(Mis)Allocation, Market Power, and Global Oil Extraction

By JOHN ASKER AND ALLAN COLLARD-WEXLER AND JAN DE LOECKER*

We propose an approach to measuring the misallocation of production in a market that compares actual industry cost curves to undistorted (counterfactual) supply curves. As compared to traditional, TFPR-based, misallocation measures, this approach leverages cost data, such that results are readily mapped to welfare metrics. As an application, we analyze global crude oil extraction and quantify the extent of misallocation therein, together with the proportion attributable to market power. From 1970 to 2014, we find substantial misallocation, in the order of 744 billion USD, 14.1 percent to 21.9 percent of which is attributable to market power.

JEL: D2, L1, L4, L72
Keywords: Market Power, Productive Inefficiency, Misallocation, Oil.

I. Misallocation and productive inefficiency

The aggregate impact of the misallocation of production across an economy's productive units has attracted considerable attention in recent years.[1] This production misallocation, and the resulting welfare loss, occurs through more production being allocated to less productive (higher-cost) units of production, and less production to the most efficient production units in the economy. Much of the extant literature on misallocation focuses on measuring misallocation by examining dispersion in establishment-level revenue total factor productivity (TFPR) within industries or entire economies.

Restuccia and Rogerson (2013, 2017) decompose this literature into the *direct approach*, which looks at evidence of misallocation arising from specific observable sources, and the *indirect approach*, which identifies distortions as deviations, or wedges, from a specific model, to evaluate the size of overall misallocation in a economy or market. In this paper we develop a hybrid of these two approaches

* Asker: Columbia University, Department of Economics, 420 West 118th Street New York, NY 10027. Email: johnasker@gmail.com. Collard-Wexler: Duke University and NBER. 230 Social Sciences Building, Durham, NC 27708. Email: collardwexler@gmail.com. De Loecker: KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium. Email: jan.deloecker@kuleuven.be. We would like to thank Russell Cooper, Martin Hackmann, Volker Nocke, Michael Whinston, Rob Porter, Ryan Kellogg, 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, Andrew Steck, Sherry Wu and Yilin Jiang, Cameron Birchall, and Enrico Camarda. Financial Assistance from Princeton University's Industrial Organization Group, NSF (SPS #220783) and Odysseus Type 1 (FWO) Grant is greatly appreciated.

[1] Contributions to this literature include Banerjee and Duflo (2005); De Mel et al. (2008); Restuccia and Rogerson (2008); Hsieh and Klenow (2009); Bartelsman et al. (2013); Asker et al. (2014); David et al. (2016); Gopinath et al. (2017). Surveys of this literature include Restuccia and Rogerson (2013, 2017) and Hopenhayn (2014).

by estimating the aggregate extent of misallocation and then allowing the aggregate misallocation measure to be decomposed into specific sources (here, market power and other factors). The approach also shifts the focus of measurement away from TFPR toward the cost of production.[2] Focusing on costs provides an alternative, data driven, approach to measuring misallocation and allows results to be readily mapped to welfare metrics.

We apply our approach to the global oil industry, using detailed information on production and costs from 13,248 oil fields, covering 92% of world production, from 1970 to 2014. The contribution of market power to aggregate misallocation is investigated by considering the production patterns of OPEC and its member states. Before delving deeper into the specifics of our application, it is helpful to describe our approach to measurement, and its relation to the broader misallocation literature, in general terms.

The point of departure of this paper is to focus on the area measured by comparing the realized resource cost of production (the area under the actual marginal cost curve) to the efficient resource cost of production (the area under the marginal cost curve achievable in absence of any distortion). To illustrate why this is a measure of welfare loss arising from misallocation, consider Figure 1. Figure 1 presents a stylized, graphical, representation of a market in which market power is the sole imperfection. (The example is easily adapted to any distortion that creates a tax wedge. What is required are wedges that differentially impact productive units.).

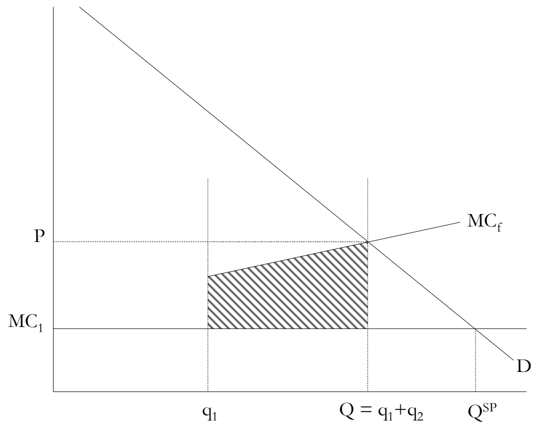
In Figure 1 there exists a producer with market power, with constant marginal cost MC_1. Also present in the market is a competitive, price-taking, fringe that has an aggregated marginal cost curve given by MC_f. The market price is equal to P, and the quantity produced by the (low-cost) producer with market power, q_1, is less than total production $Q = q_1 + q_2$, where q_2 is the production of the fringe.[3] 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. It is this welfare cost, the *rectangle*, that we take as our measure of the full extent of misallocation. This measure is referred to as *productive inefficiency*.

To measure the full extent of productive inefficiency in a market (the shaded rectangle in Figure 1) it is necessary to observe the realized level of production and the marginal costs of the infra-marginal units of production. In our setting, these are data. The central challenge is to construct the marginal cost schedule absent distortions. For this, a model of efficient production is required, as given by firm behavior in a competitive equilibrium, or equivalently, as the solution

[2] TFPR is revenue-based TFP (Foster et al., 2008). That is, if production (Q), for a given producer f, relies on a bundle of inputs (X) using the production function $Q_f = A_f X_f$, then A_f is quantity-based TFP (TFPQ) and $P_f \times A_f$ is TFPR (that is, TFP when outputs are measured via revenue).

[3] 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.

Figure 1. : Production misallocation (resulting from market power)



Notes: q_1 indicates total production from the cartel, while q_f indicates production from the competitive fringe. The producer with market power 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.

to a social planner problem. Such a model, when combined with data, can deliver the required marginal cost schedule. Once this is completed, the extent of misallocation, measured in welfare relevant terms, can be recovered.

A typical approach to measuring misallocation has been to assume that one market or economy (often the U.S.) is undistorted, and use this as a benchmark against which to measure the extent to which other economies suffer from misallocation (see, for example, Hsieh and Klenow (2009) and Restuccia and Rogerson (2008)). By leveraging detailed micro-data on costs, what is described here is a model-based alternative to this benchmark approach, which has the benefit of being sufficiently micro-founded as to have a clear welfare interpretation.

This approach to measurement can be further extended if a plausible source of a particular distortion is observed, or alternatively by explicit modeling the distortion. Once the full extent of misallocation is recovered, this lets additional structure, separating out (say) market power, to be imposed to asses the extent to which that distortion contributed to the overall level of misallocation. In this sense, the cost-based approach to measurement explored in this paper allows Rogerson's direct and indirect approaches to studying misallocation to be combined. In applications, a persistent challenge is allowing measurement to account for the interaction between observed and unobserved sources of distortion. The approach illustrated in this paper accounts for this interdependence, and so relates the theory of the second best directly to the misallocation literature.

Lastly, the distinction between the productive inefficiency arising from misallocation, and the more familiar dead weight loss triangle arising from market imperfections is made readily apparent in Figure 1. Both welfare losses are, of course, well appreciated.[4] This paper distinguishes and quantifies the former source of welfare loss, as this is that part of the total welfare loss arising from market imperfections that speaks directly to the extent of misallocation.

A. The link between misallocation and productive inefficiency

A direct link exists between the use of plant-level dispersion in TFPR as an indicator of misallocation, as employed by Hsieh and Klenow (2009) and much of the literature that follows, and the measures explored here, which exploits dispersion in (plant-level) marginal costs of production to estimate productive inefficiency. Let the relevant productive-unit (firm, plant or similar) be indexed by f and the degree of misallocation be indexed by M. In Hsieh and Klenow (2009), this given by

$$M = \sqrt{\sum_f \left(\mathrm{TFPR}_f - \overline{\mathrm{TFPR}_f}\right)^2}, \tag{1}$$

where TFPR is revenue-based productivity (Foster et al. (2008)), and $\overline{\mathrm{TFPR}_f}$ is average TFPR. That is, the standard deviation in TFPR is used as a measure of misallocation.[5]

To illustrate the link between this measure of misallocation, and that levered here, consider the following simple model. The production function is $Q_f = A_f X_f$.[6] Further, in keeping with Figure 1, assume that output is homogenous over productive units, and that the market for inputs is competitive, implying a common output, P, and input prices, P_X. Noting that $\mathrm{TFPR} = P \times A_f$, this implies that dispersion in TFPR is proportional to dispersion in TFPQ, quantity based productivity A_f. Marginal cost is given by $c_f = P_X / A_f$. It follows that, after a small amount of algebra, the degree of misallocation M can be computed as

$$M \;=\; P \times P_X \times \sqrt{\sum_f \left(1/c_f - \overline{1/c_f}\right)^2}. \tag{2}$$

Hence, dispersion in the inverse of marginal cost is proportional to dispersion in TFPR, so examining dispersion through the lens of TFPR or marginal cost is equivalent.

In the Hsieh and Klenow (2009) setting, the mapping from TFPR to TFPQ relies

[4] Borenstein et al. (2002) and Cicala (2017) quantify the extent of productive inefficiency, due to market power in the wholesale electricity market. In electricity markets, since retail prices are fixed, the demand curve is completely inelastic, and thus there is no quantity distortion, only productive distortions.

[5] Hsieh and Klenow (2009) also use several other measures of dispersion. We focus on the standard deviation of TFPR as it seems to be the one most often used.

[6] X_f can be thought of as an input index for a constant returns to scale technology.

on model assumptions on conduct and the shape of demand curve. Therefore, any model misspecification on those two dimensions may lead the analyst to infer the presence of distortions. Haltiwanger et al. (2018) note that the impact of any model misspecification in the Hsieh and Klenow (2009) setting enters through either the implied misspecification of the elasticity of output price to marginal cost or the elasticity of marginal cost to TFPQ.

