price following 2000 brings an increase in the dispersion of costs, albeit in a much more muted way relative to the US and Canada.

The comparison between the dispersions in production costs in Saudi Arabia, Kuwait and the other six countries in Figures 3 and 4 illustrate the considerable scope for reallocation that exists. Dispersion in production costs is ubiquitous, and has been documented in a variety of settings ranging from manufacturing to services – see Syverson (2011) for an overview of the literature. Compared to the reported dispersion in productivity (measured by TFP) in the studies cited in Syverson, the dispersion in oil production is high; there is a 1:9 ratio between the $10th$ and $90th$ percentiles of cost. This is markedly higher than in most industries and is especially surprising since, for the oil industry, measurement is not contaminated by variation in output prices, which is a common issue in the literature (see De Loecker (2011)). Further, the

[25]See Asker (2010) for another example of an inefficient real-world cartel. Marshall and Marx (2012) provide an overview of a large number of cartels and the related theoretical and empirical work that organizes our understanding of them.

[26]The lumpiness observed here is inevitable. A large offshore project will involve many wells coming on line in the same year. If production starts late in the year, little production will be recorded, despite a large expenditure on infrastructure.

low costs that Saudi Arabia and Kuwait enjoy make it clear that, in a competitive equilibrium, these countries would be exhausting their deposits, subject to physical limits on extraction speeds, before the more speculative fields observed in the upper portions of other countries' costs distribution come online (see Proposition 1).

The extent to which the result in Proposition 1 is useful in interpreting the data rests on the plausibility of the following counterexample: if the low-cost fields in Kuwait are constrained by reserves, while those in Canada are not, then it is not surprising that there is no scope for low-cost countries to expand production. In Table 3, we show reserves in different regions of the world, as well as the ratio of reserves to production (that is, the number of years that a region could produce at the current rate) for 2014. Outside OPEC, the ratio of reserves to production is 10, while in OPEC countries, the ratio is 19. Hence, the data are consistent with the members of the cartel restricting production, relative to reserves, more than producers outside the cartel. As one might expect from the literature and historical commentary surveyed in Section 2, the data are consistent with OPEC, and Saudi Arabia in particular, having market power.[27]

The patterns observed in comparing the eight countries in Figures 3 and 4, are reflected in Table 4, which compares production, reserves and costs over time for Saudi Arabia, OPEC and all non-OPEC countries. Unit costs are reported using both the baseline specification, which omits taxes and royalties, and the alternative specification that includes taxes and royalty payments. Considerable scope for reallocation exists in each period, with the scope increasing as time goes on. This is not surprising, as distortions persist and low-cost OPEC deposits remain significant; in later periods, higher-cost deposits should come online as lower cost, non-OPEC, deposits get exhausted. Hence, the potential for gains from reallocation should get larger over time. This is the case regardless of whether the baseline or the alternative cost specification is used.

5 Quantifying the extent of misallocation

This section quantifies the extent to which market power can plausibly account for the misallocation observed in the production of crude oil. We do this by using the model described in Section 3 to compute a counterfactual production path, which we then compare with the actual production path to quantify the cost of misallocation. This requires the model to be parameterized. The details of this parameterization are found in the subsection below. Following that, we discuss the logic by which misallocation is attributable to market power. The results then follow, together with a series of robustness tests.

5.1 Model parameterization

The *Sorting Algorithm* described at the end of Subsection 3.2 is used to compute the competitive allocation (production path) in the counterfactual model described in Section 3. The inputs required are the aggregate production levels, Q_t, field-level total reserves, $R_{ft=1}$, and field-

[27] Other Middle East states, like Kuwait, also behave in ways consistent with market power. We focus on OPEC and Saudi Arabia since OPEC is the joint vehicle and Saudi Arabia has the largest reserves and production.

year costs, $c_f\mu_{ft}$. The remaining element required is a social annual discount factor, needed to compute a net present value of any accumulated distortions. This is set at 0.95.

Aggregate production is observed in each year from 1970-2014, and it is assumed that markets clear within the year, so that annual demand and production are equivalent. For years following 2014, global production (equivalently, demand) is assumed to grow at a rate of 1.3 percent per year, which is the (geometric) average growth rate observed for 1970-2014.

Reserves, as described above, are measured using P50 reserve figures, assessed at a price per barrel of $70 in 2014 dollars. Since reserves fluctuate somewhat over time for a given field, the actual production up to 2014 is added to the P50 reserve level in 2014 to give the reserve level for a given field available in 1970.[28]

Field-level costs are the central input required by our algorithm, and recovering $c_f\mu_{ft}$ from the cost data is the central aspect of generating this input. However, some auxiliary modeling elements, that bear on costs, are also relevant. The auxiliary elements are dealt with first. Then the recovery of $c_f\mu_{ft}$ is discussed.

The first auxiliary element is that the path of field discovery is assumed to be exogenous. Hence, for a field discovered in 1980, the cost of production is infinite prior to that date. Similarly, fields that are never observed to have produced between 1970 and 2014 are excluded. This is equivalent to assuming that the cost of these fields are infinite.

The second auxiliary element is the imposition of a limit on the proportion of $R_{i,t=1}$ that can be extracted in each year. The model in Section 3 assumes that any amount of oil may be extracted, up to the limit of available reserves, in any year. This is clearly a simplification. A range of engineering and geological factors can limit the proportion of reserves that can be extracted from a field in any given year, not least of which is the need to maintain a minimum level of pressure in the well so as to make extraction feasible – extraction that is too fast can lead to sharp drops in well pressure. The median producing field extracts 1.9 percent of its maximal reserves per year, and the 95th percentile field extracts 25.5 percent. The mean extraction rate in non-OPEC countries was 10 percent in 2014 (see Table 3). Given that in the main specification, the upper limit on the rate at which a field of can extract reserves is given by $\max\{x_f, 10\%\}$, where x_f is the maximal proportion of reserves extracted, in any year, for that field. The algorithm is easily adjusted to accommodate these auxiliary model elements, and we will present robustness checks where the extraction rate is alternatively chosen to be two percent, or unrestricted.[29]

We now turn to recovery of $c_f\mu_{ft}$ from the cost data. Unit costs for a field-year are measured as described in Section 4.1. These unit costs, denoted c_{ft}, need to be decomposed into three elements: 1) the time-invariant marginal cost, c_f; 2) a technology-year specific cost shifter, μ_{st}, where s indexes the technology (onshore and offshore); and 3) measurement error, $\exp(\epsilon_{ft})$. That is,

$$c_{ft} = c_f\mu_{ft} = c_f\mu_{st}\exp(\epsilon_{ft}). \tag{11}$$

[28] For some fields, we see reserves increasing over time, most likely because of new discoveries inside the field, and improvements in technology that make more oil recoverable. If we had used reserves reported in 1970, this would led us to the uncomfortable position of having more oil extracted in the period 1970-2015 than reported reserves in 1970, at least for certain regions of the world.

