Journal of International Economics 125 (2020) 103323

Contents lists available at ScienceDirect

Journal of International Economics

journal homepage: www.elsevier.com/locate/jie

Full length articles

Global sourcing in oil markets☆

Farid Farrokhi *

Purdue University, USA

ARTICLE INFO

Article history:
Received 28 May 2018
Received in revised form 14 March 2020
Accepted 16 March 2020
Available online 24 March 2020

Repository: Mendeley Data https://data.mendeley.com/datasets/s6kdxyn3s7/draft?a=f3ceeb0f-0898-45ba-8f3a-846c8ed5556e

Keywords:
Global sourcing
Crude oil markets
Refining industry
Maximum likelihood
General equilibrium

ABSTRACT

This paper develops a multi-country general equilibrium model that incorporates crude oil global sourcing by refineries and refined oil demand by downstream industries that are connected through international trade and input-output linkages. I exploit detailed data on crude oil imports of US refineries, and derive a new procedure that allows for an all-in-one estimation of which suppliers refineries select and how much they buy from each. Using the estimates to evaluate counterfactual policies, I find: (i) A boom in crude oil production of a source changes the relative prices of crude oil across countries modestly which I interpret as the extent to which the behavior of crude oil markets deviates from an integrated global market. (ii) Lifting the ban on U.S. crude oil exports creates notable distributional impact across American crude oil producers and refineries. (iii) Due to the importance of oil trade and consumption, gains from trade are larger compared to the existing benchmarks.

1. Introduction

Trade in oil accounts for a large share of world trade, but occupies a small part of the trade literature. The literature on international trade has included the oil industry only in multi-sector frameworks designed for manufacturing rather than natural resources. The fields of industrial organization and energy economics lack a general equilibrium framework to put the oil industry into global perspective. Both the specifics of this industry and a general equilibrium analysis must come together to address trade-related questions about oil markets.

This paper develops a general equilibrium framework to study how changes in oil markets, such as a boom in U.S. crude oil production, affect oil prices and trade flows across the world. Specifically, I use the framework to examine a few key applications. First, I study the extent to which crude oil markets behave as one integrated global market. To do so, I explore how much a shock to crude oil production of a source changes the relative prices of crude oil across the world geography. Second, to demonstrate how the model can be used to evaluate policy, I examine the implications of lifting the ban on U.S. crude oil exports. This exercise asks: how much does the price of U.S. crude oil rise when it can be sold in global markets? What distributional gains does it create between crude oil suppliers, refineries as consumers of crude oil, and downstream consumers of refined oil? Lastly, I study the welfare implications of ceasing international trade between countries or regions of the world. I compare my results with benchmarks in the literature to shed light on the importance of oil trade and consumption for gains from trade.

To address these questions, I develop a model of crude oil global sourcing by refineries and refined oil demand by downstream users. Heterogeneous refineries decide which international suppliers to select and how much crude oil to purchase from each. Global trade in crude oil and supply of refined oil are the endogenous outcomes of the aggregation of refineries' sourcing and production. Refined oil demand is generated by final consumers as well as downstream industries that are connected through international trade and input-output linkages in the style of Caliendo and Parro (2015). The framework is designed for a medium-run horizon in which crude oil production flows and incumbent refineries are taken as given, and equilibrium determines prices and quantities of all goods in the locations of supply and demand.

The production of crude oil is concentrated in a relatively small number of sources from where it flows to numerous refineries around the world. I document the main patterns of these flows by exploiting data on the imports of American refineries. In particular, (i) most refineries import from a few supplier countries, (ii) refineries with similar observable characteristics allocate their total crude oil purchases across suppliers in different ways.

☆ I am grateful to my former advisors Jonathan Eaton and Stephen Yeaple for their guidance and support, and to Paul Grieco and James Tybout for helpful discussions. I am thankful to two anonymous referees and the editor Andres Rodriguez-Clare as well as Thomas Covert, Ed Green, Cecile Gaubert, Marc Henry, Russ Hillberry David Hummels, Barry Ickes, David Jinkins, Amir Khosrashahi, Kala Krishna, Ahmad Lashkaripour, Michael Plante, Heitor Pellegrina, Anson Soderbery, Tim Schmidt-Eisenlohr, Felix Tintelnot, Chong Xiang and many seminar participants for helpful comments and suggestions. I gratefully acknowledge the support through the Center for Research on International Financial and Energy Security at Penn State. All errors are my own.
* Corresponding author.
E-mail address: ffarrokh@purdue.edu.

https://doi.org/10.1016/j.jinteco.2020.103323
0022-1996/

I model refineries' procurement by focusing on the logistics of crude oil sourcing. Transport costs not only vary across space due to distance and location of infrastructure, but also fluctuate over time due to availability of tankers and limited pipeline capacity. Because of costs fluctuations, refineries –which operate continuously– lower their input costs when they spread their purchases across a greater number of suppliers. Offsetting this benefit, sourcing from each supplier creates fixed costs associated with establishing business relationships. The trade-off between these gains and fixed costs explains pattern (i).

Using the *observed* characteristics of refineries and suppliers, I specify variable costs that a refiner faces to import from a supplier (including price at origin, distance effect on transport costs, and a cost-advantage for complex refineries). This specification alone fails to justify pattern (ii). To accommodate it, I introduce *unobserved* variable costs of trade to the pairs of refiners and suppliers.

I then develop a new procedure to jointly estimate fixed and variables costs that refineries face in their international crude oil sourcing. The task has proved challenging because a refiner's buying decisions are interdependent, e.g. adding a supplier may lead to dropping other suppliers or purchasing less from them. This interdependency is absent from typical export participation models such as Melitz (2003) which could be dealt with by, for example, a Tobit formulation. In addition, on the import side, estimating a structural gravity equation at the firm level requires a strong timing assumption by which a firm learns about its unobserved components of variable trade costs only after selecting its suppliers, e.g. see Antràs et al. (2017).[1] I depart from this timing assumption by deriving a likelihood function that combines data on whom refineries select and how much they buy from each. This likelihood function lets a refiner not only buy less from its higher-cost suppliers but also select them with lower likelihood from an econometrician's point of view. As a result, the estimation *jointly* identifies parameters that affect trade quantities and selections.

This methodological departure is crucial to my results. Either large diversification gains with large fixed costs or small diversification gains with small fixed costs could explain why refineries typically buy from a few suppliers. Compared to independent estimations of quantities and selections, my all-in-one estimation generates smaller gains from global sourcing and smaller fixed costs, and it delivers a largely better fit to data. There is information in selections about the elasticity of substitution across suppliers, which we miss if we estimate this elasticity using data on only quantities. There is also information in quantities about fixed costs, which we miss if we estimate fixed costs using data on only selections.

I embed the model of refineries' sourcing, with the parameter estimates, into a general equilibrium trade framework that features multiple industries, input-output linkages, and allows for a less-than-unity elasticity of demand for refined oil. I complete my empirical analysis by calibrating the framework to aggregate data from 2010 on 33 countries and 6 regions covering all flows of oil from production of crude to consumption of refined and using parameters from world input output database (WIOD). The popular method of exact hat algebra cannot be readily applied to this setup, because refineries' trading relationships endogenously change in response to price shocks. I develop a hybrid procedure to compute counterfactual equilibrium in which I solve oil-related variables in *levels* and downstream variables in *changes*.

The estimated model fits well out of sample. While I use cross-sectional data from 2010 to estimate the model, I check its predictions for changes during 2010 to 2013. To do so, I re-calculate the equilibrium by updating crude oil production and refining capacity of all countries to their factual values in 2013. The new equilibrium gives a reasonable prediction compared with factual changes to prices of crude oil in the U.S. relative to the rest of the world, and the pass-through to the price of refined oil. In addition, the model closely predicts crude oil imports, number of suppliers, and total input purchases of refineries in the U.S. economy.

I use my framework as a laboratory to simulate counterfactual experiments. First, I focus on a counterfactual world where only U.S. crude oil production changes. Specifically, I consider a 36% rise in U.S. production corresponding to its rise from 2010 to 2013. The price of crude oil at refinery drops by 11.5% in the U.S., 10.4% in other countries of Americas, 9.8% in African countries, and on average 9.0% elsewhere. The production boom changes relative prices of crude oil spatially across countries to a modest degree which can be interpreted as the extent to which the behavior of crude oil markets deviates from an integrated global market. In particular, compared with Americas and Africa, countries in Europe, Russia, and part of Asia are less integrated with the U.S. market.

To show how the model can be used to study counterfactual policies I explore the implications of lifting the ban on U.S. crude oil exports. I find that had the ban been lifted when U.S. production rose from 2010 to 2013, average prices of U.S. crude oil would have risen by 4.05%, profits of U.S. refineries would have decreased by 6.97%, and American consumers would have faced a negligibly higher prices of refined oil. These changes translate to \$7.49 billion increase in annual revenues of U.S. crude oil producers, and \$7.76 billion decrease in annual profits of U.S. refineries. Furthermore, I compare the results of this experiment to the factual data after the US government lifted the ban in 2015. The model predictions of changes to US crude oil exports in 2010–13 are reasonably close to actual changes to US exports in 2015–18.

Lastly, I study the importance of oil trade and consumption for gains from trade. I find gains to U.S. consumers as the change to their real wages when trade in all goods between the U.S. and the rest of the world is prohibitive. I compare my results to a benchmark provided by Costinot and Rodriguez-Clare (2014). This benchmark is appropriate since it is the same as the downstream part of my model, with crude oil to be part of mining and refined oil included in fuels. I find that gains from trade are notably larger here than those in the benchmark. By moving US to autarky, US real wage falls by 28.8% here compared to 8.0% in the benchmark.

In addition, my analysis highlights the importance of distinguishing between rents on natural resources and wages paid to labor. Such distinction is absent in widely used labor theories of trade. For example, I find that by moving US to autarky real GDP per capita would fall by 11.2% compared to 28.8% fall in real wage. The reason is the dramatic rise of oil prices, and hence, oil revenues in the US. Therefore, who receives the rents would be crucial for distributional welfare implications.

This paper fits into the literature on international trade in two broad ways. First, it closely relates to the literature on quantitative models of trade in intermediate goods. The seminal work of Eaton and Kortum (2002) and its extensions such as Caliendo and Parro (2015) address trade in intermediates at the level of country or industry. Several recent studies emphasize the importance of sourcing at the firm level in manufacturing or grid-cell level in agriculture to examine gains from input trade (Gopinath and Neiman, 2014; Halpern et al., 2015; Antràs et al., 2017; Blaum et al., 2018; Farrokhi and Pellegrina, 2019). I develop new tools that can be used to estimate sourcing models and other setups in which individual buyers face both discrete and continuous choices. I then apply this new estimation procedure as well as existing tools from the general equilibrium tradition in the trade literature to a carefully specified model of oil trade.

This carefully specified model serves two main purposes. First, it helps me exploit rich micro data on refineries' sourcing to identify parameters that control margins of crude oil imports. These parameters govern the extent to which refineries' location with respect to suppliers as well as their complexity and capacity constraints matter for transmission of price shocks in oil markets. In addition, the heterogeneity in variable costs of refineries' imports helps me identify the substitution

[1] In addition to papers that use firm-level import data, another set of studies use product-level import data, e.g. Broda and Weinstein (2006) among many. What makes these two bodies of literature comparable is a similar demand system that gives rise to micro-level gravity equations conditional on trading relationships. Based on such a gravity equation, these studies estimate the elasticity of substitution across suppliers or goods by using only trade quantities and independently from buyers' selection decisions.

elasticity across crude oil suppliers. This identification is key to examine how large gains from global sourcing are through the input-saving channel, and how a local supply shock affects relative prices of crude oil across the world geography.

Second, this microfoundation helps me conduct counterfactual exercises tightly connected to supply and demand variables in crude oil markets. For example, standard assumptions in existing multi-industry models attribute variations in cross-country crude oil production to labor productivity shifters, capture all frictions in crude oil trade by iceberg trade costs, and assume zero profits or constant markups in the oil industry. In contrast, this framework brings data on the magnitude and quality of crude oil production and trade flows, allows for an array of frictions in the logistics of crude oil imports, makes room for endogenous rents in the refining industry, and distinguishes between rents on crude oil resources and wages paid to workers.

On the methodological side, models of firms' sourcing, in contrast to typical models of export participation such as Melitz (2003), feature interdependent decisions for selecting suppliers. In explaining selections into import markets, Antràs et al. (2017) is the closest to my model of sourcing. While they allow for a more general structure of fixed costs, I allow for a richer specification of variable trade costs. I use this alternative specification to deal with a sample selection bias in estimating trade elasticity at the level of individual buyers. Another difference is that in Antràs et al. (2017) firms can largely grow by global sourcing, while in my model refineries face a limit to the amount they can produce. This difference highlights the medium-run horizon of my model as opposed to the long-run horizon of theirs.

The trade literature has studied manufacturing more than natural resources or agriculture. One well-known result that holds across benchmark trade models is that gains from trade are often small (Arkolakis et al., 2012). These small gains are at odds with the critical role of trade in natural resources. Results in this paper confirm that gains from oil trade are in comparison notably large. Trade in oil benefits net importers of crude oil as their production costs would increase in autarky due to higher oil prices. In addition, it can benefit large net exporters of crude oil as they would face lower oil prices, and so, receive lower oil revenues if they were in autarky. The gains countries receive from these channels are sizable because the elasticity of substitution between oil and other inputs is small, meaning that input-users can hardly substitute oil with other factors of production or intermediates.[2] A recent study by Fally and Sayre (2018), which relies on similar mechanisms, also finds large gains from trade in commodities.

This study relates to a large body of economic research on the oil industry. A literature aims to identify causes and consequences of oil shocks using econometric techniques and time-series oil price data (Kilian, 2009; Hamilton, 2011). These papers do not incorporate a room for the geography of oil production or logistics of oil trade, and do not model economic decisions made by oil producers and consumers. My paper complements this literature by (i) studying oil prices and trade flows spatially across the geography of the world, and (ii) since I model economic decisions that underlie oil trade, I can explicitly address a wide set of counterfactual policies. This paper also complements a recent study by Çakir Melek et al. (2017) who employ a two-country DSGE model to examine the macroeconomic impact of the shale boom. Several papers study economics of refineries focusing only on U.S. market (Chesnes, 2015; Sweeney, 2015). I complement these studies by modeling, estimating, and solving for a general equilibrium incorporating the oil industry in *global* markets.

2. Background & empirical patterns

The purpose of this section is to motivate main features of the model based on evidence. I provide background on the refining industry, and document detailed features of international oil trade data. Then, I explain how the observed patterns motivate the model.

Fig. 1. Refinery process flow chart. Source: Simplified illustration based on Gary et al. (2007).

2.1. Background

2.1.1. Structure of a refinery

A refinery is an industrial facility that converts crude oil into refined oil products. