Online Appendix to
(Mis)Allocation and Market Power, and Global Oil Extraction
Not For Publication

John Asker and Allan Collard-Wexler and Jan De Loecker*

October 22, 2018

Contents

*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.

1   Data: Collection, sources and measurement

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 was obtained from from Rystad Energy (Rystad hereafter), an energy consultancy based in Norway. The specific data product is called the U-Cube, or upstream, database.[1] The data cover the operations of each oil field around the world, and is documented in Rystad Energy (2015). For each field, the data includes production and different operating and capital costs, as well as characteristics such as geology and reserves. Various parts of Rystad's data product has been used in other economic studies, including Bornstein et al. (2017) and Bartik et al. (2016).

1.1   Data collection and sources

As the Rystad data covers the global oil industry. As a result it is collected from a variety of original and secondary sources. These sources range from high quality government reports in countries such as Norway and the United States, through company reports for large private companies, to interviews with companies in the shipping and oil services industries. Collating and reconciling these sources is a difficult process, particularly in politically unstable areas such as Syria from 2012 onward. Thus in some countries, while aggregate production at the country level may be observable from things like tanker movements, Rystad uses engineering models to approximate micro production and costs. These compromises are unavoidable when assembling consistent micro data for a global industry.

Compiling data on the oil industry involves confronting issues common to evaluating the performance of any global industry. Oil extraction is completed by many different government and non-government entities, across many different countries, in a wide variety of geological and environmental settings (e.g. on-shore and off-shore extraction). In addition, sales of crude oil are made between many different buyers and sellers in a largely decentralized market. As a consequence, there is no centralized data collection protocol that leads to a unified dataset of the sort commonly used in industry studies that focus on a specific product class in a specific geographic location, or studies based on census datasets collected in some countries by statistical agencies.

Nonetheless, many high quality data sources exist describing aspects of the oil industry. These data sources include government reports, company reports, regulatory filings (financial, environmental and otherwise), records of royalty payments to governments, press releases, analyst reports, tanker movements and world of mouth reports from on-the-ground operators. A central challenge in building a global data structure is to collate and cross-check these data sources. A second challenge is to handle data quality that varies across countries (ranging from countries that have extreme levels of transparency, to countries that view oil production, revenue and reserves as matters of national security and shroud their activity in considerable secrecy). This second challenge requires imputing missing data, reconciling contradictory data sources and cross-checking questionable data with multiple sources. These challenges are comparable to those confronted by national statistical agencies when compiling measures of aggregate economic activity, like GDP figures.

---

[1]See https://www.rystadenergy.com/Products/EnP-Solutions/UCube/Default accessed March 6 2018.

1.1.1   Collection

In this section we provide a short description of how this dataset is assembled, as detailed in Rystad Energy (2015). The U-Cube is a bottom-up database: one that starts from individual oil fields and aggregates them up to obtain country and global production. The data from Rystad concerns a large number of fields from 1970 to 2014. Since this dataset is used to forecast future oil prices, it includes currently producing fields, as well as fields that could begin producing in the future.

Table 1 shows the data cleaning steps we perform, along with the total number of fields after each step. The data has 66,920 unique fields, but only 21,233 produce at any point in time. The non-producing fields are used by Rystad for forecasting purposes — they are estimates of the production of a field that has not yet started production. Of the 21,233 producing fields, about 1,408 are gas-only fields which we drop from our analysis.[2] Linking the remaining fields to reserves leaves 13,248 fields, which account for 92 percent of (non-gas) production.

Table 1: Data Cleaning and Sample Frame

	Fields	Total Production in Trillions of BOE (incl. Gas)
All Fields	66,920	1.81
Drop non-production fields	21,233	1.81
Drop non-oil fields	19,803	1.16
Drop missing reserves	13,248	1.07
Drop fields with missing discovery year	11,455	1.07

Note: 13,248 fields used in the paper except in section 5 and 6 — the structural model, where 11,455 fields are used. Total Production is over the 1970 to 2014 period.

1.1.2   Data Sources

There are two types of data that are used to construct field-level data. Geological and lifecycle data, and economic data. Rystad keeps a database of the type and geology of the field. Some of this information is about the physical aspects of the field, such as whether it is an oil sands, shale, or an offshore play of a certain depth, and details on the hydrocarbons produced, e.g. gas or oil, density, and sulfur content. This information is complemented with other geological details. As well, Rystad keeps track of the discovery date and depletion of these fields.

Much of the field data is obtained from government and company sources. For instance, in the United States, governments and private mineral rights owners collect royalties from firms, and as a consequence collect detailed field-level production data. In addition, some companies publish information on production levels and reserves in different fields for the benefit of investors.

The second type of data are economic. These data are mainly sourced from three different places: company reports, government agencies, and oilfield service firm interviews and publications. Where details are unavailable, the economic and production data is extrapolated from similar oil fields, and cross-validated against aggregated statistics.[3]

---

[2]A gas-only field is a field that does not produce any crude oil, but only natural gas, including liquified or flared.

[3]In construction field level data that aggregates to the global oil supply, Rystad needs to confront similar issues

1.2   Measurement of main variables

1.2.1   Production

For the analysis, we convert all units of fuel to crude barrel of oil equivalent (BOE) – different hydrocarbon mixes, e.g. condensate or gas have different energy contents by volume. We also drop production of gas products, namely gas, LNG and NGL, and flared or injected gas. While gas and oil are often recovered jointly in one oil field, we observe expenditures of the various categories broken down by oil and gas type.

   Some of these fields are fairly close to production units, such as individual offshore oil rigs, while other fields are quite large, such as the Ghawar field, with hundreds of rigs. Indeed, in 2014, the tenth percentile oil field produces 8,000 barrels per year, while the 90th percentile produces 1.9 million barrels per year.[4] Figure 1 shows the time series of total production from 1970 to 2014, as well as the total number of producing fields.
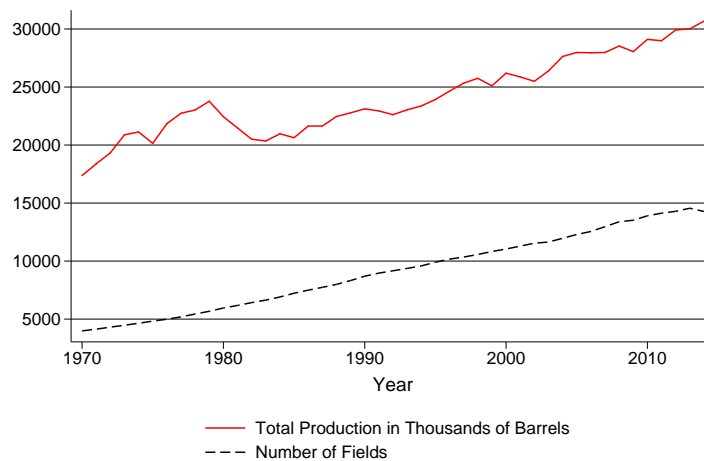

Figure 1: Total Production and Number of Oil Producing Assets

1.2.2   Measurement of reserves

Central to much of the discussion in the paper is the notion of reserves. Reserves are the un-extracted quantity of oil remaining in a field that is recoverable at a given price. Reserves are distinct from the unrecoverable 25% to 75% of the oil that will remain in the field even after production has concluded, depending on geology. Reserves are difficult to accurately predict; the best measure of reserves is simply the total quantity extracted after production has been completed.

---

as those faced by statistical agencies that construct national accounts. See White et al. (Forthcoming) and White and Rotemberg (2017) for a discussion of the role of imputation in constructing manufacturing microdata, both in the U.S. and in other countries.

   [4]For some countries, industry measures of production differ from official publications, with the industry believing official figures to be wrong. In these instances, Rystad provides an industry measure. See **?** for a contemporary discussion of this type of measurement error in Saudi Arabia's official figures.

Unfortunately, most fields have not been fully exploited. Therefore we need to use industry estimates of reserves. The oil industry reports reserves at different levels of extraction probability. There are three levels. P90 (or P1) is the quantity able to recovered with a 90% probability given current technical and economic conditions. The P90 reserve is the asset value able to be reported on company balance sheets under U.S. GAAP. This definition means that reserves fluctuate with the prevailing 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% probability. Finally P10 or (P1 + P2 + P3) are total reserves recoverable with a 10% chance. The level of P90, P50 and P10 can vary significantly within a field. For instance in the North Ward Estes field discussed below, P90, P50 and P10 in 1975 were estimated at 26.6, 56.4 and 66.4 million barrels. In this paper, unless stated otherwise, the reserve number used for a field is P50. The precision with which reserves are measured varies by a field's country and stage of production. Untapped deposits have less precise reserve estimates, since no production or well pressure data are available. Once a field starts producing, reserves become easier to estimate, particularly as pressure starts to change, as the pressure gradient of a field as the resource is depleted is relatively well understood, conditional on geology. A further confounding factor is the oil reserves are strategic assets, with most industry sources commenting that various countries will inflate reserve figures for political reasons. This is particularly relevant for OPEC countries due to the way OPEC has computed quotas at various times in its history. This paper takes the Rystad reserve data as the best estimate available.