[29] All that is required is that the algorithm keep track of activity in a year and set prices to be infinite once the relevant field-level limits are reached. Proposition 1 and Corollary 1 are similarly unaffected.

In the counterfactual, production undertaken by field f in year t is taken to have occurred at cost $c_f \mu_{st}$ per barrel. The technology-year specific cost shifter, μ_{st}, is estimated as

$$\ln \hat{\mu}_{st} = \sum_{f \in s} \kappa_{ft} \ln c_{ft}, \tag{12}$$

where κ_{ft} is the quantity weight of a field in a given year's total output, $\kappa_{ft} = \frac{q_{ft}}{\sum_{f \in s} q_{ft}}$. Observations are weighted by production, as opposed to giving all fields equal weighting, since a field is an already aggregated unit of production, with the extent of aggregation varying across fields.

The time-invariant marginal cost, c_f, is then estimated, allowing for measurement error, using the following (within-field) regression:

$$(\ln c_{ft} - \hat{\mu}_{st}) = \ln \hat{c}_f + \epsilon_{ft}. \tag{13}$$

Estimation is conducted using weighted least squares, with the weights being the proportion of total field output done in that year.

Where confidence intervals are reported, they are computed via a bootstrap. Specifically, we employ a two-step bootstrap routine. In the first step, for each resample k, we take the true dataset and resample field-year observations ft and compute μ_{st}^k. In the second step, for each field in the true dataset, the field-years are resampled. This allows us to estimate c_f^k using the μ_{st}^k from the first step.[30] This, in turn, allows $c_f^k \mu_{st}^k$ to be computed using the algorithm to compute counterfactual predictions. The goal of this procedure is to capture the estimation error in not only the field-technology coefficient $\mu_{s,t}$, but also in the field-specific coefficient c_f. Fifty bootstrap iterations are used.

5.2 Identification of misallocation costs attributable to market power

To quantify the role of market power in distorting the efficient allocation of resources, the (counterfactual) path of extraction when firms are undistorted price takers needs to be computed. This is done with the sorting algorithm, using the cost measures described above. We then compute the total cost of production distortion by comparing the net present value of the costs of production from the observed cost of production to that from the counterfactual path.

There are two challenges to identifying the economic impact of misallocation plausibly attributable to market power in the oil market. First, it is unlikely that every instance of misallocation can be attributed to market power. Second, the data do not extend past 2014, which means that we do not see extraction paths in the data beyond this point.

In the absence of any other source of distortion, measuring distortions due to market power would be straightforward. The net present value, at 1970, of the cost of the observed production path would be compared to the net present value of the competitive equilibrium production path. The difference between the two would be the misallocative effect of market power

[30] The field-year observations used to compute μ_{st}^k are resampled independently from those used to compute each c_f^k. The practical reason to do this is that there a few large fields composed of tens of thousands of individual oil wells, such as Saudi Arabia's Ghawar fields, that have large, and central, effects in the counterfactual exercises.

measured as a stock in 1970 (we will present numbers deflated to 2014 dollars to make dollar numbers comparable across the paper). This total cost of production distortion must be a weak upper bound on the impact of market power, as it ignores any other source of distortion.

To focus the measurement on market power, it is necessary to articulate where market power is held. In the context of the global oil market, given the evidence presented in Sections 2 and 4, market power could be exercised by Saudi Arabia, by some intermediate subset of OPEC or by OPEC as a whole. When, for illustrative purposes, OPEC is considered the repository of market power, this still leaves distortions outside and within OPEC to consider. Given this, we proceed by solving a series of constrained social planner problems.

First, we solve for the competitive allocation, holding each country's production level in each year fixed. This removes internal distortions likely not attributable to market power. The second set of distortions to be removed are production distortions across both OPEC and non-OPEC countries. We remove these by computing the sorting algorithm, imposing the constraint that total non-OPEC production each year must be that observed in the data. The NPV of the cost of production from this path can then be compared to that from the unconstrained solution to the sorting algorithm. This gives the cost of two types of misallocation: the misallocation of production across OPEC and non-OPEC countries; and the misallocation of production within OPEC and non-OPEC countries. Third, holding OPEC production fixed, we solve for the competitive allocation again. This means that the undistorted market is free to reallocate production both within a country and across countries, subject to keeping OPEC production in each year the same as is observed in data. Lastly, the unconstrained competitive allocation is computed, which we call the (world) optimal solution. This allocation is required to deliver only the global production observed in each year in the data. Holding OPEC production constant, we take the difference between the competitive allocation and the optimal solution to be the distortion attributable to OPEC.

Almost surely, this measure of misallocation is conservative, and, thus, we call it a lower bound. In particular, it removes the distortions that emerge within OPEC itself that may be due to the political constraints that need to be met for OPEC to exercise any market power. That is, Saudi Arabia and Kuwait likely need to assign a positive quota to Venezuela in order to give them some rents from complying with the overall OPEC production plan. In most years, an efficient cartel would not have Venezuela producing. In the computation described above, distortions of this sort are not considered. Given that many real cartels are observed to use inefficient mechanisms to coordinate, at least some of the misallocation within OPEC should be attributable to the coordinated exercise of market power. See Marshall and Marx (2012) for an extensive overview, and Asker (2010) for a specific example. In addition, some of the within- or across-country distortions seen in countries outside OPEC may be due to strategic responses to OPEC production plans. To understand the extent to which this can further increase the misallocation attributable to market power, a competitive allocation in which only the country allocations within OPEC are held fixed is computed. This is then compared to the world optimal solution. The difference provides an upper bound for the measure of inefficiency due to market power, in which misallocation across OPEC countries is assumed to be entirely caused by the inefficient cartel mechanism.

Finally, we need to address the censoring in the data, such that production paths past 2014

are not observed. Given that oil is a finite resource, the central source of misallocative cost will be due to fields that are cheap to exploit being delayed, such that the resulting gains from trade occur in the future are discounted. This means that the future actual path of production matters for a measure of misallocation, as the more the exploitation of cheap resources are delayed, the greater the misallocation.[31] In the face of the inevitable censoring, we take a conservative approach. To project the path of "actual" production out past 2014, we compute the competitive solution, taking the stocks in each country at the end of 2014 as initial state variables. This means that there is no new distortion introduced to the path of actual production after the end of the data. As a result, the misallocation numbers we report are an underestimate of the true magnitude.[32]

The approach to isolating the impact of market power performed here measures the extent to which market power, on its own, moves the market away from prefect competition. In this sense, it measures the infra-marginal impact of market power. An alternative would be to attempt to model all the other sources of distortion in the market and then to estimate the marginal impact of market power on market outcomes, conditional on all other distortions. This would be a measurement exercise in the spirit of Lipsey and Lancaster (1956) and Buchanan (1969). Both the infra-marginal and marginal approaches are complementary in deriving an understanding of the force of market power in shaping the world oil market. The infra-marginal approach to measurement is the primary measure employed in this paper, as it keeps the analysis closer to the core data on costs. Measurement of the marginal impact of market power is explored in Section 6.1.

5.3 Results

We report results for paths from two different sample periods, 1970-2014 (the range of observed data) and 1970-2100 (when all fields active during the period covered by the data are exhausted). These dynamic measures collapse a lot of economic richness into a single NPV calculation. For this reason, we also discuss a static decomposition, as it gives some insight into the changing nature of distortions over time and the underlying mechanism of the model.

5.3.1 Dynamic productive inefficiency

We first compute productive inefficiency from the full dynamic model. We calculate the net present value of the cost of production of the entire observed quantity path in our sample, 1970-2014, starting in 1970. We also consider a longer time period, 1970-2100, for which we

[31] Our measure of distortions is also distorted even further because we consider only the contribution of fields that have produced in the data from 1970 to 2014. There are many fields that have yet to come online, and the costs of these fields, based on Rystad's estimates for unexploited oil reserves, are reported as being considerably cheaper in Saudi Arabia and Kuwait than in the rest of the world. However, incorporating these fields would require us to take a different approach to measurement, relying strongly on the accuracy of Rystad's cost forecasting model.

[32] Another option would be to use a fully specified structural model of OPEC and other producers to simulate this forward. Among other things, this would require capturing, in a parsimonious model, the geo-political aspects of OPEC, and world oil production generally, which seem beyond the scope of the current exercise. Instead, we opt for an approach that introduces a clear conservative bias.

forecast our demand for oil beyond 2014 using a 1.3 percent annual growth rate.[33]

We begin by examining the counterfactual path computed by the unconstrained sorting algorithm. This path is compared to the actual path in Figure 5, which plots the market share of OPEC in the actual and counterfactual paths over time. On the actual (observed) path, the production share of OPEC fluctuates by around 50 percent between 1970-1980. On the counterfactual path, this share jumps to to over 90 percent. This reflects the inter-temporal substitution of low-cost production (OPEC) for higher-cost production that is the source of production misallocation in this industry. Diving more deeply into which fields would produce in the competitive equilibrium in these early years, over 90 percent of world output in the 1970s would come from three fields: Ghawar Shedgum and Ghawar Uthmaniyah in Saudi Arabia, and Greater Burgan in Kuwait. Unsurprisingly, it is Saudi Arabia and Kuwait that are often pointed to by industry commentators as leaders in the OPEC cartel. It takes until the mid-1990s for the production share of OPEC in the counterfactual and the actual path to converge, suggesting a substantial amount of misallocation.

The extent of this misallocation is reported in Table 5. The left column reports results for the years 1970-2014, the extent of our data, while the right column reports results for 1970-2100, which corresponds to the exhaustion of all resources in our sample. The 1970-2100 results allow for the inter-temporal substitution of production to be fully incorporated into the calculation, but in doing so, we make the conservative assumption that after 2014, the actual path of extraction is determined by the solution to the social planner's problem, taking conditions at the end of 2014 as initial conditions. The 1970-2014 results are provided to give a sense of the influence of this assumption.

Focusing on the 1970-2100 results, the cost of the actual path of extraction (actual up to 2014 and the social planner's thereafter) is 2.499 trillion in 2014 dollars. The cost of the counterfactual path in which all market actors are undistorted price takers is 1.757 trillion (2014) dollars. That is, the counterfactual costs are measured to be 70.3 percent of the actual costs. The difference between the two, 742 billion (2014) dollars, is the extent of the total distortion in the market. This is decomposed into within-country distortions for non-OPEC and OPEC countries (38% and 21% of the total distortion, respectively); across-country distortions between non-OPEC countries (19% of the total distortion); across-country distortions between OPEC countries (7.8% of the total distortion); and the distortion between OPEC and non-OPEC countries (13.9% of the total distortion).

Within-country distortions can be attributed to wedges that move national production away from cost minimization. Examples of the sources of such distortions might include: political economy forces directing production to specific regions to, for instance, promote employment; region taxation (e.g., different payroll tax rates in different U.S. states); risk factors not fully captured by input cost measures (e.g., armed conflict in specific regions of a country); natural events (e.g., the impact of hurricane Katrina would be counted as a distortion in this framework); or environmental restrictions, or other regulatory frictions, that are location-specific.

Across-country distortions can come from similar sources, albeit acting at the cross-country level. For instance, a national oil production tax could impact all national production equally

[33] In practice, given that the sample does not include untapped fields as of 2014, oil production ceases in around 2035, depending on the exact model specification.

but could drive a wedge between national production and competing international production. Across-country distortions for OPEC countries are particularly interesting as an additional source of distortion, such as the quota-like agreements that OPEC has periodically used to coordinate production cuts across its members. To the extent that these arrangements restrict the low-cost producers while giving freedom to the high-cost producers, they contribute to distortions in the rest of the market. Thus, at least some of the across-country distortion in OPEC countries is likely attributable to OPEC's coordinated exercise of market power. Finally, OPEC's self-imposed production restrictions distort production, leading to a higher proportion of production coming from non-OPEC countries. This gives rise to the final source of distortion.

Of these sources of distortion, a lower bound is derived by focusing on the distortion between OPEC and non-OPEC countries. This is a lower bound because it ignores the inefficiency of the OPEC mechanism itself. An upper bound is derived by adding the across-country distortions between OPEC countries to this lower bound. This leads to a lower and upper bound of 103 and 161 billion dollars, respectively, or 13.9 percent and 21.6 percent of the total distortion. Of note is that misallocation resulting from the internal structure of the cartel is estimated to account for up to 36 percent of the overall production distortion generated by OPEC's exercise of market power.[34]

To give a sense of scale, recall that the NPV of actual costs is a (conservative) estimate of the full realized resource cost of production. The estimate of the distortionary impact of market power represents 6.4 percent of the full resource cost of production. By comparison, 29.7 percent of the total resource cost is attributable to some form of distortion. Given that the total distortion measure includes acts of nature and other acts (such as wars) that lie beyond the reach of mainstream economic policy, the fact that market power can plausibly account for 21.6 percent of the total distortion in the market seems to suggest that market power is a significant policy-relevant source of distortion. This is further emphasized by noting that, at times, countries outside OPEC have coordinated with OPEC to guide world price (notably Russia and Norway in the late 1990s), suggesting that market power distortions may also be found in places other than OPEC in this market.

5.3.2 Static productive inefficiency

The measure from the full dynamic model, in able 5 above, provides a quantification of misallocation taking into account the full extent of the inter-temporal substitution of production. In doing so, it collapses everything into a single NPV computation. While this is model-consistent, it hides aspects of the model mechanics and also anchors everything to the starting date of our setting, 1970. In this section, "static distortions" are reported to complement the results from the full dynamic model. These static distortions are computed by taking the observed initial conditions at the start of each year as given, and computing a counterfactual path from those initial conditions. The distortion for only that year is reported. That is, for 2014, we take as initial conditions the state of the global market at the end of 2013 and run the sorting algorithm from that starting point. This gives us the counterfactual production for 2014, which we compare to actual production for 2014. This computation gives a sense of the extent to

[34] Asker (2010) finds a similar magnitude of cartel inefficiency.

which misallocation varies by year, taking as initial conditions the actual market conduct in all previous years.

We begin by examining distortions in quantity space. Table 6 presents the market share of the 20 largest oil-producing countries in 2014, as well as those that our competitive model would predict with price taking to start on Jan 1 2014 (hence the term static). The measures presented here incorporate all distortions, rather than focusing specifically on those that can be attributed to the exercise of market power by OPEC.

As might be expected, the market share of the Gulf countries increases significantly, from 25.8 percent to 74.4 percent. Saudi Arabia increases its share by 28.1 percentage points, and Kuwait increases by 12.5 percentage points. This mirrors contemporary commentary casting these two counties as key players in the OPEC cartel.

All other Gulf countries would increase production, but not to the same extent. Other, non-Gulf, OPEC members would cut back on production, reducing market share by a cumulative 9.1 percentage points. This is consistent with OPEC not being an efficiently run cartel internally, and allocating more share to these countries than would be consistent with joint surplus maximization.

Production by non-OPEC countries would decrease substantially. The large non-OPEC producers would decrease their share from 60.7 percent to 21.2 percent, while the rest of the world would decrease share from 13.6% to 4.4%. In particular, Russia and the U.S. would both see large share reductions of 9.7 and 11.9 percentage points, respectively.

These quantity changes are significant, physically, economically and, geo-politically. This underscores the significant extent to which distortions shape world production. It should be recognized that shifting shares to this extent is unlikely to be technically feasible in one year, but then neither is it likely that all distortions present in the global oil market will be removed in one year. Rather, the results in Table 6 give a clear picture of the significance of the cumulative distortions and the influence of these distortions in shaping the world as we find it.

In Figure 6, static distortions, reported as costs, are shown for each year from 1970 to 2014, together with decompositions into components that mirror those discussed in Table 5. The time-series of the size of the overall distortion follows the rise, fall and rise of the oil price. This is to be expected, as a higher oil price will attract entry by marginal producers, with these marginal producers having higher costs as the oil price rises. Hence, as the oil price rises, the marginal unit of withheld production attracts a higher-cost substitute. Given this, it is unsurprising that recent higher oil prices coincide with higher distortions. This effect is compounded, at least to some extent given other distortions, by the mechanical process by which lower-cost non-OPEC reserves are depleted earlier, resulting in the marginal producer in later years having a higher resource cost. Nonetheless, in years post-2008, over 25 percent of the total static distortions shown here are attributable to OPEC's exercise of market power (the combined black and grey components of each bar).

5.3.3 Alternative Specifications

The baseline model underlying the results reported above incorporates a set of model and parameter assumptions. These include the following: field extraction in a given year is capped

at 10 percent of the maximal reserve level; reserves are measured using the P50 metric assessed at a price per barrel of $70; the resource cost of extraction does not include payments made in the form of taxes or royalties; and fields are available for exploitation after the date of discovery. In this section, the sensitivity of the baseline results to these assumptions is explored.

Table 7 shows the dynamic counterfactuals for various other alternative specifications for the timespan 1970-2100 (exhaustion). Column (1) shows results for the baseline specification and merely reproduces the results reported in Table 5. Columns (2) and (3) show results when field extraction in a given year is capped at the maximum of 2 or 100 percent of the field's maximal reserve level (respectively) and the maximum extracted proportion of maximal reserves observed in data for the field in any year.[35] All measures of distortions and costs are similar across specifications (1) through (3).

Column (4) switches attention to the measure of reserves, substituting a P90 reserve measure for the P50 measure used in the baseline. Recall that P90 is a more restrictive definition of reserves. The use of this alternative measure makes almost no difference to the results. This is because, for many fields and almost all the larger ones, we see a long history of production, which means that reserve numbers comprise only a component of our measure of recoverable reserves, and these reserves are reported on fields for which the underlying geology is well understood. Any remaining differences are further absorbed through the exercise of taking an NPV. Hence, any differences that occur toward the end of the time period have little impact on the discounted sums.

Column (5) adds the tax items in Table A.1 to costs. These costs are then interpreted as the resource costs that are relevant for welfare measures, and the cost measures that determine the path of extraction in the undistorted, price-taking, counterfactual. This measure sees a significant jump in the extent of within-OPEC distortion. This occurs because distortionary taxes within OPEC are higher in the high-cost countries (Saudi Arabia, for instance, extracts oil payments from Saudi Aramco profits rather than from revenue, and so incurs no distortionary taxation). Hence, when including tax measures in the measure of resource costs, this exacerbates existing distortions within OPEC. In the rest of the world, there is a slight negative correlation between tax levels and costs, and so this effect is not present there. Column (6) lets observed taxes influence the behavior determining the paths of production, but evaluates the resource cost without including taxes. The impact of taxes on the estimates are explored further in Section 6.1.1.

Column (7) restricts the sample to include only those fields active in 1970. Column (8) restricts fields to be available in the counterfactual after the first date of observed production, rather than from the date of discovery. As can be seen, this makes little difference to the upper and lower bounds on the impact of market power. Together, these simulations provide preliminary evidence that the results are not being driven by the treatment of any start-up costs that may be present. These results are discussed more thoroughly in Section 6.2, together with a more detailed discussion of the impact of start-up costs.

[35]The actual and counterfactual costs of extraction vary slightly across specification, depending on when final extraction occurs. This is true for the actual path, in addition to the counterfactual path, since it needs to be projected out past 2014.

6 Modeling alternatives

In this section, we discuss additional factors that may impact the results reported above. In Section 6.1 we drop the infra-marginal approach to measuring welfare impact adopted in the rest of the paper (in which other distortions are removed before the impact of market power is calculated). Instead, we infer the size of wedges resulting from distortions other than market power. These wedges are then used to infer the marginal impact of market power, in the spirit of Lipsey and Lancaster (1956) and Buchanan (1969). Following that, we return to the the infra-marginal measurement model and consider the impact of start-up costs (section 6.2), curvature in fields' marginal cost curves (Section 6.3) and heterogeneity in the discount factors of market actors (Section 6.4).

6.1 Taxes and Wedges

The analysis proceeds in two steps. The first, in Section 6.1.1, examines the impact of market power conditional on observable taxes. Aside from accounting for the impact of observable taxes on behavior, it stays close to the marginal approach to measuring welfare impact. The second step, in Section 6.1.2, estimates the marginal impact by first inferring the distribution of wedges in non-OPEC countries and then exploiting the assumption that, in the absence of OPEC market power, these wedges reflect the wedge distribution that would arise within OPEC. This lets world outcomes be compared to a counterfactual that models non-market-power distortions, allowing us to estimate the marginal impact of market power.

6.1.1 Taxes and Royalties

As documented in Table 2, there are many different taxes that governments levy on the oil industry. These taxes include royalties, taxes on revenue, income taxes, forms of production-sharing agreement that act like royalties, and operating expenditure taxes. Depending on the country, these taxes can generate revenues that are up to double the resource cost of extracting oil. Some taxes, such as income taxes, are, in principle, non-distortionary — they should not affect production choices. Other taxes, including royalties, taxes on revenue, taxes on operating expenditures, and production-sharing agreements, will alter production decisions. The data contain records of many of these forms of taxation. Given this, we incorporate those tax elements that are clearly distortionary (listed in Table 4) into costs. These costs are then used to compute counterfactual paths. However, in evaluating the economic (resource) costs of different allocations, we use only economic costs and not costs inclusive of taxes. That is, we compute competitive allocations, under different constraints, under the assumption that the sorting algorithm operates on costs inclusive of observed distortionary taxes.

These results are included in column (6) of Table 7. The results suggest that the presence of these observed distortionary taxes, if anything, slightly increase the impact of market power in this market. Since OPEC countries typically have nationalized oil production, they tend not to raise government revenues through taxes on oil producers.

6.1.2 Wedges

While the previous section discussed the effect of observed distortionary taxes on our measurement of the effect of market power, there are still likely other deviations present in the data that drive a wedge between price and (social) marginal costs. For instance, in the United States, we frequently observe cheaper fields producing after more expensive fields have been used, which violates the logic of the sorting algorithm. As in the previous discussion of taxes, we wish to evaluate the effect of market power in the presence of these unobserved "wedges" on production choices. That is, we want to derive a measure of the marginal impact of market power conditional on these distortions.

The first step in doing so is to infer the size of these wedges for the oil reserves in our data. This is done by computing the wedges required to transform the marginal costs observed in our data into marginal costs (inclusive of wedges) that are consistent with the order of extraction actually observed. The idea is that if two fields have marginal costs of \$6 and \$10 per barrel, but the \$10 field is extracted first, then there must be a wedge of at least \$4 imposed on the cheaper field. Subject to some details in implementation, this observation lets the wedges that apply to non-OPEC fields be recovered, given that we observe actual marginal costs and the order of extraction.[36] Wedges are inferred such that the marginal costs inclusive of wedges generate the observed production path of non-OPEC fields when applied to the sorting algorithm. Given that production paths can be rationalized by a range of wedges, we select the vector of taxes that minimizes the sum of the absolute size of wedges. Where the solution is not unique over this set of vectors, we choose the solution that sets the tax on the median barrel in that interval equal to zero.

The above approach results in a vector of implicit taxes that apply to barrels that are not subject to the market power distortions resulting from OPEC. This being said, the timing of extraction in non-OPEC countries, is clearly dependent on the market power exerted by OPEC, so the estimation of these wedges cannot be completely divorced from market power.

In implementing this approach on the full dataset some model choices are required to ease computation and enhance transparency. First, no annual extraction limit is imposed on a field ex ante. The implicit taxes that are derived absorb this feature at a field-year level. Second, μ is common to all fields and is not technology-specific. This allows the taxes to be derived deterministically since, with this adjustment, the ordering of the production sequence in perfect competition is independent of μ realizations.

The above process gives a set of wedges for non-OPEC fields (equivalently, reserves or barrels). Recovery of non-market-power related wedges for OPEC fields is not possible in our setting since the path of OPEC extraction is a function of the exercise of market power. This means that OPEC wedges cannot be separately identified.

Hence, the second step in the process is to use the set of wedges recovered from non-OPEC fields to inform an understanding of the wedges that would be present in OPEC fields in the absence of OPEC exercising market power. To do this, we sample from the distribution of wedges inferred on non-OPEC barrels and apply these to OPEC fields to form the marginal costs that determine the production paths of OPEC fields in the absence of market power.

[36] A detailed discussion of the algorithm used to recover wedges is contained in Appendix B.

This sampling process is conducted as follows. The first step, on non-OPEC fields, gives a distribution of wedges. Each wedge τ_k in this distribution is mapped to a subset of the barrels of size q_k in a given field (if a field produces across multiple years, the barrels in each year will have different wedges since the c_f's are common to all barrels in a field). We sample i.i.d. with replacement from the distribution of wedges, and apply each wedge τ_k that is sampled to a quantity q_k of barrels in an OPEC field. This process is continued until all OPEC reserves have a wedge attached to them. This results in a sampling procedure that applies wedges weighted by the quantity of oil to which the wedge applies in the non-OPEC wedge distribution. We will refer to this as the unconditional sampling procedure.

A notable feature of the data is that wedges are strongly negatively correlated with actual marginal costs. That is, the correlation coefficient between c_f and the inferred wedges in the non-OPEC fields is -0.99.[37] To accommodate this, we employ an alternative sample procedure, in which we require that wedges applied to OPEC reserves be drawn from non-OPEC fields that have a c_f within \$5 of the c_f of the OPEC field. We will refer to this as sampling conditional on marginal costs.

Table 8 shows the results, and columns (1) and (2) reproduce the baseline and 100 percent extraction limit cases in Table 7 (columns (1) and (3)). These two columns are for comparison with the results conditional on wedges, and the reported measurements have the same definition as those reported in Tables 7 and 5.

Columns (3) and (4) of Table 8 contain the results derived using the procedure described above. Column (3) uses the unconditional sampling procedure. The reported actual cost is the NPV of the resource cost of the actual path of extraction, evaluated using actual resource costs.[38] The counterfactual reported for column (3), denoted C_2, takes marginal costs including wedges, and computes the extraction path of global oil reserves assuming that all actors are price takers. That is, the sorting algorithm is applied with marginal costs including wedges as an input, and not further constraints. Note that this is different from that reported for columns (1) and (2). As in earlier tables, the reported counterfactual there, denoted C_1, removes all sources of distortion and corresponds to a perfectly competitive equilibrium path.