To the extent that we depart significantly from the existing misallocation literature, it is in using productive inefficiency (the shaded area in Figure 1) as a measure of the quantitative extent of misallocation. Merely examining the dispersion in marginal costs may be misleading in comparing different markets. Consider two markets, indexed by $j \in \{1, 2\}$, each with two firms. Firm A_j has marginal cost of 5, and firm B_j has marginal cost of 10. Both markets have the same dispersion of marginal costs. However, if in the first market, firm A_1 has a market share of 90 percent, while in the second market firm A_2 only has market share of 50 percent, we would say there is more misallocation is the second market.[7]

Focusing explicitly on productive inefficiency also avoids confounding the welfare loss due to output restrictions (the deadweight loss triangle in Figure 1) from arising from a distortion, with its impact on misallocation. For instance, a pure monopoly in a market, like that represented in Figure 1, will have no TFPR or marginal cost dispersion, but will still impose a welfare cost (the dead weight loss triangle). Hence, it is the productive inefficiency that speaks directly to misallocation.

B. Application to global oil extraction

A predicate for misallocation is inter-firm heterogeneity in productivity, or the costs of production. Given this, the application in this paper is the global upstream oil industry – an industry with notable exogenous variation in costs across productive units, in large part attributable to differences in geology.[8] For instance, the world's largest oil field, the Ghawar field in Saudi Arabia, has average costs (in 2014 USD) of approximately $3 per barrel over the timeframe covered by our data. 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.

The oil fields in low-cost, OPEC, countries have very large reserves and are

[7] Hsieh and Klenow (2009) measure the full extent of misallocation by exploiting their general equilibrium framework to optimally redeploy resources and measuring the resulting output expansion. In the partial equilibrium setting here, this metric would be difficult to employ as redistributing resources within a single industry may result in more output being generated than is demanded, resulting in a perverse welfare loss.

[8] The literature has established that producers that compete in narrowly defined product markets can have very different levels of productivity (see Syverson (2011, 2004); Foster et al. (2008)). 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 individual producers, and industry-wide performance. See Olley and Pakes (1996); Melitz (2003); Syverson (2004); Goldberg et al. (2010); Holmes and Schmitz (2010); Edmond et al. (2015); Atkeson and Burstein (2010); De Loecker (2011); Collard-Wexler and De Loecker (2015) and De Loecker et al. (2016).

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.[9] This implies production 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 arising from misallocation.

We leverage a major feature of the oil industry that allows us to measure misallocation without being subject to the standard model misspecification concerns: oil is by and large a homogeneous good, a commodity. This makes the presence of differences in the observed cost of production immediately informative about differences in welfare-relevant resource costs. Second, the data set we utilize allows us to directly observe the cost of production by oil field, and, as such, gives a data environment in which to deploy the approach described above without significant additional modeling.

That is, in terms of Figure 1, the object to be estimated is the proportion of production of the low-cost firm 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 observed market equilibrium. In particular, by holding the (observed) market-level output fixed, this approach avoids having to model the existence and workings of the OPEC cartel, which, in the context of the world oil market, is a complicated matter. It can, similarly, take an agnostic approach to the way that other distortions manifest.

As foreshadowed in the previous paragraph, OPEC is a notable feature of the global oil industry. For our purposes, OPEC gives rise to an observable driver of market-power and, hence, misallocation (providing the link to the direct approach discussed above). Obviously, any possible impact of OPEC likely interacts with the many other likely sources of distortion in the global oil industry. It is well understood that there are a variety of other sources potentially giving rise to misallocation, including (and by no means limited to) geopolitics, within-country corruption, taxation, and an oft-expressed desire for self-sufficiency on the part of many sovereign states.[10] The decomposition of the full extent of misallocation into that attributable to OPEC-related market power, and that attributable to other channels, raises conceptual issues closely related to the issues raised in Lipsey and Lancaster (1956)'s articulation of the theory of the second-best.

Despite all these these attractive features, a complication arising in measuring misallocation in the the oil market is the finite resource extraction problem embedded in oil production. This creates inter-temporal linkages of supply. By leveraging rich micro-data and a flexible dynamic framework, productive ineffi-

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

[10] A particularly interesting factor in the context of developed countries, like the US, is the incentive to demonstrate technological progress in developing oil production in what were thought to be infeasible locations in order to raise proven reserves.

ciency can be computed accounting for these dynamics

Applying this approach the preferred specification leads to a net present value measure of misallocation, from 1970, of 744 billion in 2014 USD, where the net present value measure of the realised cost of extraction is measured at 2,499 billon. Of this misallocation, we attribute 14.1 to 21.9 percent to market power (lower and upper bounds). The remaining misallocation likely comes from a wide variety of sources – taxes are significant in this industry in many jurisdictions, and political economy distortions and disruptions due to wars are also notable. We discuss evidence relating to all these sources of misallocation.

While we tilt the measurement approach to be somewhat conservative, measurement error and expectational errors remain potentially problematic for our estimates. Measurement error is a potential property of the data, and may result in the analysis finding misallocation where the is none. This would arise if two fields are such that field A has a lower extraction cost than field B, but in our data measurement error leads us to infer the opposite.[11] As an indication of the extent of robustness to this, if the measured cost of every field is multiplied by an i.i.d. random variable uniformly distributed between [0.5,1.5], and the analysis re-done, 50 iterations of this procedure generate a 10-percentile misallocation number of 692 billion and 90-percentile misallocation number of 950 billion.[12]

Expectation error is similarly confounding. A concern is that, since reserves and other technical features of a deposit are somewhat uncertain at the point at which a field is first exploited, it may be that while ex ante a field is cheap, ex post, in our data, it is expensive (the converse also applies). This would lead to a finding of misallocation, where such a finding is not warranted when viewed ex ante. The exercise described in the previous paragraph, in which a random variable is added to the measured costs, suggests that our qualitative results would be similar, even if field operators had somewhat imprecise measures of the costs of a field when deciding on operating plans. Additionally, the influence of expectation errors may also be tempered, at least for large conventional fields, by the ability of operators to scale production up and down over time. That said, we have no precise way to pin down the extent to which our numbers are confounded by either measurement error or expectational errors.[13]

This paper is organized as follows. Section 2 presents a short description of the oil market, to which we apply our empirical framework, 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

[11] Almost certainly, our data has some measurement error. Given the variety of sources from which the data is drawn, and the range of countries covered, some measurement error is inevitable (see the Online Appendix for more details.). This is a feature of much of the data used in the misallocation literature more generally, see White et al (2018).

[12] That is, we simulate the impact of a random measurement error at the field-level equivalent to a maximal change in costs of +/- 50%. Note that the unit of observation in our data is the field-year, so the measurement error in this exercise is persistent over years. See the Online Appendix for more details.

[13] Note that a field can contain many wells so that, even if the cost of capping a well is prohibitive given market conditions, production can be scaled by adjusting the rate at which wells are replaced and added.

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.

II. 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.[14]

A. *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 oil production area.[15] Fields vary considerably in the number of wells and the associated infrastructure.[16]

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

[14] The Online Appendix (`http:public.econ.duke.edu/~ac418/OnlineAppendix_Misallocation_Oil.pdf`) provides the reader with a more detailed discussion of the data sources, measurement and on the specifics of oil production. The code that was used for this project, but not the proprietary Rystad data used in the analysis, can be found at (`http://public.econ.duke.edu/~ac418/Replication_Files_No_Data_Oil_Misallocation.zip`).

[15] Often coinciding with common management and ownership.

[16] 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.

series of measurement issues. The first is how to measure the quantity associated with a deposit in units 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.[17]

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 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 pre-determined by geology; the fundamental starting point for the analysis. Section IV.A 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.

B. *The global oil market*

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

[17] 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 a discussion.

many private firms, while Saudi Arabia has a nationalized oil company (Saudi Aramco).

Table 1—: Largest crude producers, % of global production 1970-2014

OPEC		Non-OPEC	
Saudi Arabia	11.8%	United States	14.4%
Iran	5.4%	Russia	13.0%
Venezuela	3.8%	China	4.1%
UAE	3.1%	Mexico	3.7%
Nigeria	2.8%	Canada	3.3%
Iraq	2.7%	UK	2.4%
Kuwait	2.6%	Norway	2.4%

Notes: Global production from 1970-2014 was 1,156 billion barrels. Collectively, these 14 countries account for 85.4% of global production.

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.[18] 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 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".[19] 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.

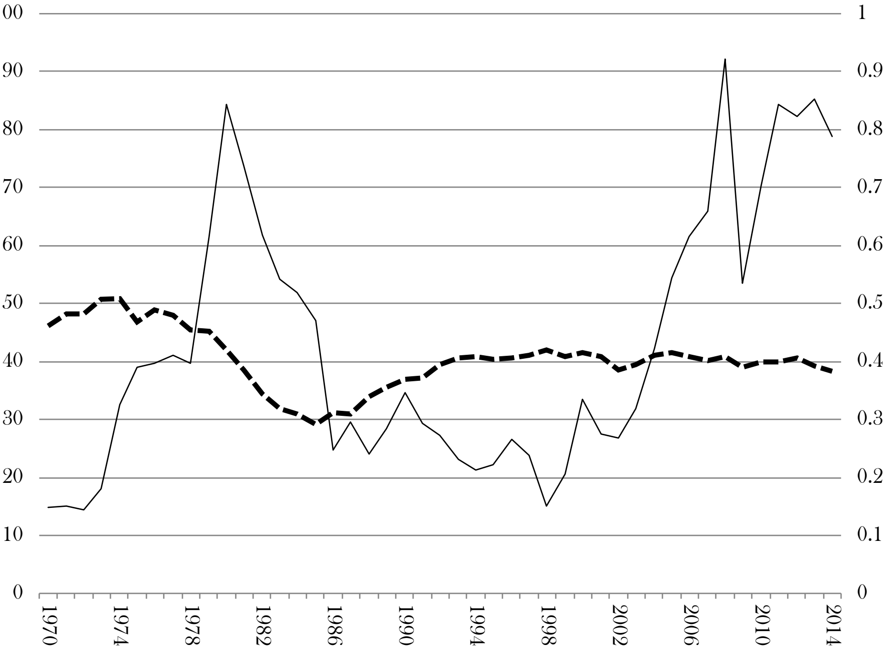
Figure 2 shows OPEC's market share and the price of crude from 1970 to 2014. Before OPEC started coordinating extensively on price reductions, it had a global production share fluctuating around 48 percent. This fell to a low point of 29.2

[18] 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).

[19] `www.opec.org` accessed 4/10/17.

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.[20]

Figure 2. : OPEC market share and oil price (1970-2014)



Notes: The vertical axis on the left is in dollars and corresponds to the annual average oil price, which is indicated by the black line. This price series is deflated with the US GDP deflator (base year 2009). The OPEC market share in each year is indicated by the dashed black line. The vertical axis on the right indicates the level of the market share. Countries are included in OPEC in all years if they had ever had active membership between 1970 and 2014.

III. 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.[21] Thus, we need to adopt a definition of productive inefficiency appropriate for a dynamic context.

[20] For more detail on OPEC and its history see the online appendix.

[21] 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. (2018).

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.

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 (or, equivalently, the policy of the social planner), this section builds the underlying cost function that is used to guide measurement and modeling. It also provides the algorithm used to compute the solution to the social planner's problem (equivalently, a competitive equilibrium). 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.

A. 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:

$$\begin{aligned} q_{ft} &= \min\{\alpha_{ft}K_{ft}, \gamma_{ft}L_{ft}\} \\ &s.t. \quad q_{ft} \leq R_{ft},\ R_{ft} = R_{ft-1} - q_{ft-1} \\ &\qquad R_{f0} > 0\ R_{ft} \geq 0, \end{aligned} \tag{3}$$

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}.[22]

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:

$$C\left(q_{ft}\right)=\left(\frac{w_{ft}}{\gamma_{ft}}+\frac{r_{ft}}{\alpha_{ft}}\right)q_{ft}. \tag{4}$$

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

$$\frac{w_{ft}}{\gamma_{ft}}=\frac{w_f}{\gamma_f}\mu_{ft} \quad \text{and} \quad \frac{r_{ft}}{\alpha_{ft}}=\frac{r_f}{\alpha_f}\mu_{ft}. \tag{5}$$

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:

$$C\left(q_{ft}\right)=\left(\frac{w_f}{\gamma_f}+\frac{r_f}{\alpha_f}\right)\mu_{ft}q_{ft}. \tag{6}$$

Marginal cost is then given by:

$$c_{ft}=MC\left(q_{ft}\right)=AC\left(q_{ft}\right)=\begin{cases}c_f\mu_{ft} & \text{if } q_{ft}\leq R_{ft}\\ +\infty, & \text{otherwise,}\end{cases} \tag{7}$$

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 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 $k\geq 1$.[23] 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

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

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

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}.

B. 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:[24]

$$\begin{aligned} & \max\nolimits_{\{q_{ft}\}} \textstyle\sum_{t=1}^{T} \delta^{t-1} \left(p_t - c_f\right) q_{ft} \\ \text{s.t.} \quad & R_{f0} \geq \textstyle\sum_{t=1}^{T} q_{ft}, \quad \text{and} \quad q_{ft} \geq 0 \;\; \forall t \in \{1, \cdots, T\}. \end{aligned} \tag{8}$$

Proposition 1 and corollary 1 together establish that the lowest-cost fields are extracted first in any competitive equilibrium.[25]

PROPOSITION 1: *Let marginal costs be described by equation 7. 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, $E\left(\mu_{ft} \,|\mu_{f1}\right) = 1$. Thus $E\left(c_f \mu_{ft} \,|c_f \mu_{f1}\right) = c_f$. We assume by contradiction that $\overline{q}_1 = 1$ and $\underline{q}_1 = 0$. Then, there must exist periods such that

$$\delta^{t-1}\left(p_t - \underline{c}\right) \geq \left(p_1 - \underline{c}\right) \tag{9}$$

and

$$\delta^{t-1}\left(p_t - \overline{c}\right) \leq \left(p_1 - \overline{c}\right), \tag{10}$$

[24] 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 VI.B.

[25] 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.

where at least one inequality is strict. Assume, for exposition, that the inequality in equation 10 is strict.

From equation 9,

$$\delta^{t-1}\left(p_t - \overline{c}\right) + \delta^{t-1}\left(\overline{c} - \underline{c}\right) \geq \left(p_1 - \overline{c}\right) + \left(\overline{c} - \underline{c}\right), \tag{11}$$

Since $\delta^{t-1}\left(\overline{c} - \underline{c}\right) < \left(\overline{c} - \underline{c}\right)$, this implies that $\delta^{t-1}\left(p_t - \overline{c}\right) \geq \left(p_1 - \overline{c}\right)$, which is a violation of equation 10.

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 when low-cost fields are not being exploited prior to higher-cost fields coming on line, this is an indication of the presence of market distortions.[26] The implication that in a competitive equilibrium, lower cost resources are extracted first has previously been noted by, for instance, Herfindahl (1967) and Solow and Wan (1976).

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 (or equivalently, the social planner's solution), 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.

[26] Consistent with this, 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. See Sweeney (1993) for an extended discussion of the comparison of competitive equilibrium and equilibrium with market power in these settings.

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.

IV. Descriptive evidence of productive 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.

A. 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.[27,28]

Given the Leontief production function, yielding the cost function given by equation (7), 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:

$$c_{ft} = \frac{\sum_h \text{ Expenditure}_{hft}}{q_{ft}}, \tag{12}$$

where the various expenditure categories are h={ Well Capital, Facility Capital, Abandonment cost, Production Operating,Transportation Operating, and SGA},

[27] 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 the Online Appendix.

[28] Almost certainly, our data has some measurement error. Given the variety of sources from which the data is drawn, and the range of countries covered, some measurement error is inevitable (see the brief discussion in the introduction and longer discussion in the Online Appendix for more details).

Table 2—: Summary statistics, by field-year

Variable	mean	median	5%	95%
Field-year characteristics:				
Production (mB/year)	3.43	0.22	0.00	10.92
Reserves (mB)	99.49	3.71	0.03	239.78
Discovery Year	1965	1967	1911	1999
Startup Year	1971	1974	1916	2005
Off-shore	0.19			
Costs: ($m)				
Exploration Capital Expenditures	0.61	0.00	0.00	0.41
Well Capital Expenditures	9.10	0.49	0.00	35.32
Facility Capital Expenditures	5.14	0.21	0.00	16.85
Production Operating Expenditures	10.41	0.46	0.00	38.47
Transportation Operating Expenditures	2.27	0.13	0.00	7.01
SGA Operating Expenditures	2.65	0.22	0.00	8.85
Taxes Operating Expenditures	1.41	0.00	0.00	1.09
Royalties	18.19	0.40	0.00	45.36
Government Profit Oil	15.59	0.00	0.00	21.00

Notes: Only fields with active production during 1970-2014 are included. There are 66,920 fields in the Rystad data. 13,248 of these fields have active production. Reserves data exists for 13,298 fields. As a result, in Section V, 11,457 fields are used. All numbers are in $US deflated by the US GDP deflator for 2009. mB indicates million barrels. The unit of observation is the field-year.

and all expenditures are deflated by the US GDP deflator with 2009 as the base year.[29] All expenses are recorded in the year they are incurred.[30] 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.[31] The Leontief technology assumption makes this supply curve a step function. Modeling each well

[29] It is notoriously difficult to accurately represent the economic relevant capital expenditures in any industry, and the oil industry is not different. We therefore subject our analysis to a variety of robustness checks. See, in particular, section VI.B. Robustness of results with respect to the inclusion or omission of capital expenditures is discussed in the online appendix.

[30] Note that our approach to estimating costs used in the quantification of misallocation in section 5 (see, in particular, section 5.1) implicitly averages capital costs over the observed life of a field.

[31] In Section V 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.

and aggregating up would, at best, put a small amount of curvature in each step, which, given the level of aggregation, 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 VI.C).

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 V,) P50 values at an oil price of $70 a barrel are used to report reserves. In section V, 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.

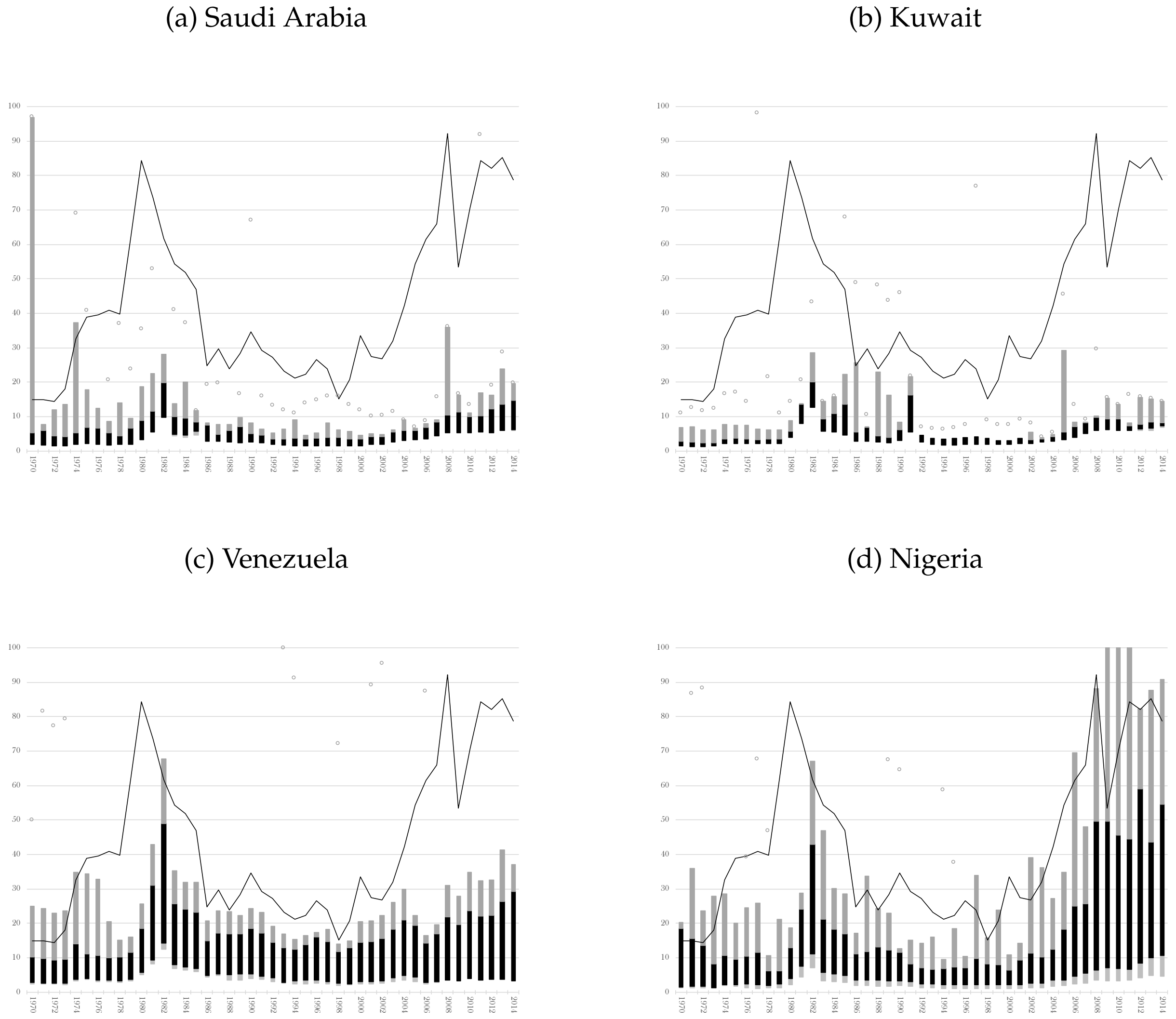
B. Preliminary Evidence

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). In the context of the wider misallocation literature, these figures are the analog of the standard deviations of TFPR reported in, for instance, Hsieh and Klenow (2009) (see section I.A).

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

Figure 3. : Production costs and price (1970-2014):
OPEC countries



(a) Saudi Arabia

(b) Kuwait

(c) Venezuela

(d) Nigeria

Notes: Each panel plots the dispersion of the costs of production (by barrel) in a country, and the price of oil. The vertical axis is $/barrel, from 0 to 100 in increments of 10. The horizontal axis is in years, from 1970 to 2014. Costs are indicated by the bars and circles. The (grey and black) bar indicates the range of costs within the 1st and 99th percentiles of production. That is, the cheapest, and most expensive, 1% of barrels produced in the year are excluded. The black portion of the bar indicates the 5th to 95th percentiles range. Circles indicate the maximum cost per barrel incurred in a year. Where a cost exceeds $100 per barrel, it is not shown (the vertical axis is truncated at 100) – this accounts for many of the maxima not being visible, for instance. The oil price is indicated by the black line. All series have been deflated with the US GDP deflator (base year 2009). All costs are measured according to the baseline specification.

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.[32]

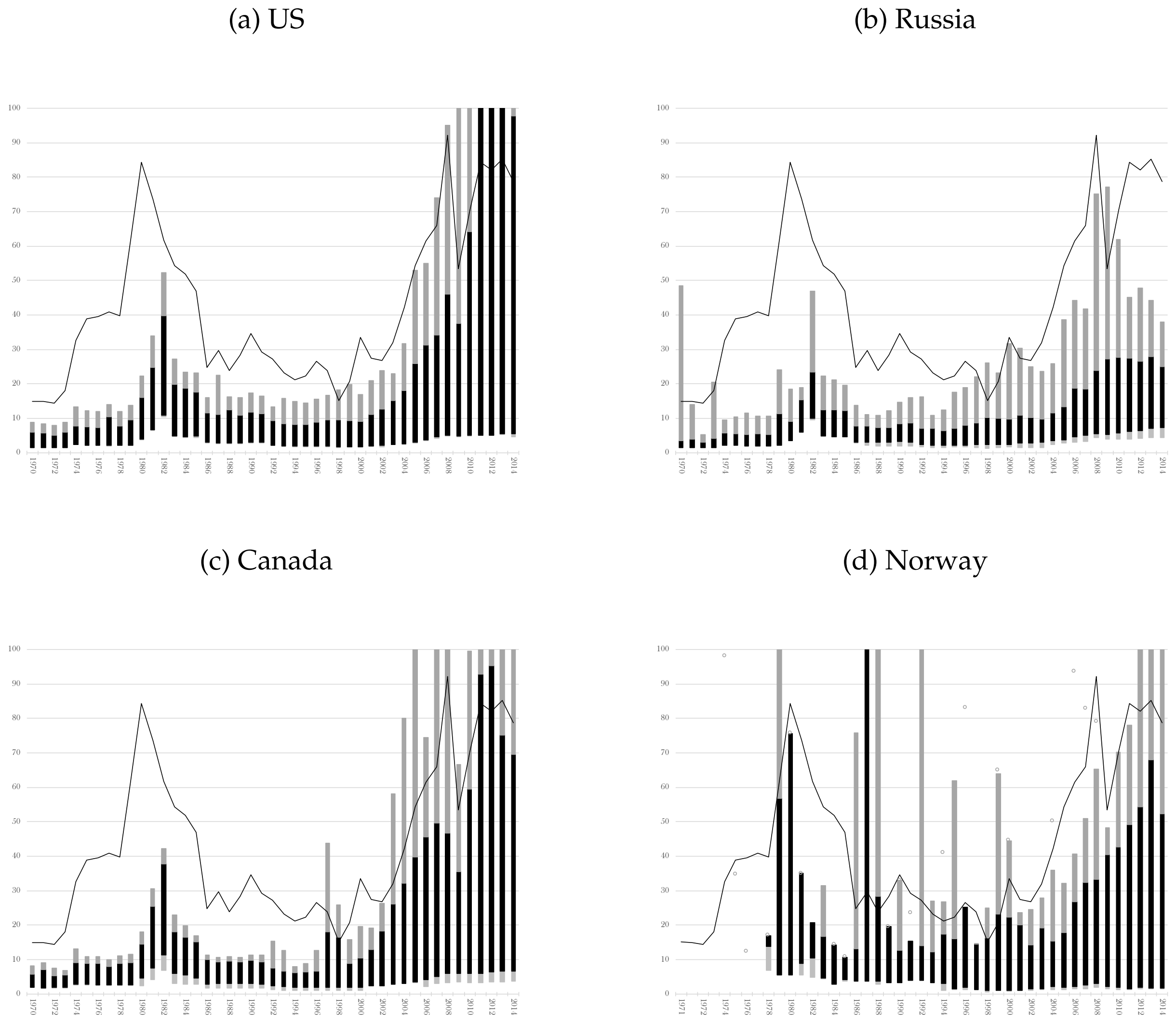
Figure 4 allows us to compare the within-OPEC patterns in Figure 3 with those in non-OPEC countries. Panels (a) and (b) of Figure 4 show the US and Russia, the two biggest oil producers between 1970 and 2014.

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

[32] 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.

Figure 4. : Production costs and price (1970-2014):
Other Countries



(a) US

(b) Russia

(c) Canada

(d) Norway

Notes: The notes for Figure 3 also apply to this figure. The Norwegian panel reflects little meaningful production prior to 1978.

often brings, and the large scale of the infrastructure involved.[33] 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 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 OPEC restricting production, relative to reserves, more than producers outside the cartel. As one might expect, the data are consistent with OPEC, and Saudi Arabia in particular, exercising market power.[34]

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 re-

[33] 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. Setup costs are discussed in Section VI.B.

[34] 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.

Table 3—: Reserves and production, 2014

	Reserves	Share of world reserves	$\frac{\text{Reserves}}{\text{Annual production}}$
	(mB)	(%)	(%)
Non-OPEC	218,054	50	10
Russia	46,134	11	12
Canada	36,622	8	43
United States	31,735	7	7
Norway	6,962	2	10
OPEC	220,561	50	19
Saudi Arabia	74,194	17	18
Venezuela	17,523	4	19
Kuwait	15,723	4	16
Nigeria	7,952	2	10

Notes: Data are for 2014. Total reserves for the world in 2014 were 438 billion barrels. The ratio of reserves-to-production was 14. OPEC countries are listed in Section II.B. Countries are included in OPEC in all years if they had ever had active membership between 1970 and 2014. Reserves are reported using P50 measures at a world price of $70 per barrel.

allocation should get larger over time. This is the case regardless of whether the baseline or the alternative cost specification is used.

V. Quantifying the extent of misallocation

This section quantifies the extent to which production is misallocated, followed by quantification of market power as a specific source of misallocation. We do this by using the model described in Section III 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.

A. *Model parameterization*

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

Table 4—: Unit costs across the global oil industry (1970-2014)

	1970-1979	1980-1989	1990-1999	2000-2014
Number of active fields	4,766	7,088	9,760	12,085
Mean oil price	20	40	21	59
Mean global production (mB/year)	20,861	21,489	23,984	26,298
OPEC	9,979	7,289	9,606	11,249
Mean global reserves (mB)	737,928	728,532	661,815	517,559
OPEC	392,912	365,891	328,914	254,730
Unit costs (Baseline specification):				
95th percentile Saudi Arabia	5.8	13.6	4.4	10.4
Median Saudi Arabia	2.3	5.6	2.3	5.4
95th percentile OPEC	6.7	18.6	7.6	20.1
Median OPEC	2.4	5.9	2.8	6.1
95th percentile non-OPEC	6.7	15.6	9.2	28.2
Median non-OPEC	3.6	7.0	4.1	9.7
Unit costs (including taxes and royalty payments):				
95th percentile Saudi Arabia	5.8	13.6	4.4	10.4
Median Saudi Arabia	2.3	5.6	2.3	5.4
95th percentile OPEC	30.2	53.6	21.1	79.1
Median OPEC	2.8	13.6	6.5	12.0
95th percentile non-OPEC	26.3	40.1	20.3	75.3
Median non-OPEC	9.1	14.8	9.1	24.0

Notes: The unit cost is computed as per Section IV.A and top-coded at $100. The unit of observation for unit cost is the barrel. Percentiles and medians are calculated at the barrel level. All prices and costs are deflated with the US GDP deflator (base year 2009). Reserves are reported using P50 measures at a world price of $70 per barrel. OPEC countries are listed in Section II.B. Countries are included OPEC in all years if they had ever had active membership between 1970 and 2014. Only fields active between 1970 and 2014 are included.

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.[35]

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 III 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.[36]

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 IV.A. 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

[35] 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.

[36] 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.

(onshore and offshore); and 3) measurement error, $\exp{(\epsilon_{ft})}$.[37] That is,

$$c_{ft} = c_f \mu_{ft} = c_f \mu_{st} \exp{(\epsilon_{ft})}. \tag{13}$$

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

$$\hat{\mu}_{st} = \sum_{f \in s} \kappa_{ft} \ln c_{ft}, \tag{14}$$

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:

$$(\ln c_{ft} - \hat{\mu}_{st}) = \ln \hat{c}_f + \epsilon_{ft}. \tag{15}$$

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^k_{st}. In the second step, for each field in the true dataset, the field-years are resampled. This allows us to estimate c^k_f using the μ^k_{st} from the first step.[38] This, in turn, allows $c^k_f \mu^k_{st}$ 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 μ_{st}, but also in the field-specific coefficient c_f. Fifty bootstrap iterations are used.

Our estimates of marginal costs effectively spread capital expenditures over the observed production lifetime of a field. The time-specific unit cost of production (as measured in equation (12)) is used to obtain a time-averaged field-specific marginal cost estimate. We then add back the production-weighted technology-time specific average cost. This approach, therefore, takes into account the use of capital over the asset's observed lifetime (taking a field fixed effect smoothes capex spikes out over the observed life of the asset), adjusting

[37] As noted in the introduction, measurement error more broadly defined may be a confounding factor. This specification addresses measurement error that manifests according to a stationary and ergodic process on the domain ft.

[38] The field-year observations used to compute μ^k_{st} are resampled independently from those used to compute each c^k_f. 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.

for some fluctuation across years in input prices.[39]

B. *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 measured as a stock in 1970 (we will present numbers deflated to 2014 dollars to make dollar numbers comparable across the paper).

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 II and IV, 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

[39] For assets with only a short history in the data this may introduce an upward bias in costs, but 58% of production is done by fields that are already producing in 1970. The approach we take avoids undue adhoc modeling and, by biasing fields exploited in later years toward having higher measured costs, pushes our analysis toward finding a lower misallocation measure. That is, these fields, which are observed to only produce in later periods, will also be biased toward comparatively later production in the counterfactual, by virtue of having higher measured costs, thus producing less measured misallocation. Specification 7 of Table 7 as discussed in section V.C addresses this concern by focusing only on fields already active in 1970.

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.[40] In the face of the inevitable censoring, we take a conservative approach. To project the path of "actual" production out past

[40] 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.

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.[41]

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 VI.A.

C. 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.

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 forecast our demand for oil beyond 2014 using a 1.3 percent annual growth rate.[42]

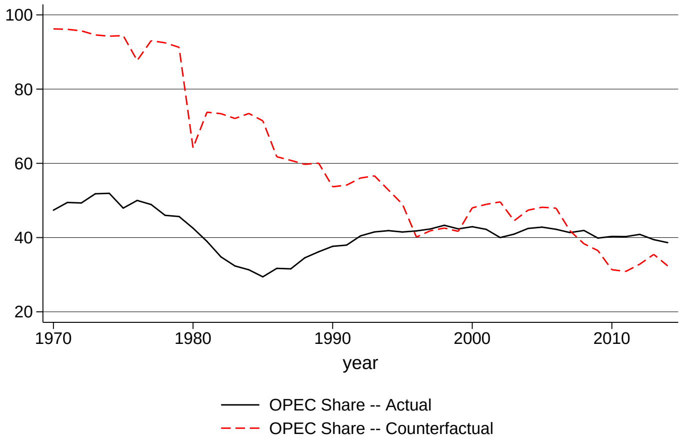
We begin by examining the counterfactual path computed by the unconstrained sorting algorithm. This path is compared to the actual path in Figure 5, which plots the market share of OPEC in the actual and counterfactual paths over time. On the actual (observed) path, the production share of OPEC fluctuates

[41] 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 geopolitical 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.

[42] 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.

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.

Figure 5. : OPEC Market Share 1970-2014



Notes: OPEC Share - Counterfactual presents the share of production accounted for by OPEC.

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

Table 5—: Dynamic counterfactual results (NPV of costs in billions of 2014 dollars)

	Timespan			
	1970-2014		1970-2100	
Actual (A)	2184	(125)	2499	(130)
Counterfactual (C)	1268	(76)	1756	(79)
Total distortion (A - C)	916	(124)	744	(112)
Decomposition of total distortion				
Within country (non-OPEC)	329	(80)	284	(41)
Within country (OPEC)	192	(46)	157	(72)
Across country (within non-OPEC)	163	(18)	139	(17)
Across country (within OPEC) (X)	85	(22)	58	(21)
Between OPEC and non-OPEC (Y)	148	(29)	105	(25)
Production distortion due to OPEC market power				
Upper bound (X+Y)	233	(42)	163	(38)
Lower bound (Y only)	148	(29)	105	(25)

Notes: The NPV of costs from 1970 to 2014, and to 2100 (exhaustion of all fields), are reported in billions of 2014 dollars (assuming a 5 percent discount rate). Results are for the baseline specification: a field extraction rate of 10 percent of reserves is imposed in the counterfactual: the p50 measures of reserves are used where needed, and a demand growth rate of 1.3 percent per year after 2014 is assumed. The Actual path is that observed in the data. The Counterfactual path is that computed using the unconstrained sorting algorithm. The within-country (non-OPEC) decomposition takes the path from the sorting algorithm in which all non-OPEC countries are constrained to produce their actual production. OPEC fields produce as in the data. The reported number is A - [the NPV of the costs of this path] = D1. The within-country (OPEC) decomposition is the mirror of this for OPEC countries (= D2). The across-country (within non-OPEC) decomposition takes the path from the sorting algorithm in which non-OPEC production is constrained to match the observed amount. OPEC fields produce as in the data. The reported number is A - D1 - [the NPV of the costs of this path] = E1. The across-country (within OPEC) decomposition is the mirror of this for OPEC countries (= E2). The Between OPEC and non-OPEC decomposition takes the path from the unconstrained sorting algorithm. The reported number is A - D1 - D2 - E1 - E2 - C = F1. Bootstrapped standard errors in parentheses using 50 bootstrap replications.

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. These distortions are explored empirically in section V.C.

Other potential sources of wedges, that may manifest at the within country level, are measurement error and expectation error. As noted in the introduction, measurement error can result in measured misallocation that is not actually present, while expectational error can result in ex post misallocation that may be absent when examined from an ex ante perspective. Accurately accounting for both these sources of measured misallocation is a challenge for the methodology exploited here, and and methodologies exploited elsewhere in the literature, since misallocation is obtained as a residual after imposing a particular model of production on the data. The major departure of this paper is that we bring to bear observable sources of misallocation at the micro level.

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 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.[43]

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.

Static productive inefficiency

The measure from the full dynamic model, in Table 5, 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 which misallocation varies by year, taking as initial conditions the actual market conduct in all previous years.

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

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

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).

Next, as noted in section V.C, within-country distortions can be attributed to wedges that move national production away from cost minimization. Figure 7 shows within-country variation across time in the extent of deviations from cost minimization. More precisely, the ratio of total observed cost and cost-minimizing (optimal) cost, taking the actual production up that year as an initial condition (as per static distortion calculations) and holding country production fixed, is computed for each year (i.e., annual within-country deviations from cost minimization). This ratio is then indexed to the level in 1970. The indexed ratio (vertical axis) is graphed by year (horizontal axis). Indexing allows time variation to be compared across countries on a convenient scale.

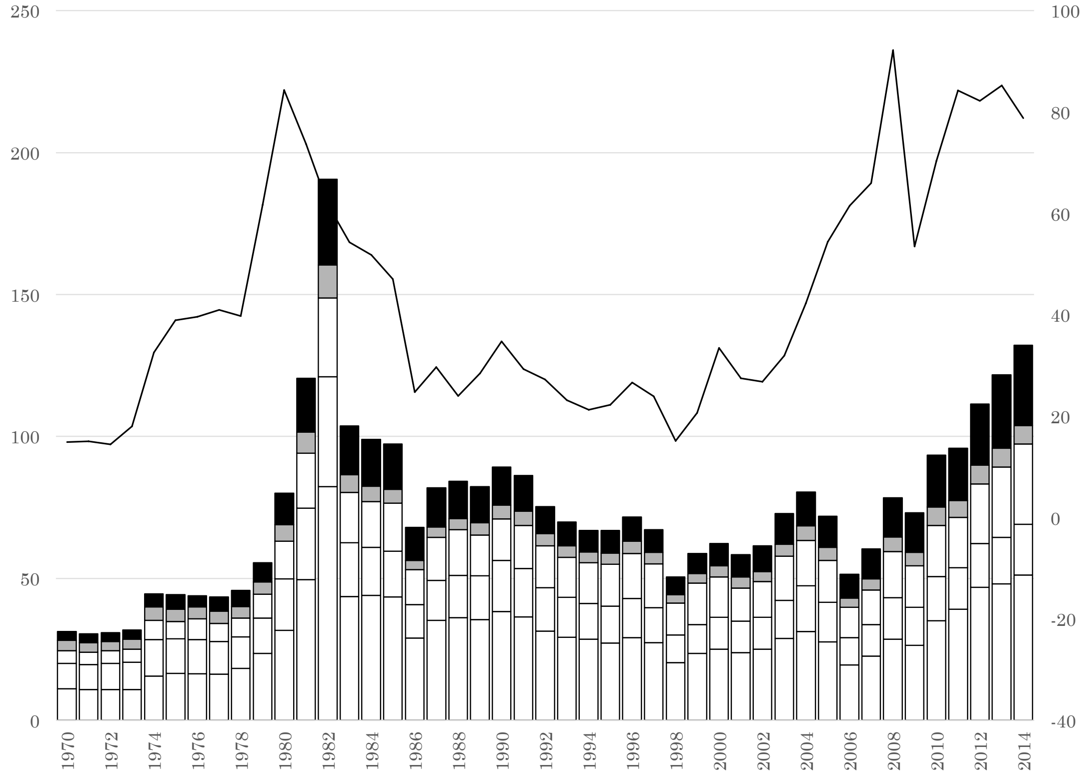
The top panel of Figure 7 shows the deviations over time for Iraq, Iran and Kuwait. Deviations from cost minimization increase in the 1980s, during the Iran and Iraq war. Subsequently, deviations decrease for Iran. By contrast, deviations in Iraq spike in the two gulf wars and the incursion by ISIL. Interesting, the first gulf war also sees a spike in deviation for Kuwait, reflecting the destruction of infrastructure by retreating Iraqi forces. These time-series here illustrate the influence of war in generating misallocation.

The bottom panel of Figure 7 shows the deviations over time for the USA, China and Russia. Deviations in the USA are stable until the late 2000s, when unconventional drilling increases substantially. We speculate that the increase in deviations are suggestive of expectational errors (or at least a divergence in expectations of profitability within the US oil sector). Russia and China both show marked declines in divergence from cost-minimization, consistent with their long-run transition to market based economies.

Last, we examine 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.

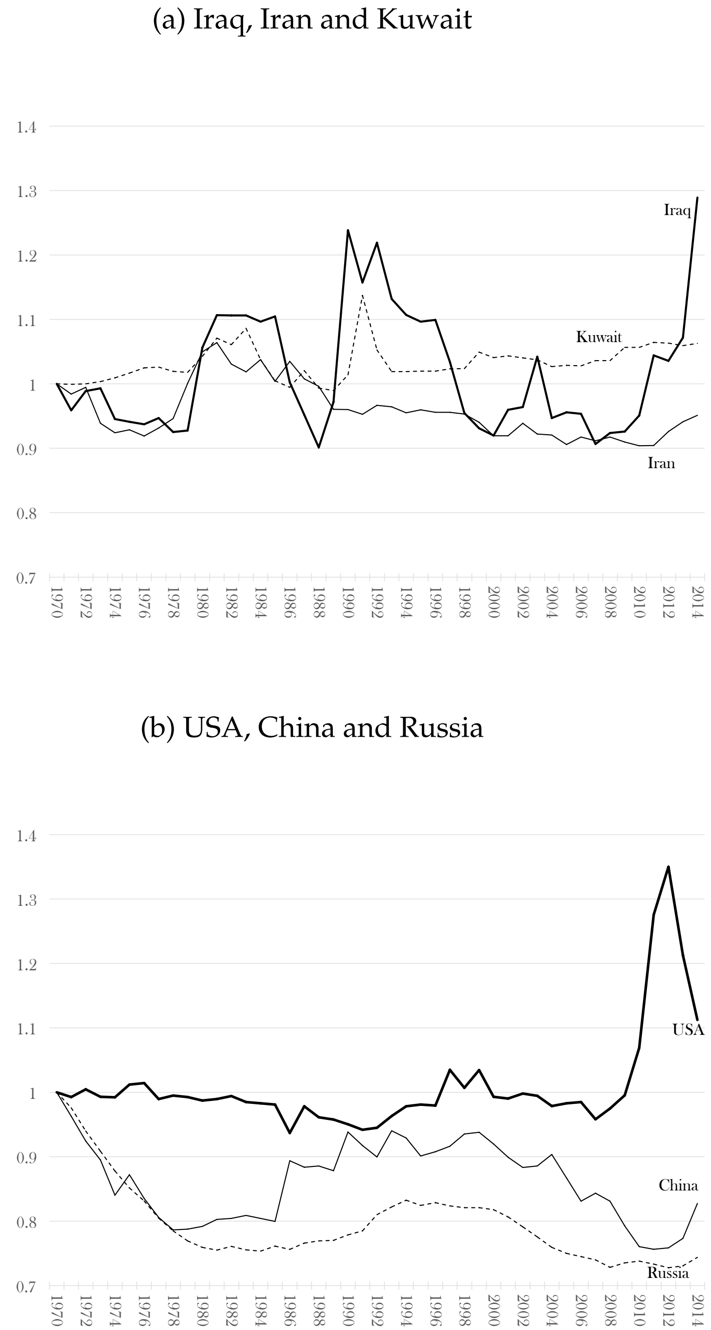
As might be expected, the market share of the Gulf countries increases significantly, from 25.8 percent to 74.4 percent. Saudi Arabia increases its share by 28.1 percentage points, and Kuwait increases by 12.5 percentage points. This mir-

Figure 6. : Decomposing Static Distortions



Notes: Static distortions for each year are presented in 2014 dollars (left vertical axis), with the total height of each bar representing the difference between the actual cost of production and the optimal cost of production (the total distortion). Each bar is decomposed into the following distortions (from bottom to top): Within-country (non-OPEC); Within-country (OPEC); Across country (Within non-OPEC); Across-country (within OPEC, in grey); Between-OPEC and non-OPEC (in black). Definitions of distortions decompositions mirror those in Table 5, although only applying to the individual year of interest. The oil price is shown using the black line dollars corresponding to the right vertical axis.

Figure 7. : Within-Country deviations from cost-minimizing production: Selected countries (1970-2014)



(a) Iraq, Iran and Kuwait

(b) USA, China and Russia

Notes: In panel (a), Iraq is heavy bold, Iran is light bold line and Kuwait is the dashed line. In panel (b), USA is heavy bold, China is light bold line and Russia is the dashed line. The ratio of total observed cost and cost-minimizing (optimal) cost, taking the actual production up that year as an initial condition (as per static distortion calculations) and holding country production fixed, is computed for each year (i.e., annual within-country deviations from cost minimization). This ratio is indexed to the level in 1970. The indexed ratio (vertical axis) is graphed by year (horizontal axis). The ratio in 1970 (i.e. the base year) for Iraq, Iran, Kuwait, USA, China, and Russia is 1.385, 1.563, 1.037, 1.401, 1.956 and 1.742 respectively.

Table 6—: Static counterfactual for 2014: Top 20 producers

Country	Actual output share	Counterfactual output share	Δ Share
Persian Gulf OPEC	0.258	0.744	0.486
Iran	0.057	0.091	0.034
Iraq	0.029	0.069	0.040
Kuwait	0.030	0.155	0.125
Qatar	0.009	0.015	0.006
Saudi Arabia	0.133	0.414	0.281
United Arab Emirates	0.031	0.075	0.044
Other OPEC	0.135	0.044	-0.091
Algeria	0.021	0.015	-0.006
Indonesia	0.020	0.002	-0.018
Libya	0.025	0.012	-0.013
Nigeria	0.028	0.006	-0.022
Venezuela	0.041	0.009	-0.032
Non-OPEC	0.607	0.212	-0.395
Brazil	0.014	0.001	-0.013
Canada	0.023	0.006	-0.017
China	0.045	0.002	-0.043
Kazakhstan	0.010	0.000	-0.01
Mexico	0.023	0.013	-0.01
Norway	0.027	0.009	-0.018
Russia	0.144	0.047	-0.097
United Kingdom	0.022	0.001	-0.021
United States	0.132	0.013	-0.119
Rest of the World	0.136	0.044	-0.092

Notes: Reported results are for the top 20 producers between 1970 and 2014. Initial conditions are the state of the global market at the end of 2013. Application of the sorting algorithm gives counterfactual production for 2014. In every other respect, the baseline specification is used: a field extraction rate of 10 percent of reserves is imposed in the counterfactual; the p50 measures of reserves are used where needed; and a demand growth rate of 1.3 percent per year after 2014 is assumed.

rors 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.

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.[44] This gives a basis for assessing the implications of different assumptions around the physical extraction limits and the implications of cost increases due to increases in production intensity.[45] All measures of

[44] 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.

[45] The extraction limit results in a hockey-stick cost function with respect to production intensity within a year. For instance, in the 2% extraction limit case, as the production intensity goes beyond $\max\{x_f, 2\%\}$ the cost of additional production intensity becomes infinite. This hockey-stick feature is extreme, but allows an opinion to be formed as to the implications of alternate assumptions. This is further explored in section 3.3 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 VI.A.

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 VI.B, together with a more detailed discussion of the impact of start-up costs.

VI. Modeling alternatives

In this section, we discuss additional factors that may impact the results reported above. In Section VI.A 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

the Online Appendix where we vary the extraction limit by type of oil field (offshore, onshore, shale (tight) oil). The results are robust to these perturbations.

Table 7—: Dynamic counterfactual results, alternate specifications

	Specification							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Actual (A)	2499	2467	2507	2499	4484	2474	1465	2500
Counterfactual (C)	1756	1804	1713	1757	2839	1703	1021	1797
Total distortion (A - C)	744	664	793	742	1645	771	444	703
Proportion: (A - C)/A	0.298	0.269	0.316	0.297	0.367	0.312	0.303	0.281
Distortion due to OPEC								
Upper bound (X+Y)	163	148	150	161	747	196	179	188
Lower bound (Y only)	105	89	95	104	225	99	120	125
Proportion: (X+Y)/(A-C)	0.219	0.224	0.189	0.218	0.454	0.255	0.404	0.268
Proportion: Y/(A-C)	0.142	0.134	0.120	0.140	0.137	0.128	0.271	0.178

Notes: Select results for Table 5 are reported for different model and parameter specifications. The units are billions of 2014 dollars or proportions. Results correspond to the 1970-2100 (exhaustion) timespan. Specifications are: (1) the baseline specification; (2) baseline, but with the limit on the proportion of reserves extractable in a given year changed to $\max\{x_f, 2\%\}$; (3) baseline, but with a no limit on the proportion of reserves extractable in a given year; (4) baseline, but using a P90 reserve measure; (5) baseline, adding the distortionary tax items in Table A.1 to costs; (6) has behavior computed with the competitive solution with wedge inclusive costs, but the costs of a particular allocation are evaluated with respect to economic costs only; (7) baseline, but restricting the sample to include only fields in active production in 1970; (8) baseline, but constraining fields to be usable in and after the first year of observed production, rather than discovery.

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 VI.B), curvature in fields' marginal cost curves (Section VI.C) and heterogeneity in the discount factors of market actors (Section VI.D).

A. *The marginal impact of market power*

The analysis proceeds in two steps. The first, in Section VI.A, 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 VI.A, 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.

