Online Appendix to *Market Power, Production (Mis)Allocation and OPEC* – Not For Publication

OA.1 Oil: product and production

A Crude oil

Crude oil is the oil that is delivered to refineries for processing into the various hydrocarbon products used by end consumers.[1] 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, and crudes that higher sulphur content/sour, 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 is 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^o and 50^o API.[2] Most refineries are geared toward processing crudes in the 30^o to 40^o API range, with some variation across refineries. Different refineries will have different sulphur tolerances as well. Hence, crudes that lie outside the 30^o to 40^o 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.

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. 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 need 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.

B Crude oil production

Once a deposit is discovered it needs to be exploited. A deposit will be located in a field, which is a deposit, or set of deposits, sitting within a common geological structure. 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.

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 will have 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. Other forms of onshore deposit require other 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 mixed into loose rock or sand. These crudes are recovered through surface mining 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 production costs.

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

[2] Common benchmark crudes are Arabian Light, Brent and WTI which have densities of 34^o, 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.

Offshore production. Off-shore deposits have additional 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 (at 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 heated to keep the crude at a steady temperature. In deep water wells anti-freeze is also often added at the well head. Distance from land affects the way labor can be housed, transported and rotated and how the crude can be stored and transported back to land. The weather presents a series of additional challenges related platform stability, production interruptions and safety due to variation in storms, hurricanes and sea states across the globe.

Examples. The cost differences that arise from different operating environments are well 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 (comprising all of operating and capital expenditures) of the onshore field from 1970 onward was $7.57 per barrel, and computed by dividing total expenditures by total production from 1970 (inclusive). The offshore field had a unit cost of $19.74., and both fields were selected to be in the southern US states to keep currency and input market conditions as similar as possible.

OA.2 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. The data cover the operations of each oil field around the world, and is documented in Rystad Energy (2015). For each field the data include production and different operating and capital costs, as well as the characteristics of the field such as the 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).

A Data collection and sources

The Rystad data covers the oil global industry. As a result it is collected from a variety of original and secondary sources, ranging from high quality government reports in countries such as Norway and the United States, through company reports for large private companies, to interviews with shipping companies and oil service companies. Collating and reconciling these sources is a difficult process, particularly in politically unstable areas such as, for instance, Syria from 2012 onward where ISIL controlled portions of Syria in this period, including oil producing assets. 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 where numbers are not directly reported. This is unavoidable when attempting to study a global industry that has strategic importance.

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 (companies and otherwise), 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.

By contrast, in an industry, like upstream oil production, where production occurs in many sovereign countries in a decentralized manner, data will be similarly fractured. Thus, any effort to study this globally important industry requires confronting this challenge. Despite the fractured nature of data on the oil industry, many high quality data sources exist. These data sources include government reports,

Table OA.1: Data Cleaning and Sample Frame

	Fields	Total Production in Trillions of Barrels
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

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 similar to those confronted by national statistical agencies when compiling measures of aggregate economic activity, like GDP figures.

A.1 Collection

In this section we provide a short description of how this dataset is assembled by Rystad. 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, but also fields that could possibly come online in the future.

Table OA.1 shows the data cleaning steps we perform, along with the total number of fields after each step. The data has 66,000 unique fields, but only 21,000 produce at any point in time. These non-producing fields are used by Rystad for forecasting purposes — they are estimates of the production of a field that has not started production yet. Of the 21,000 producing fields, about 2,000 are gas only fields which we drop from our analysis. Linking these fields to measures of reserves leaves us with 13,000 fields, but 92 percent of production.

A.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 assets such as whether the asset is an oil sands, shale, or an offshore oil platform of a certain depth, or wether the oil field produces gas, or heavy or sour crude oil. This information is complemented with more detail on the exact geology of the oil field. As well, Rystad keeps track of the discovery date and depletion of these oil fields.

Much of the field data is obtained from government and company sources. For instance, in the United States, there is detailed data on field level production from information on the royalties that oil firms pay for their leases. As well, many companies publish information on production levels and reserves in different fields. The second type of data are economic. These data are mainly sourced from three different places including company and government Reports, and information from oil service firms through documents and interviews. Table OA.2 shows information on the data sources for different regions in the world. While data in the United States comes from direct measurement, data in Saudi Arabia is mainly extrapolated from the geological attributes of fields in this country.

While there is a considerable amount of data on oil fields, most of the economic and production data is extrapolated from similar oil fields. For instance, there is no data on fields in Saudi Arabia since Saudi ARAMCO does not publish information on it's operations. Therefore, Rystad uses information on costs of oil fields that are comparable, principally those in Iraq, to infer the costs of production in Saudi Arabia.

Table OA.2: Data Source by Region of the World

Region	Government webpage	Annual Report	Investor Presentation	Company Press Releases	Articles on oil	Other document	Analyst estimate	Modelled
Australia	57	1	1	11			18	11
Caribbean		9					3	88
Central America	57							43
Central Asia		8	1	6	1	6	1	77
East Africa		6		3	7		1	20
East Asia		26	3	2		4	2	62
Eastern Europe	23	3		1			1	45
Middle East	2	11	1	1	4	2	14	64
North Africa	3	13	3	3	4	3	9	63
North America	26	8	1	1		1	4	60
Oceania		9						10
Russia		31	1	2	1	6	12	47
South Africa								100
South America	65	2		1	2			30
South Asia	21	1	8	2	11		2	54
South East Asia	17	12	2	1		1	1	66
Southern Europe	57	1	1	9				32
West Africa	23	23	2	3	5	3	1	40
Western Europe	87				1		1	11
Total	29	12	1	2	1	1	4	49

B Measurement of main variables

B.1 Production

All units of fuel are converted to crude barrel equivalent, since different fuels like Condensate and Gas have different energy production by volume. Thus, in the paper we refer to production as a homogeneous product measured in crude oil equivalents. We 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 assets are quite large, such as the Ghawar field, with hundreds of rigs. Indeed, in 2014, the tenth percentile oil asset produces 8,000 barrels per year, while the 90th percentile produces 1.9 million barrels per year. Figure OA.1 shows the time series of total production from 1970 to 2014, as well as the total number of producing fields.

B.2 Measurement of reserves

Central to much of the discussion in the paper is the notion of reserves. The reserve is the unextracted, but recoverable, quantity of oil remaining in the ground in a field. Depending on the geology, between 25% and 75% of the crude will remain in the deposit after production has concluded. 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 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 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. 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% 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 above,

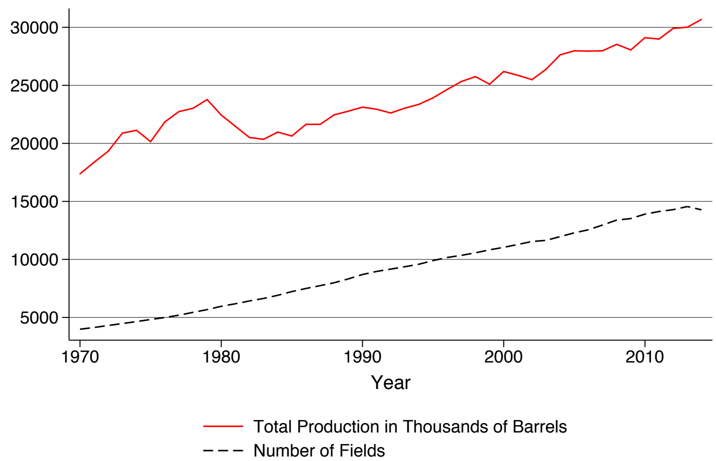

Figure OA.1: Total Production and Number of Oil Producing Assets

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 the production stage of a field, and the country in which the field exists. Untapped deposits have less precise reserve estimates, since there no actual production data or well pressure data to rely on. 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 OA.3 shows total reserves in the world in 1970, 1990, 2000, and 2010. They stand at 439 trillion barrels in 2014, if one considers P50 and a forecasted price of oil of $70 a barrel.

Table OA.3: 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 (prior to section 5) P50 values at an oil price of $70 a barrel are used to report reserves. In section 5, a field's reserves in 1970 are computed as the sum of: i) the actual production history from 1970 to 2014, and ii) the P50 value at an oil price of $70 a barrel in 2014.