These differences in measurement of the counterfactual results in column (3)'s counterfactual (C_2) giving similar total distortions ($A - C_1$ or $A - C_2$) to those in columns (1) and (2). That said, the difference is not as large as one might expect. This is due, in part, to the unconditional sampling procedure. This procedure means that OPEC fields do not have the same negative correlation between c_f and the wedge size observed outside OPEC. This results in OPEC have a much more efficient extraction path, absent market power, than the rest of the world, allowing large welfare gains to be realized from moving to price-taking behavior.

Given this, column (4) reports results using sampling conditional on marginal costs. As can be seen, this reduces the impact of removing market power distortion and allows other

[37] Since wedges are constructed to explain why lower-cost fields are not extracted first, it is not surprising that wedges move precisely in the opposite direction from costs.

[38] These differ across specification due to different extraction rate assumptions in columns (1) and (2) and differences in wedges sampling between (3) and (4). This means that each specification has slightly different extraction paths post-2014. In keeping with the necessary structure of inferring wedges, paths of extraction are computed taking costs as $c_f \mu_t$, but to minimize departures from the results in the rest of the paper, resource costs are evaluated using cost as measured by $c_f \mu_{st}$

distortions to have a greater influence on the counterfactual paths.

For columns (3) and (4), the OPEC distortion is simply the difference between the actual cost and the counterfactual costs. This is because all distortions unrelated to OPEC are captured by the imputed wedges imposed on the marginal costs used by firms to compute the extraction paths. For columns (1) and (2), the definitions of upper and lower bound correspond to those used in Tables 7 and 5.

Comparing the distortion due to OPEC across columns indicates that the infra-marginal approach to measurement (columns (1) and (2)), if anything, understates the impact of market power in this setting. The impact of market power as reported in columns (3) and (4) tends to be larger, suggesting that, at the margin, market power combines with other distortions to further amplify the extent of production misallocation, and the associated welfare loss therein. Thus, a marginal approach to measurement, accounting for the impact of the theory of the second best, in the spirit of Lipsey and Lancaster (1956), suggests that our baseline estimate of the impact of market power in the oil market is a conservative estimate of the impact of OPEC countries shifting to price-taking behavior.

6.2 Start-up Costs

Start-up costs, expenditures linked to "switching the field on" and, therefore, sunk in or before the first year of operation, can be an issue in two parts of the analysis. First, they may alter the counterfactual production path. That is, a high start-up cost, low marginal cost, field may delay its initial production date in a competitive equilibrium relative to that predicted in the baseline model. Second, these fixed costs are welfare-relevant and, to the extent to which they are not being counted in the measurement, they may offset the otherwise conservative nature of the calculations executed using the baseline model.

As a preliminary measure, we examine the proportion of production between 1970 and 2014 that is provided by fields that were producing in 1970. For these fields, all start-up costs will already be sunk. These fields collectively are responsible for 58 percent of total production between 1970 and 2014, over 45 percent of total costs. Hence, for a large proportion of the fields and costs, the presence and magnitude of any start-up cost is irrelevant.

Figure 7 investigates the size of these fixed costs for fields that started production after 1970, by computing, within a field, cumulative costs and cumulative production. The magnitude of these start-ups costs can be identified from the proportion of expenditures incurred prior to the start of production. The figure then aggregates this relationship, weighting by field-level production, and breaking out fields by onshore, offshore shelf, offshore deepwater, and shale/oil sands. For a conventional onshore field, just over 20 percent of costs are incurred before the first barrel is produced, while for an offshore deepwater field, this number is closer to 30 percent. By contrast, shale has much smaller start-up costs (even if this relationship is far noisier given the more limited number of shale fields).[39]

[39] In Figure 7, constant marginal costs and no start-up costs would imply that cumulative costs lie on the 45-degree line. In the presence of start-up costs, the line is rotated upwards, but should remain linear thereafter. In figure 7 the right tail of the cost distribution is slightly convex, suggesting that there are some increasing marginal costs at the end of a field's life. As is discussed in Section 6.3, this is due to rising industry-wide input costs at the end of our sample period.

Given the presence of start-up costs, we evaluate their impact on our results by running two alternate simulations. In the first, we use only fields active in 1970, which have already sunk any start-up costs. In the second, we restrict the first year of production of fields to occur in or after the first year that we observe production in the data (as opposed to discovery, as in the baseline specification). This allows the process governing the imposition of start-up costs to be held constant between the actual path and any counterfactual. Hence, start-up costs can be canceled. Results from these simulations are found in columns (7) (only fields active in 1970) and (8) (restricting start year) of Table 7. In Column 7, actual costs are only 1465, rather than 2499 in Column 1. This is because we also restrict demand to account only for consumption from pre-1970. We find a 30 percent difference between the actual and competitive counterfactual costs, which is identical to the baseline results in Column 1. Moreover, the upper and lower bounds on the contribution of OPEC to this gap are 40 and 27 percent, respectively, larger than the 22 and 14 percent in the baseline of column 1. Likewise, in Column 8, the actual costs are similar to the baseline, but the counterfactual costs are higher at 1,797 billion, rather than 1,756 billion, since we have prohibited fields from starting production before the first year we see them produce in the data. Again, the productive inefficiency due to OPEC is also larger, with upper and lower bounds of 188 and 125 billion, respectively, versus 163 and 105 billion in column 1 (baseline).

There are larger effects when we focus attention on fields producing in 1970 or restrict counterfactual production paths to start off new fields no sooner than their first year of production in the data, since most of the low-cost fields in OPEC are the super-giant fields in the Persian Gulf, such as Ghawar or Burgan, and these fields have been in operation since the 1950s and 1960s. Therefore, the reallocation of production from non-OPEC fields to super-giant fields in the Persian Gulf is unhindered by start-up costs since these fields were already producing in 1970.

To explain the similarity between these alternative simulations and the baseline results, we examine the correlation between start-up costs and our estimates of marginal costs. Fields with high start-up costs and low marginal costs are problematic for the way competitive equilibrium is modeled. These fields may delay activation relative to what is predicted by the sorting algorithm we employ. To investigate the correlation between start-up and marginal costs, we measure start-up costs as the sum of expenditures in years prior to and including the first year of production, and compute the correlation between these start-up costs and unit costs c_{ft}.

The resulting correlation coefficient is 0.47, while the Spearman rank correlation, on the same sample, is 0.86. This suggests a strong positive relationship between a field's initial start-up cost and subsequent total cost of production, so start-up costs would not, on average, reverse the order of extraction from the sorting algorithm on marginal costs.

6.3 Marginal cost curvature

Recall that we use the following specification for costs (in equation 11):

$$c_{ft} = c_f \mu_{st} \exp(\epsilon_{ft}).$$

This specification assumes constant marginal costs, conditional on the realization of μ_{st}. If costs have curvature, this will be absorbed into the ϵ_{ft}, and this ϵ_{ft} will be correlated with the stage

of the field in its life cycle (the proportion of recoverable reserves that have been extracted).