Table 2 shows total reserves in the world in 1970, 1990, 2000, and 2014. They stand at 439 trillion barrels in 2014, if one considers P50 and a forecasted price of oil of $70 a barrel.

Table 2: Reserves, Probability of Recovery and Forecasted Price

Year	Reserve in Trillions of Barrels				
	P10 $70 barrel	P50 $70 barrel	P90 $70 barrel	P10 $100 barrel	P10 $130 barrel
1970	932	724	354	1138	1172
1990	914	699	385	1206	1264
2000	797	609	342	1121	1194
2014	572	439	248	1020	1126

In the paper, in descriptive discussions, P50 values at an oil price of $70 a barrel are used to report reserves. When simulating counterfactual production paths, 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.

1.2.3 Further details on cost measures

In what follows we pull information from the data documentation and other internal documents presenting additional detail on how Rystad handles the various components of capital expenditures (this expands on Appendix A in the paper) .[5]

---

[5]The documents we draw from are *Rystad UCUBE Technical Handbook* p 40-41 and p46 onward, and *UCUBE Technical Presentation 2015*, dated 4 August 2015, p 30-31.

- **Facility Capex**: "includes costs to develop, install, maintain and modify surface installations and infrastructure. As reported by operators, field partners or officials, or modeled." Facility cost consists of (discussion arises in the context of a discussion of forecasting the Gundrun Field):

    – Development Costs: "Costs associated with the construction and implementation of the facility required for the processing and production of the field. This cost depends on location, facility type, and resources. The cost will be between 2 to 30 \$/bbl."

    – Modification, Maintenance & Operation Costs: "Expenditures related to maintenance and improvements required to keep the facility operational. The value is calculated as a share of the development costs, and reaches its peak after decline starts."

- **Well Capex**: "Well capex is capitalized costs related to well construction, including drilling costs, rig lease, well completion, well stimulation, steel costs and materials.", and discussing the construction of this when estimated: "The total well capex for an asset is estimated by looking at the field type and estimated drilling cost per boe of resources. The drilling costs are determined by several dimensions like water depth, reservoir depth, recovery, region, recovery method and facility type. The total cost is then distributed among life time of a projects based on predefined profiles. These profiles are determined by the facility type. The figure below shows how the well capex is estimated for the Gudrun field."

1.2.4   Data quality

When sourcing the data, we investigated the various data vendors in this industry and our judgement was that Rystad was the best for our purpose, evaluated on transparency, coverage and quality. That said, the other two vendors, IHS and Wood Gundy, are also widely used and the three firms are broadly comparable. The data we acquired does not include a full set of the internal metrics that Rystad uses to log the raw sources and manipulations/reconciliations of the various data sources data before it is sold to industry participants. That said, quality can be reviewed along several dimensions. These are discussed below.

The clearest insight that we have into data collection is with respect to production. We have, for 2013, global production classified into bins. These quality bins are (with complete definitions pulled from documentation provided for external use):[6]

- High : Data point updated directly based on information reported by the company (annual reports, quarterly reports) and by official institutions (government reported data)

- Medium : Additional assessments and/or estimations done to the data reported by the company or official institutions.

- Low : Estimated data point based on insight on the company or geological area.

- Modeled : the data point is the result of the algorithm.

---

[6]No further explanation is provided in the external documentation.

The proportion of 2013 global production data (counted in barrels) falling into each of these bins is High - 35%, Medium - 10%, Low - 18%, and Modelled - 37%. By way of comparison, White et al (2018) report that in the 2002 US Census of Manufacturers 79% of observations had imputed data for at least one variable used to compute TFP. In the same Census, 42% of observations of the cost of materials needed to be imputed.

For cost variables, we are not in a position to provide this level of detail. In the Rystad external documentation, they do, by country, describe the data quality of the 'economic variables' – costs, revenues, etc – by country, but do not provide a clear indication of what the classifications (High, Medium, Low and Very Low) mean or the time period to which the information applies. The issue of time period is relevant here as much of the commercial value of Rystad's product is in forecasting and performing diligence on asset acquisitions. To this end, the documentation is focused on that part of the Rystad platform that provides forecasts and current data – these are parts of the Rystad platform that we do not use (our data acquisition occurred in 2016 and we only used data up to 2014 to allow sources to report data with a delay). What is encouraging is that where estimation and forecasting occurs, the documentation is clear that it is done using historical data, which is contained in that part of the Rystad product that we utilize. Given these qualifiers, the documentation reports that the countries with both production and economic data classified as 'High' are the USA, Canada, Norway, Azerbaijan and Angola.

In line with the finding of White et al, we conjecture that modeling/interpolation would bias us toward lower levels of misallocation by tending to push numbers closer to the populations means than is actually the case. Nonetheless, to guide judgement as to the extent to which anything we report in the paper is impacted by imputed or modeled data, we re-ran the exercise on just Norway, Canada and the USA (the OECD countries that Rystad reports as having the best data). That is, we considered a scenario in which the world consisted of only these three countries. We then reproduced the first three rows of table 5, in which the baseline misallocation numbers are reported. The resulting numbers are in the table 3.

Note that for the USA-Canada-Norway sample the total distortion (misallocation) is 17% of the actual cost of extraction. This compares with 29% for the full sample. Given that the impact of OPEC is estimated to be 6.5% of the actual cost of sample, this suggests that non-OPEC related distortions in the full sample are of the same rough order of magnitude as in the USA-Canada-Norway sample.

2 Cross-validation of data against alternative sources

The data in our paper comes from Rystad, but there are other actors that also produce estimates of the cost of oil extraction. At the global level, there are three firms that produce estimates of oil field costs: Rystad, IHS, and Wood Gundy. There are also a number of published estimates of oil costs from sources such as the World Bank, the IMF, and various news agencies using underlying data from (one of) the three mentioned firms.

The World Bank produces estimates of the cost of oil extraction at the country level in its report, The World Bank (2011). Lin Lawell (2017) relies on these estimates to study the global oil market. Using the procedure in Lin Lawell (2017) we recover a country-wide measure of the cost of production by country from 2000 to 2014 from The World Bank (2011).[7] In the report,

---

[7]This report states that it sources oil information from "Petroleum, national sources Costs: IEA, World Bank,

Table 3: Dynamic counterfactual results
(NPV of costs in billions of 2014 dollars)

| | Timespan | |
	USA, Canada and Norway only 1970-2100	As reported in paper 1970-2100
Actual (A)	661	2499
Counterfactual (C)	545	1756
Total distortion (A - C)	116	744

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 world bank produces a measure of oil rents (π), and using the identity that rents are profits from oil, along with readably available data on oil production and global oil prices, the cost of oil can be recovered.[8]

Figure 2 plots the costs estimates from the World Bank's report against those from Rystad (computed as the total expenditures in a country divided by total production of oil in that country). There are 340 country-year pairs, and the filled red circles indicate countries in the Persian Gulf (Iran, Iraq, Kuwait, Saudi Arabia, and the United Arab Emirates). There is a high degree of agreement between these two measures, with a correlation of 0.84 between Rystad and the World Bank. There is also strong agreement that low cost fields are primarily located in the Persian Gulf.

We also searched for other data sources to assess the cost differences across countries that we document in the paper. We found several sources for country-level costs: Argaam, the IMF, Citibank, Goldman Sachs, and Reuters. Argaam, a Saudi Financial News website, used information from Rystad, our data provider.[9] The IMF collected its own data as well as from the International Energy Agency.[10] Citibank used data from Wood Gundy, as well as internal analysis.[11]

Goldman Sachs also constructed its own data on breakeven costs of oil fields in many dif-

---

national sources" in Table A.2. on page 153.

[8] We rely on the following relationship: $\pi = \frac{(P-c)Q}{GDP}$ to obtain a measure for cost of production at the country level: $c = P - \frac{\pi GDP}{Q}$.

[9] https://www.argaam.com/en/article/articledetail/id/404677, accessed February 12, 2018.

[10] https://knoema.com/vyronoe/cost-of-oil-production-by-country , accessed February 12, 2018.

[11] https://www.citievents.com/Metron/common/downloads/attachments/files/global%20oil%20and%20gas%20volatile%20times%20ahead.aspx?id=046399f7-2aa9-4d6b-8414-c43c4497c147&ext=.pdf&type=1, accessed February 12, 2018.
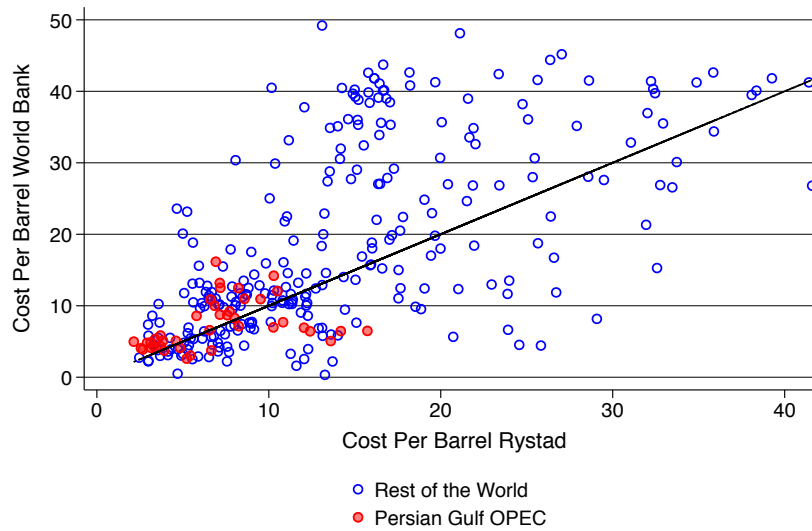
Figure 2: Alternative Data Sources (World Bank and Rystad)

ferent countries.[12] Reuters assembled estimates by surveying different consulting firms.[13]