Taxes and Royalties: Allowing for interactions with other observed distortions

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.

Wedges: Allowing for interactions with unobserved distortions

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.[46]

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

[46] This approach, while not addressing the total level of misallocation, provides an alternate route to insulating the level of misallocation due to market power from contamination from measurement error and expectation error.

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.[47] 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, the time varying component of costs, 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.

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

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

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.[48] 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.[49] 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 distortions to have a greater influence on the counterfactual paths.

[48] 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.

[49] 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 $c_f \mu_{st}$

Table 8—: Dynamic counterfactual results, conditional on inferred wedges

	Specification			
	(1)	(2)	(3)	(4)
Actual (A)	2498	2492	2670	2596
Counterfactual (C_1)	1757	1757	-	-
Counterfactual (C_2)	-	-	1825	2452
Total distortion (A - C_1)	741	735	-	-
Second-Best distortion (A - C_2)	-	-	845	144
Distortion due to OPEC	-	-	845	144
Upper bound	174	158	-	-
Lower bound	117	100	-	-

Notes: The units are billions of 2014 dollars. Results correspond to the 1970-2100 (exhaustion) timespan. Specifications are: (1) the baseline specification; (2) baseline, but with no limit on the proportion of reserves extractable in a given year; (3) average over 20 iterations of a counterfactual computed conditional on inferred wedges for non-OPEC field-years, and wedges for OPEC reserves sampled (with replacement) i.i.d from all non-OPEC inferred wedges; (4) average over 20 iterations of a counterfactual computed conditional on inferred wedges for non-OPEC field-years, and wedges for OPEC reserves sampled (with replacement) i.i.d from all non-OPEC inferred wedges inferred for fields with c_f within $5 of the OPEC field's c_f.) The actual costs 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, all 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 (in each specification) are evaluated using cost as measured by $c_f \mu_{st}$.

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 inframarginal 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.

B. 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.

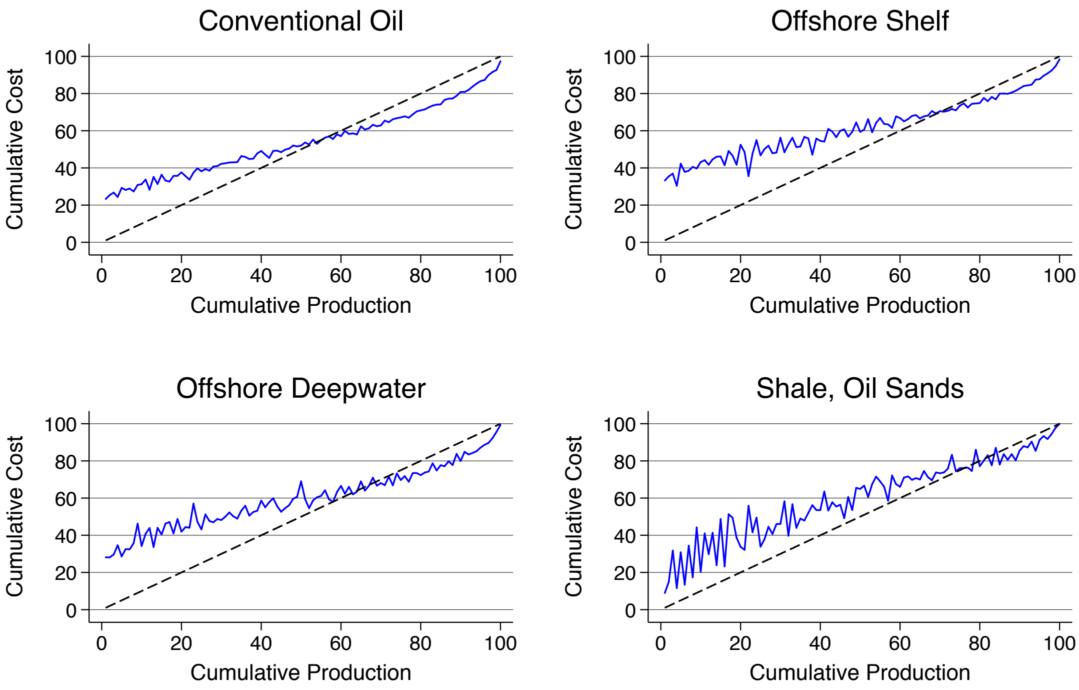
Figure 8 investigates the size of these fixed costs for fields that started production after 1970, by computing, within a field, cumulative costs and cumulative production. The magnitude of these start-ups costs can be identified from the proportion of expenditures incurred prior to the start of production. We report aggregates, 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).[50]

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

[50] In Figure 8, 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 8 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 VI.C, this is due to rising industry-wide input costs at the end of our sample period.

Figure 8. : Costs over the Field Lifecycle



Notes: Cumulative Production is measured as cumulative production for a field in year *t* divided by the total production observed over a fields lifespan. Cumulative Costs are defined likewise for costs. Conventional Oil is 72% of production; Offshore Shelf is 21%; Offshore Deepwater is 6%; Shale is 1%. Only fields that start producing after 1970 are used for this figure.

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.

C. *Marginal cost curvature*

Recall that we use the following specification for costs (in equation 13):

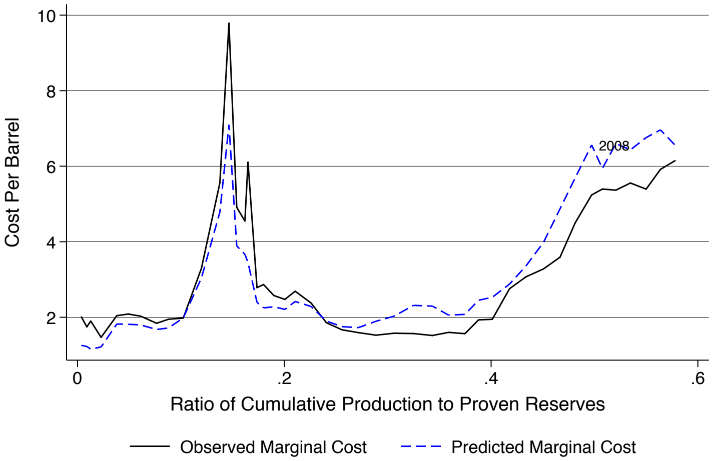
$$c_{ft} = c_f \mu_{st} \exp\left(\epsilon_{ft}\right).$$

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 9, 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 13.

Figure 9. : Observed and Predicted Marginal Cost
Ghawar Uthmaniyah (SA)



Notes: Observed and predicted marginal cost, using the cost specification in equation 13, is plotted against cumulative production. The vertical line indicates the proven reserves, and we insert the production year 2008, the year with the highest oil price in the sample period 1970-2014.

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, Figure 10 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 10 shows that ϵ_{ft} is unrelated to the the proportion of recoverable reserves that have been extracted. In particular, there is no systematic 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 9 and 10 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.

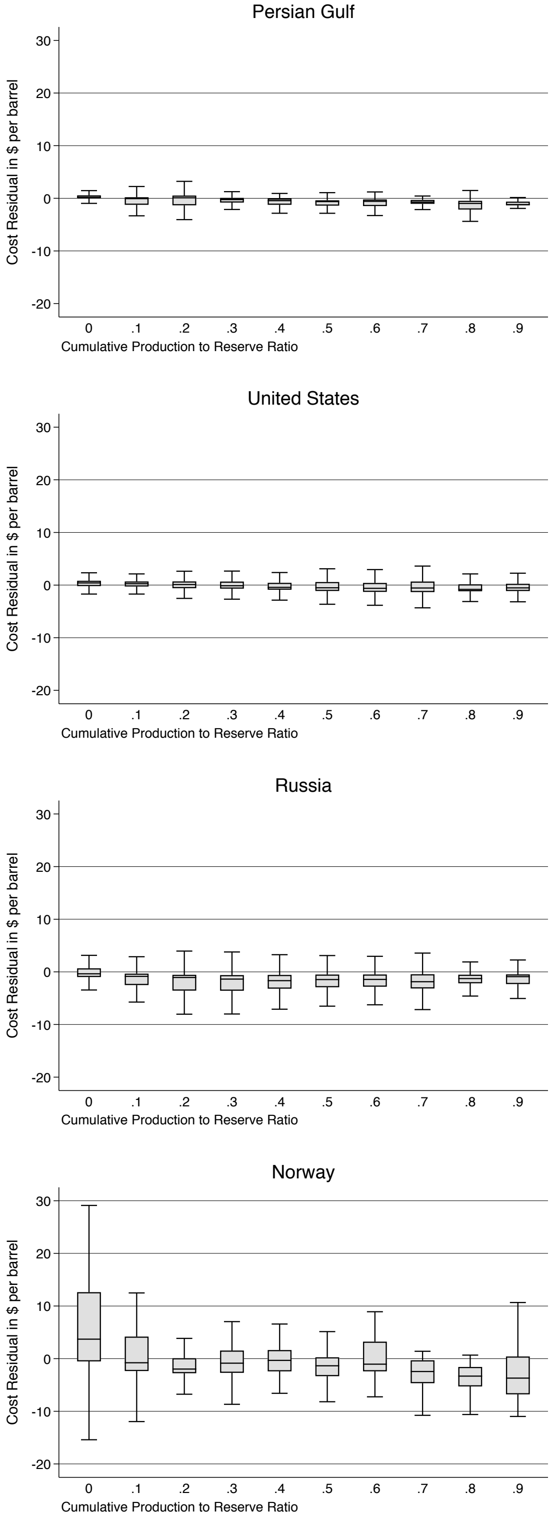
D. Discount factors

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 VI.A 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.

Figure 10. : Deviation of marginal cost specification and output



Notes: The residuals from the cost specification in equation 13 are plotted against quantiles of the ratio of cumulative output-to-reserves, and weighted by the production of a field in total (country-level) production. The whiskers present the upper and lower adjacent values, while the box shows the 25th, 50th, and 75th percentile of the distribution respectively. We consider only fields with reserves that are equal to or higher than total recorded production over the observed life cycle of a field.