To assess the impact of any field-level curvature that may be present, it is useful to keep in mind two features of our approach. First, we consider the entire global market for crude oil, and aggregate production across many thousands of oil fields and hundreds of thousands of individual wells, which implies that within-field curvature is likely to be less important to the extent that it merely smooths the transitions in an aggregate supply curve that resembles a step function. Second, we already impose a form of curvature by limiting the speed of extraction in a given year.

Nonetheless, we present two pieces of evidence to inform an evaluation of the likely impact of curvature on the results. First, in Figure 8, we present the observed marginal cost schedule for the largest oil field in our data, Ghawar Uthmaniyah in Saudi Arabia, with cumulative production on the horizontal axis and costs on the vertical axis. We also plot the predicted marginal cost derived from estimating the cost specification in equation 11. The wedge between the two curves indicates the extent to which the cost specification, which combines constant marginal cost with technology-year specific cost shocks, is violated. Reserves are shown with a vertical line. This figure indicates that predicted and actual marginal costs are very close to each other and that the main source of variation in the observed costs is due to the cyclicality in input prices that is correlated with oil prices, most likely reflecting input market tightness during periods of high prices. This source of cost variation is picked up by the μ_{st} in the cost specification, and the technology s subscript allows the share of energy used in production (e.g., fuel costs) to vary across technology types.

Second, in Figure 9 presents the relationship between the error term ϵ_{ft} and cumulative output over reserves, for all onshore fields in the Gulf states, the U.S. and Russia, as well as for Norway's offshore fields. Figure 9 shows that ϵ_{ft} is unrelated to the the proportion of recoverable reserves that have been extracted. In particular, there is no obvious pattern as reserves near depletion. For Norwegian offshore production, an initial high marginal cost of production is observed, reflecting the start-up costs discussed in the previous section, but following these early periods, ϵ_{ft} is centered around zero.

Together, Figures 8 and 9 indicate that any violations away from constant marginal costs are not substantial, and this visual intuition is confirmed by regression analysis on the entire sample of fields. Moreover, the presence of the technology-year fixed effects absorbs most of the observed variation in the levels of marginal cost curves.

6.4 Discount factors

A maintained assumption in the analysis presented in this paper is that all actors in the market have the same discount factor (δ), and, indeed, any change in a common discount factor does not alter the ordering of production in a price-taking equilibrium. Many features of the framework explored here would be infeasible were this feature to be arbitrarily relaxed. For instance, the wedges explored in Section 6.1 could be substantially rationalized by different discount factors existing across different firms and regimes.

That said, the model accommodates some flexibility in discount rates with no change to the results. In particular, Proposition 1 and Corollary 1 remain valid if fields with lower costs

have a lower discount factor (value the future less) – that is if $\frac{\partial \delta}{\partial c} > 0$. If this is true, then the sorting algorithm is preserved without alteration. Given this, and that many of the lower-cost reserves in the world are located in arguably less stable geographic regions, it may be that, even if the common discount factor assumption is unreasonable, the sorting algorithm still provides a useful framework through which to model the price-taking counterfactual. In such a setting, all that would be required is that the discount factor that is used, 0.95, be an appropriate social discount factor for making global welfare calculations.

7 Conclusion

This paper provides estimates of the extent to which market power is a source of production misallocation. This production misallocation, and the resulting welfare loss occur through less production being allocated to low-cost producers, and more to high-cost producers. In quantifying the potential magnitude of this productive inefficiency, we examine the global market for crude oil, in which OPEC, a notable international cartel, is alleged to exert considerable market power at times.

We find evidence for substantial productive inefficiency due to market power, of which we attribute up to 163 billion 2014 USD, in NPV terms, to the activities of the cartel, depending on model specification. The results from this study indicate that market power can affect aggregate outcomes – here, the total cost of production in the world oil market (and, hence, the price of oil) – which, in turn, affect a host of economic decisions and macroeconomic aggregates. Harberger (1954) estimated the welfare losses due to monopoly for the entire US economy in 1921 at two billion dollars in today's currency. From the study of a single industry, oil, we provide evidence suggesting that monopoly may have welfare effects that are several orders of magnitude larger than previously thought.

The point of the departure of this paper is to focus on the so-called *rectangle* – i.e., the welfare due to misallocation of production, given the quantity produced. This is in contrast to the traditional focus on the so-called *triangle*– i.e., the welfare loss due to the quantity reduction implied by market power. The distinction between both welfare losses is, of course, well known, but this paper distinguishes and quantifies the former source of welfare loss. It does so by combining high-quality micro-level data on cost of production with a novel empirical framework, which takes this cost data as an input to simulate the counterfactual allocation in the absence of market power.

The framework introduced in this paper is quite general and can be applied studying misallocation in other contexts. The sorting algorithm in this paper addresses the problem of inter-temporal misallocation, and the approach of providing lower bounds to misallocation by looking at competitive solutions constrained to limit production within and between countries could be used to provide more conservative, and perhaps also more plausible, estimates of economic distortions. In industries in which production dynamics are not present, but for which similar, detailed production and cost data exist, the static analog of this approach is similarly feasible.

However, many other environments will imply the cost of production to be linked over time; these includes but are not limited to learning by doing, adjustment cost of factors of

production, time to build, technology adoption, and research and development. While in this paper, the approach is tailored to the specifics of the oil market, it is more general and applies whenever producer-level costs of production is observed and an indication of which market participants are, potentially, believed to execute market power.

Finally, thinking more carefully about distortions in the presence of other economic frictions opens the door to incorporating the theory of the second-best into work on misallocation.

References

Ackerberg, Daniel, C Lanier Benkard, Steven Berry, and Ariel Pakes, "Econometric tools for analyzing market outcomes," in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 6B, Elsevier, 2007, pp. 4171–4276.

Anderson, Soren T, Ryan Kellogg, and Stephen W Salant, "Hotelling under pressure," *Journal of Political Economy*, 2014, *forthcoming*.

Asker, John, "A study of the internal organization of a bidding cartel," *The American Economic Review*, 2010, pp. 724–762.

_ , Allan Collard-Wexler, and Jan De Loecker, "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 2014, *122* (5), 1013–1063.

Atkeson, Andrew and Ariel Tomás Burstein, "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, 2010, *118* (3), 433–484.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta, "Cross-country differences in productivity: The role of allocation and selection," *The American Economic Review*, 2013, *103* (1), 305–334.

Bartik, Alexander Wickman, Janet Currie, Michael Greenstone, and Christopher R Knittel, "The local economic and welfare consequences of hydraulic fracturing," Technical Report, National Bureau of Economic Research 2016.

Borenstein, Severin, James B Bushnell, and Frank A Wolak, "Measuring market inefficiencies in California's restructured wholesale electricity market," *The American Economic Review*, 2002, *92* (5), 1376–1405.

Bornstein, Gideon, Per Krusell, and Sergio Rebelo, "Lags, Costs, and Shocks: An Equilibrium Model of the Oil Industry," Technical Report, National Bureau of Economic Research 2017.

Bresnahan, Timothy F, "Competition and collusion in the American automobile industry: The 1955 price war," *The Journal of Industrial Economics*, 1987, pp. 457–482.

Bridgman, Benjamin, Shi Qi, and James A. Schmitz, "Cartels Destroy Productivity: Evidence from the New Deal Sugar Manufacturing Cartel, 1934-74," Staff Report 519, Federal Reserve Bank of Minneapolis October 2015.

Buchanan, James M, "External diseconomies, corrective taxes, and market structure," *The American Economic Review*, 1969, *59* (1), 174–177.

Byrne, David P and Nicolas de Roos, "Learning to coordinate: A study in retail gasoline," Technical Report, University of Melbourne 2016.

Collard-Wexler, Allan, "Productivity dispersion and plant selection in the ready-mix concrete industry," *Working paper, Duke University*, 2011.

_ and Jan De Loecker, "Reallocation and Technology: Evidence from the US Steel Industry," *American Economic Review*, 2015, *105* (1), 131–171.

Covert, Thomas R, "Experiential and Social Learning in Firms: The Case of Hydraulic Fracturing in the Bakken Shale," Technical Report, University of Chicago 2015.

Crémer, Jacques and Djavad Salehi-Isfahani, *Models of the oil market*, Vol. 2, Taylor & Francis, 1991.

De Loecker, Jan, "Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity," *Econometrica*, 2011, *79* (5), 1407–1451.

__ **and Jan Eeckhout**, "The Rise of Market Power and the Macroeconomic Implications," Technical Report 23687, National Bureau of Economic Research 2017.

__ **, Pinelopi Koujianou Goldberg, Amit Kumar Khandelwal, and Nina Pavcnik**, "Prices, Markups and Trade Reform," *Econometrica*, March 2016, *84* (2), 445–510.

Downey, Morgan Patrick, *Oil 101*, La Vergne, Tenesse: Wooden Table Press, 2009.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "Competition, Markups, and the Gains from International Trade," *American Economic Review*, 2015, *105* (10), 3183–3221.

Fattouh, Bassam, "The drivers of oil prices: the usefulness and limitations of non-structural model, the demand-supply framework and informal approaches," *Oxford Institute for Energy Studies Working Paper*, 2007.

Fink, Nikolaus, Philipp Schmidt-Dengler, Konrad Stahl, and Christine Zulehner, "Registered cartels in Austria: An overview," Technical Report, ZEW Discussion Papers 2015.

Foster, Lucia, John Haltiwanger, and Chad Syverson, "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *The American Economic Review*, 2008, pp. 394–425.

Genesove, David and Wallace P Mullin, "Testing static oligopoly models: conduct and cost in the sugar industry, 1890-1914," *The RAND Journal of Economics*, 1998, pp. 355–377.

__ **and** __ , "Rules, Communication, and Collusion: Narrative Evidence from the Sugar Institute Case," *American Economic Review*, 2001, pp. 379–398.

Goldberg, Pinelopi Koujianou, Amit Kumar Khandelwal, Nina Pavcnik, and Petia Topalova, "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India," *The Quarterly Journal of Economics*, 2010, *125* (4), 1727–1767.

Harberger, Arnold C., "Monopoly and Resource Allocation," *The American Economic Review*, 1954, *44* (2), 77–87.

Harberger, Arnold C, "Using the resources at hand more effectively," *The American Economic Review*, 1959, *49* (2), 134–146.

Harrington, Joseph E, "How do cartels operate?," *Foundations and Trends in Microeconomics*, 2006, *2* (1), 1–105.

Hendricks, Kenneth and Robert H Porter, "An empirical study of an auction with asymmetric information," *The American Economic Review*, 1988, pp. 865–883.

Holmes, Thomas J and James A Schmitz, "Competition and Productivity: A Review of Evidence," *Annual Review of Economics*, 2010, *2*, 619–642.

Hopenhayn, Hugo, "Firms, Misallocation, and Aggregate Productivity: A Review," *Annual Review of Economics*, 2014, *6*, 735–770.

Hotelling, Harold, "The Economics of Exhaustible Resources," *Journal of Political Economy*, 1931, *39* (2), 137–175.

Hsieh, Chang-Tai and Peter J Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Hyytinen, Ari, Frode Steen, and Otto Toivanen, "Cartels uncovered," *Available at SSRN 1793665*, 2011.

Kellogg, Ryan, "The effect of uncertainty on investment: evidence from Texas oil drilling," *The American Economic Review*, 2014, *104* (6), 1698–1734.

Kilian, Lutz, "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," *American Economic Review*, 2009, *99* (3), 1053–1069.

Kohl, Wilfrid L, "The Perfect Storm," *Harvard International Review*, 2005, *26* (4), 68.

Lipsey, Richard G and Kelvin Lancaster, "The general theory of second best," *The Review of Economic Studies*, 1956, *24* (1), 11–32.

Marshall, Robert and Leslie Marx, *The Economics of Collusion*, Cambridge, MA: MIT Press, 2012.

Melitz, Marc J, "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

Olley, G Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica: Journal of the Econometric Society*, 1996, pp. 1263–1297.

Organization of Petroleum Exporting Countries, "Monthly Oil Market Report," June 2008.

Porter, Robert H, "Detecting collusion," *Review of Industrial Organization*, 2005, *26* (2), 147–167.

Restuccia, Diego and Richard Rogerson, "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic dynamics*, 2008, *11* (4), 707–720.

Roller, Lars-Hendrik and Frode Steen, "On the Workings of a Cartel: Evidence from the Norwegian Cement Industry," *American Economic Review*, 2006, *96* (1), 321–338.

Rystad Energy, "Ucube Technical Handbook," Technical Report 2015.

Slade, Margeret and Henry Thille, "Whither Hotelling: Tests of the Theory of Exhaustible Resources," *Annual Review of Resource Economics*, 2009, *1*, 239–260.

Sweeney, James, "Economic theory of depeletable resources: An introduction," in Allen Kneese and James Sweeney, eds., *Handbook of Natural Resource and Energy Economics*, Vol. 3, CHECK, 1992, chapter 17.

Syverson, Chad, "Product substitutability and productivity dispersion," *Review of Economics and Statistics*, 2004, *86* (2), 534–550.

__ , "What Determines Productivity?," *Journal of Economic Literature*, 2011, *49* (2), 326–365.

Whinston, Michael, *Lectures on Antitrust Economics*, Cambridge, MA: MIT Press, 2006.

Yergin, Daniel, *The prize: The epic quest for oil, money & power*, Simon and Schuster, 1991.