One should be careful in interpreting these data, as many of these reports are not careful about reporting production weighted average, may only report costs for certain types of deposits (such as onshore or offshore only), or may strip capital expenditures from cost measures. Taking these caveats into account, Figure 3 plots the costs from these alternative data sources, against country level costs in 2014 from Rystad.

Again, these alternative data sources roughly correlate with our data from Rystad, with a correlation coefficient of 0.6. Indeed, there is much variation within estimates for a single country from a single source, say Citibank's estimates of Canadian oil fields at 17, 18, 20, and 22 dollars a barrel for conventional, heavy oil, offshore, and tight oil. So much of the mismatch between estimates could be due to the set of different fields being used. Moreover, the level of costs are similar between all of these sources.

3 Robustness of results to alternative specifications and measurement error

3.1 Robustness of counterfactuals to measurement error

In this section we test the extent to which our main results are sensitive to measurement error in production costs. To address this issue, we simulate the effect of measurement error on our competitive counterfactuals.

---

[12]https://www.bloomberg.com/news/articles/2017-07-12/these-numbers-show-the-huge-challenges-now-facing-opec-members, accessed February 12, 2018.

[13]https://www.reuters.com/article/oil-cost-factbox/factbox-oil-production-cost-estimates-by-country-idUSLS12407420090728, accessed February 12, 2018.
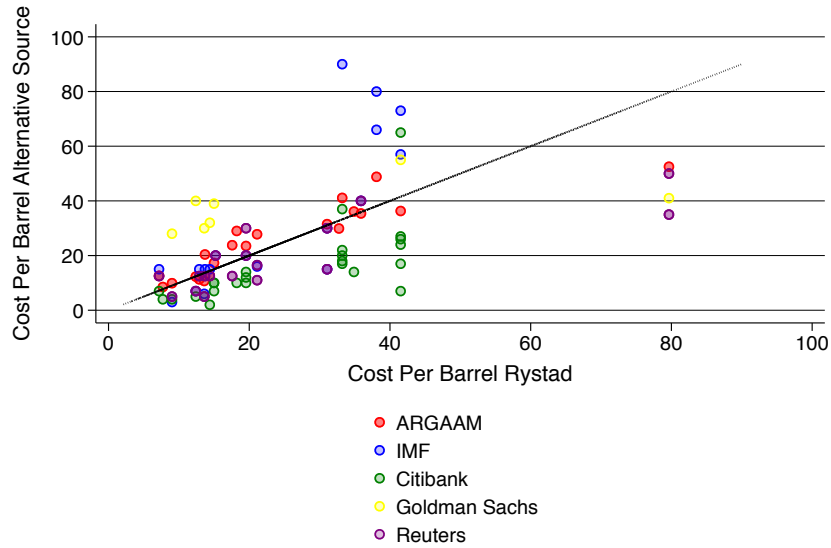
Figure 3: Alternative Data Sources

Specifically, for each field in the data, we add a measurement error ξ_{ft} to the cost of extraction c_{ft}, such that $\tilde{c}_{ft} = \xi_{ft} c_{ft}$. We then use these mismeasured $\tilde{c}_{ft}$ in the competitive counterfactual, repeating this exercise 50 times, to obtain the distribution of the welfare effects of misallocation, reporting the 10th and 90th percentiles of these distributions.

In the data, there are 11,455 fields over more than 30 years. So simply adding i.i.d. measurement error ξ_{ft} is likely to have a small aggregate effect, as it will simply wash out in a large enough sample. Moreover, much of the measurement issues, such as what are the costs of fields in Iraq, are likely to be shared among fields in the same geographic area. Instead of assuming ξ_{ft} is i.i.d., across all field-years, we assume that measurement error is common in a country across all years; i.e., $\xi_{ft} = \xi_c$ where c indexes countries. We consider two cases. First, we assume that $\xi_c \sim U[0.8, 1.2]$, so variation in measurement error lies between +/- 20 percent, and second, we consider an extreme scenario whereby $\xi_c \sim U[0.5, 1.5]$. Second, we repeat this analysis dropping countries inside the OECD.[14]

Table 4 presents the results of this exercise. Measurement error does introduce substantial variation in the estimate of the distortion, with the tenth percentile at 676 billion dollars and the ninetieth percentile at 782 billion dollars. However, the decomposition of the role of OPEC, say for the upper bound, is still between 17 and 24 percent of the total. Increasing the variance of measurement error to $U[0.5, 1.5]$ spreads out the OPEC effect from a tenth percentile of 15 percent, to a 90th percentile of 30 percent. Finally, focusing attention on only non-OECD countries also leads to similar upper bound effects of OPEC, from 17 to 33 percent across different percentiles and measurement error variances. Thus, our results on the effects of OPEC appear

---

[14]OECD countries are Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, and the United States

to be robust to a surprising amount of measurement error on the costs of oil fields.

3.2   Robustness of counterfactuals to alternative cost definitions

There is some discretion, due to the detailed breakdown of costs in the Rystad Data, of which costs to include in our measure c_{ft}. For instance, we exclude royalties and taxes, as they do not represent economic costs. As a robustness check, we repeat our competitive exercise with cost measures c_{ft} that exclude capital expenditures. The results are presented in Table 5. Capital Expenditures represent about 40 percent of costs, so excluding these lowers total expenditures on oil, in any counterfactual scenario, by approximately this amount. However, the decomposition of misallocation in terms of percentage welfare losses is virtually identical regardless of whether we included capital expenditures. For completeness, we also consider alternative cost measures, and the associated welfare analysis, eliminating each cost component one at a time (see notes in Table 5 below).[15] The role of OPEC in explaining total misallocation ranges between 19 and 21 percent depending on which exact components of cost are included; i.e., it is virtually unchanged from our baseline estimate.

3.3   Robustness of counterfactuals to alternative extraction limit specifications

Recall that for each field, extraction in each year is capped to be the maximum of either 10% of maximal reserves or the highest proportion of maximal reserves extracted in the data in a year for the field. That is, annual extraction is limited to, at most, $\max\{x_f, 10\%\}$, where x_f is the maximal proportion of reserves extracted, in any year, for that field. This keeps the level of extraction (and any related capex adjustment) in the counterfactual to reasonable levels (in specification 2 of table 7, the 10% limit is changed to 2% with little change in overall results). This effectively introduced a hockey-stick cost curve in a given year with respect to the intensity of production - once production intensity reaches a certain point, the cost of further increases to production intensity become infinite.

   To further stress test the results on this front we maintain the baseline and vary the 10% threshold for various types of fields. That is, we allow it to be 2% or 100% for each of the onshore, offshore and shale (tight) oil field types while holding everything else at 10%. The results are reported in Table 6.

---

[15]Omitting these cost components, like for instance capital expenditures, is not likely to be a feasible reallocation as more production will require, for example, more equipment and materials. This is less of an issue for those countries that actively hold reserves for immediate supply to the market, such as Saudi.

Table 4: Dynamic counterfactual results (Table 5 in paper): Measurement error

	All countries				Non-OECD only			
	$\xi_c \sim U[0.8, 1.2]$		$\xi_c \sim U[0.5, 1.5]$		$\xi_c \sim U[0.8, 1.2]$		$\xi_c \sim U[0.5, 1.5]$	
Percentile	10%	90%	10%	90%	10%	90%	10%	90%
Actual (A)	2469	2671	2291	2870	1607	1743	1536	1874
Counterfactual (C)	1763	1920	1593	1951	1179	1282	1074	1322
Total distortion (A - C)	676	782	692	950	405	468	433	566
Decomposition of total distortion in percentages								
Within country (non-OPEC)	39	42	34	40	29	33	24	30
Within country (OPEC)	20	23	16	23	33	39	27	40
Across country (within non-OPEC)	14	20	15	29	8	15	10	21
Across country (within OPEC) (X)	8	12	6	15	12	19	15	24
Between OPEC and non-OPEC (Y)	7	14	5	19	3	11	0	13
Production distortion due to OPEC market power								
Upper bound (X+Y)	17	24	15	30	17	28	18	33
Lower bound (Y only)	7	14	5	19	3	11	0	13

Table 5: Dynamic counterfactual results (Table 5 in paper) 1970-2100:
Alternative cost definitions

| | Base | Omitting cost measures: | | | |
		(1)	(2)	(3)	(4)
Actual (A)	2499	2248	1491	2525	2281
Counterfactual (C)	1756	1542	1073	1779	1563
Total distortion (A - C)	744	705	418	746	718
Decomposition of total distortion in percentages					
Within country (non-OPEC)	21	21	22	21	21
Within country (OPEC)	38	39	39	39	39
Across country (within non-OPEC)	8	8	7	7	7
Across country (within OPEC) (X)	19	18	18	18	18
Between OPEC and non-OPEC (Y)	14	15	13	13	15
Production distortion due to OPEC market power					
Upper bound (X+Y)	22	22	21	22	22
Lower bound (Y only)	13	15	13	14	15

Omitted costs are (1) SG&A, (2) All CapEx, (3) Abandonment Cost, (4) Transportation Operating Expenditures.

Table 6: Dynamic counterfactual results (Table 5 in paper) 1970-2100:
Alternative Extraction Rates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Extraction Rate Onshore	10%	10%	10%	10%	10%	100%	2%
Extraction Rate Shale	10%	100%	2%	10%	10%	10%	10%
Extraction Rate Offshore	10%	10%	10%	100%	2%	10%	10%
Actual (A)	2499	2499	2493	2501	2485	2504	2473
Counterfactual (C)	1756	1755	1749	1748	1811	1712	1812
Total Distortion (A-C)	744	745	744	752	674	793	661
Upper bound (X+Y)	163	163	162	152	138	151	148
Lower bound (Y only)	105	106	104	100	80	95	88
(X+Y)/(A-C)	0.22	0.22	0.22	0.20	0.20	0.19	0.22
Y/(A-C)	0.14	0.14	0.14	0.13	0.12	0.12	0.13

Notes: All definitions follow those used in tables 5 and 7 of the paper.

4   Further background on oil

4.1   Crude oil

Crude oil is the oil that is delivered to refineries for processing into various end-use hydrocarbon products.[16]

These products range from gasoline and other fuel oils (the majority of refinery production) to bitumen, lubricants, propane, naphthas and some waxes (such as parafin). Specifically we count as crude condensate all levels of API crude from extra heavy to light, crudes with low (sweet) and high (sour) sulphur content, as well as NGL. We exclude biodiesel, synthetic crudes, cold to liquid production and Bitumen-based production. The latter collectively account for about two percent of oil production in 2014.

Due to variation in local geology, the nature of crude in a deposit will vary. The two most important dimensions of heterogeneity in crude are density and sulphur content. Density is commonly measured in degrees API and sulphur content as a percentage by weight. Crudes are often referred to as heavy or light, and sweet or sour, referring to their density and sulphur content respectively. Crudes typically have a density between 10° and 50° API.[17] Most refineries are geared toward processing crudes in the 30° to 40° API range. Different refineries will have different sulphur tolerances as well. Hence, crudes that lie outside the 30° to 40° API range, or which have very low or high sulphur contents can trade at a discount (or a premium depending on market conditions) and may need to be mixed in with other crudes to meet refinery specifications.[18]

The heterogeneity in crudes leads to a series of measurement issues. The first is how to measure the quantity associated with a deposit in terms comparable across deposits. The data measures output in energy equivalent barrels, where the benchmark is one barrel of Brent crude oil. 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.[19]

4.2   Crude oil production

To monetize a discovered oil deposit, a firm needs to produce, or exploit it. Each deposit is located within a field, which has particular geological characteristics and may encompass multiple deposits. The manner in which a field is exploited will depend on its location and underlying geology. Every deposit will be exploited by drilling production wells. Beyond that, the most basic distinction is between onshore and offshore fields.

---

[16]The sources of this industry description are, where not otherwise noted, Downey (2009).

[17]Common benchmark crudes are Arabian Light, Brent and WTI which have densities of 34°, 38.3^o, and 39.6^o API, respectively. In measuring crude output we include conventional crudes (API < 50) and condensates, which are gaseous in the deposit but liquify after extraction. Condensates have API > 50.

[18]Sulphur needs to be removed from crude oil for the refining process to operate properly.

[19]Even on a per energy unit basis, hydrocarbons do trade at different prices due to differences in sulphur content and industry or refinery demand.

Onshore production. Production of an onshore deposit typically involves a range of stages or techniques (Downey, 2009). These are referred to as primary, secondary and tertiary recovery (methods). It is important to note that every deposit has geological features that dictate a different (and at times simultaneous) combination of primary, secondary and tertiary recovery methods over the course of a well's lifespan. Some forms of onshore deposit require substantially different production methods. For instance, tar sands, a significant proportion of which are found in Venezuela and Canada, are heavy crudes found close to the surface and mixed into loose rock or sand. These deposits are recovered through surface mining (rather than wells) and then require cleaning (to remove sand and soil) and pre-processing (to lift the API to refinery-appropriate levels). As a result tar sands can have extremely high marginal production costs.

Offshore production. Off-shore deposits have another set of production challenges.The extent of these challenges are determined by the water depth, the distance from land and the weather. Water depth creates both pressure at the well head (on the ocean floor) and temperature differentials between the subsurface deposit (hot) and the deep water (cold). Both lead to substantial engineering problems. For instance, to avoid frigid water changing the composition and viscosity of the crude, the bore is artificially heated. In deep water wells anti-freeze is also often added at the well head. Distance from land affects the way well workers can be housed, transported and rotated, and how the produced oil can be stored and transported to market. Maritime conditions and storms presents a series of additional challenges related to platform stability, production interruptions and safety.

Examples. The cost differences that arise from different operating environments are best illustrated via example. Consider the North Ward Estes field near Wickett in Ward County, Texas (an onshore conventional field) and Tract 174 in Grand Isle Block 43 located offshore of Lafourche Parish on the Louisiana coast (an offshore field in less than 100ft of water). The unit cost computed by dividing total expenditures (including operating and capital) by total production of the onshore field from 1970 onward was $7.57 per barrel. In contrast, the offshore field had a unit cost of $19.74. Note that both fields are in southern US states, and so faced similar currency and input market conditions.

4.3   The OPEC Cartel

OPEC, in 2014 (the limit of the data), 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 its inception in 1960.[20] The history of the first 30 years of OPEC is well documented in Yergin (1991) and Crémer and Salehi-Isfahani (1991).

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 pro-

---

[20]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.

ducers; an efficient, economic and regular supply of petroleum to consuming nations; and a fair return on capital to those investing in the industry".[21] 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.

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

For its first ten years, OPEC was apparently ineffective in influencing the oil price. This changed in 1973 when the outbreak of war in the Middle East allowed OPEC member countries to raise the oil price four-fold, albeit with (according to secondary sources) little coordination among themselves. Until 1982, the cartel did little more than set price guidance for its members and hope that that guidance would be followed.

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

During the 1990s, expansion in oil demand outstripped the ability of non-OPEC producers to expand supply and the power of the OPEC cartel appears to have increased.[22] By November 1997, the cartel was in a position to exert substantial influence on the market price, the cartel expanded production apparently either having broken down or mis-judged global demand in the face of the Asia crisis. This lead oil prices to decline from $35 per barrel to $10 per barrel.[23] Finally, in March 1999, with the cooperation of Russia, Norway and Oman, OPEC countries succeeded in cutting production and the price moved back into the $30-$40 per barrel range.

In 2000, the cartel explicitly announced a target price band of $22-$29 per barrel. The quota system was amended to include automatic adjustments should the reference price fall outside this interval. This change heralded the modern era of OPEC and the most sophisticated coordinating mechanism seen to date. Unfortunately for the cartel, the second Gulf War and an expansion in demand from developing countries made the price band unsustainable leading to its suspension in January 2005. The quota system, however, lived on.[24] Although the quotas are not transparent, it is clear that they are highly asymmetric.[25]

By July 2008 the Brent spot price had increased to an intra-day high of over $140 per barrel, a level in stark contrast to the price band in (ineffective) operation a mere 4 years earlier. Through to the end of 2014, oil price remained high by historical standards.[26]

The efficacy of the OPEC cartel over time has been, at best, variable. Especially in the early 2000s, its policies mirrored those of collusive structures seen in other industries. In other

---

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

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

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

[24]See Kohl (2005) and Fattouh (2007)

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

[26]Despite decreasing to below $35 a barrel in late December of 2014.

periods, it is less clear that OPEC had the ability to coordinate its members' production. The question of whether OPEC is best characterized as a political vehicle for the Gulf countries, or as a long running industrial cartel does not require an answer for the purposes of this paper. To the extent that an OPEC member has market power and distorts production, the measurement approach adopted in the simulations will account for it.[27]

References

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

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

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

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

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

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

Lin Lawell, Cynthia, "Market Power in the World Oil Market: Evidence for an OPEC Cartel and an Oligopolistic Non-OPEC Fringe," Technical Report, Cornell University 2017.

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

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

The World Bank, *The Changing Wealth of Nations Measuring Sustainable Development in the New Millennium: Environment and Development*, World Bank, 2011.

White, T. Kirk and Martin Rotemberg, "Measuring Cross-Country Differences in Misallocation," Technical Report, New York University 2017.

— **, Jerome P. Reiter, and Amil Petrin**, "Imputation in U.S. Manufacturing Data and Its Implications for Productivity Dispersion," *The Review of Economics and Statistics*, Forthcoming.

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

---

[27]This is a consequence of working with cost data directly. By contrast, if the more standard IO approach were adopted, in which demand estimates were combined with a pricing model to back out costs, then this issue would be a significant hurdle to the credibility of any estimates.