VII. Conclusion

This paper demonstrates an alternative framework for understanding and measuring the extent of misallocation, and applies this approach to the global oil industry. We focus on measuring productive inefficiency, which while closely linked to standard metrics in the existing literature on misallocation, allows new data to be brought to to bear on misallocation in a way that has a clear welfare interpretation in partial equilibrium. We also show how to extend this approach to decompose the extent of misallocation into different components, in ways that respect the interaction between different distortions as implied by the theory of the second-best.

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.

Applying this methodology to the global oil industry indicates substantial productive inefficiency. An economically significant proportion of this is due to market power – between 105 and 163 billion 2014 USD, or 14 and 21 percent of the total inefficiency, depending on the cartel model applied (and, as noted elsewhere, conditional on any confounding measurement or expectation error). 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), famously, estimated the welfare losses due to monopoly for the entire US economy in 1921 at two billion dollars in today's currency.

An observation, arising from this paper and suggestive of productive future research, is that monopoly may have welfare effects that are several orders of magnitude larger than that implied by the Harberger style of analysis.

REFERENCES

Anderson, Soren T, Ryan Kellogg, and Stephen W Salant, "Hotelling under pressure," *Journal of Political Economy*, 2018, *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.

Banerjee, Abhijit V. and Esther Duflo, "Growth Theory through the Lens of Development Economics," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, 1 ed., Vol. 1, Part A, Elsevier, 2005, chapter 07, pp. 473–552.

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.

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.

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

Cicala, Steve, "Imperfect Markets versus Imperfect Regulation in U.S. Electricity Generation," Technical Report, University of Chicago 2017.

Collard-Wexler, Allan 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.

David, Joel M., Hugo A. Hopenhayn, and Venky Venkateswaran, "Information, Misallocation, and Aggregate Productivity," *The Quarterly Journal of Economics*, 2016, *131* (2), 943–1005.

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

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

De Mel, Suresh, David McKenzie, and Christopher Woodruff, "Returns to capital in microenterprises: evidence from a field experiment," *The Quarterly Journal of Economics*, 2008, *123* (4), 1329–1372.

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

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.

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.

Gopinath, Gita, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez., "Capital Allocation and Productivity in South Europe.," *The Quarterly Journal of Economics*, 2017, *132* (4), 1915–1967.

Haltiwanger, John, Robert Kulick, and Chad Syverson, "Misallocation Measures: The Distortion That Ate the Residual," Technical Report 24199, National Bureau of Economic Research 2018.

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

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

Herfindahl, Orris C, "Depletion and Economic Theory," in Mason Gaffney, ed., *Extractive resources and taxation*, Univ. of Wisconsin Press Madison, 1967, pp. 63–90.

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.

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.

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*, 1996, *64* (6), 1263–1297.

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

_ **and** _ , "Misallocation and productivity," *Review of Economic Dynamics*, 2013, *1* (16), 1–10.

_ **and** _ , "The Causes and Costs of Misallocation," *Journal of Economic Perspectives*, August 2017, *31* (3), 151–74.

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

Solow, Robert M and Frederic Y Wan, "Extraction costs in the theory of exhaustible resources," *The Bell Journal of Economics*, 1976, *7* (2), 359–370.

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, Elsevier, 1993, 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.

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

A. Definitions of cost components

Table A.1—: Definitions of cost components

Exploration Capital Expenditures:	Costs incurred to find and prove hydrocarbons: seismic, wildcat and appraisal wells, and general engineering costs.
Well Capital Expenditures	Capitalized costs related to well construction, including drilling costs, rig lease, well completion, well stimulation, steel costs and materials.
Facility Capital Expenditures	Costs to develop, install, maintain and modify surface installations and infrastructure.
Abandonment Cost	Costs for decommissioning a field.
Production Operating Expenditures	Operational expenses directly related to the production activity. The category includes materials, tools, maintenance, equipment lease costs and operation-related salaries. Depreciation and other non-cash items are not included.
Transportation Operating Expenditures	Represents the costs of bringing the oil and gas from the production site/processing plant to the pricing point (only upstream transportation). The category includes transport fees and blending costs.
SGA Operating Expenditures	Operating expenses not directly associated with field operations. The category includes administrative staff costs, office leases, related benefits (stocks and stock option plans) and professional expenses (legal, consulting, insurance). Only exploration and production-related SG&A are included.
Taxes Operating Expenditure	Local US taxes that are directly related to production. The category includes ad valorem taxes (county-based) and severance taxes (state-based).
Royalties	The sum of all gross taxes, including royalties and oil and export duties.
Government Profit Oil	The production-sharing agreement equivalent to petroleum taxes, but paid in kind (that is, the government contracts with a company to develop and operate the field, but retains rights to a proportion of the production). Government Profit Oil reduces the company's entitlement production and is treated as a royalty effect in company reports.

Source: Rystad U-Cube External Use Documentation.

B. Constructing implicit wedges

To construct the implicit wedges on non-OPEC fields that account for the deviations in production decisions away from those consistent with perfect competition, we use the following procedure.

A wide range of implicit wedges could rationalize the observed production path for non-OPEC fields. To settle on a specific wedge vector, we assume that implicit wedges are constant over time and are attached to specific barrels. That is, different barrels from the same field may have different wedges associated with them. This allows for production from a field to be spread over many years, whereas in perfect competition, it may be compressed. Given this, we search for a vector of wedges that minimizes the absolute value of wedge payments. That is, each element of the vector is a wedge associated with a specific barrel of crude. We search for a vector that minimizes the sum of the absolute value of the elements of this vector.[51]

[51] The uniqueness of this wedge vector is not guaranteed. Where the solution is not unique over some

The solution to this minimization problem is found by using a two-step algorithm. In the first step, wedges are added to the costs of production (c_f's) for each barrel until the sequence of production predicted by the sorting algorithm matches that in the observed data for non-OPEC barrels. In the second step, these wedges from the first step are reduced, at times resulting in negative wedges, until a minimum is found. This second step finds a minimizing wedge vector by leveraging the fact that the first step creates a wedge vector that puts wedges on all barrels that are weakly too high. Thus, the direction in which to search for a minimizing vector is known.

The logic of the algorithm implementation is best illustrated via the example in Table B.1, which shows four barrels of oil, labeled A, B, C and D, which are extracted in that order. The observed unit cost of each barrel is shown in column 2. In columns 3 and 4, the first-stage calculations are shown, in which wedges (column 4) are added, as necessary, to make the adjusted costs (column 3) have the same ordering as production. Note that the sum of the absolute values of these (stage 2) wedges is $8 + 3 = 11$. Columns 5 and 6 show the stage 2 calculations. The algorithm starts with the final unit of production (D) and looks to see if the wedge on it can be reduced without violating the ordering (imposing a minimum increment of 1). In this case, it cannot, as the adjusted cost of D is 15 (column 2), while C has an adjusted cost of 14. Next, the set $\{D, C\}$ is considered: can the wedges on both these barrels be reduced by a common amount without violating ordering? Again, the answer is no, given the minimum increment of 1. Now the set $\{D, C, B\}$ is considered. Here, the stage 1 implied wedge can be reduced by up to 5 on each of $\{D, C, B\}$ without violating ordering. However, the sum of the absolute values of wedge payments in this interval is minimized when the reduction is equal to 3 or, equivalently, when the median wedge in this interval is equal to zero.[52] Note that the sum of the absolute values of these (stage 1) wedges is $3 + 5 = 8$. The algorithm implemented to find the implied wedges for all barrels in the dataset is a large-scale version of that used to solve this example.

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 wedges 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 wedges to be derived deterministically since, with this adjustment, the ordering of the production sequence in perfect competition is independent of μ realizations. More specifically, the cost specification that is used is

$$\tilde{c}_{\tilde{f}t} = \left(c_f + \tau_{\tilde{f}}\right) \mu_t \exp\left(\varepsilon_{ft}\right), \tag{1}$$

interval of barrels, we choose the solution that sets the wedge on the median barrel in that interval equal to zero.

[52] Non-uniqueness of a minimizing wedge vector arises when there are an even number of elements in a set that can be adjusted like that demonstrated here. In that instance, we choose the vector created when the median is set equal to zero.

Table B.1—: Small-scale example of calculating implied wedges

Barrel (1)	c_f (2)	Adj. cost (stage 1) (3)	Implied wedge (stage 1) (4)	Adj. cost (stage 2) (5)	Final implied wedge (stage 2) (6)
A	7	7	0	7	0
B	13	13	0	10	-3
C	6	14	8	11	5
D	12	15	3	12	0

where $\tau_{\tilde{f}}$ is the wedge to be recovered. We also add the notation $\tilde{f}$, indicating the set of barrels extracted from a specific field-year combination of field f in the data. This allows fields to experience wedges that split their production over multiple years, in a parsimonious way.

The approach above results in a vector of implicit wedges that apply to barrels that are not subject to the market power distortions resulting from OPEC. In the absence of OPEC exercising market power, it is likely that similar distortions would still impact OPEC production. To this end, we sample from the distribution of implicit wedges that distort non-OPEC production to construct a counterfactual production path for OPEC. Together with the implicit wedges added to the costs of production for non-OPEC countries, this allows us to assess the marginal distortion imposed by OPEC's exercise of market power and compare it to the infra-marginal measure discussed in the rest of the paper.

The sampling procedure is conducted as follows. Wedges are sampled with replacement from the set of all inferred wedges. Each wedge is given equal weight in the sampling. Each wedge $\tau_{\tilde{f}}$ is linked to a quantity $q_{\tilde{f}}$ of barrels extracted, in the data, in the same year from the same field. A sampled wedge is applied to an OPEC field reserves level, in the amount of the $q_{\tilde{f}}$ associated with it. Another wedge is then sampled and applied to that remaining OPEC field reserves, until all OPEC reserves are covered by a sampled wedge. Note that, through applying the wedge to reserves according to its associated $q_{\tilde{f}}$, this implicitly means that wedges associated with larger fields get larger weighting. This creates a set of counterfactual wedges for OPEC fields.

In running counterfactuals, the costs determining paths are formed using these sampled wedges for OPEC, and the (field-specific) inferred wedges, $\tau_{\tilde{f}}$, for each field outside OPEC. Resource costs are computed using the standard resource cost measure used elsewhere in the paper (not including any taxes or inferred wedges).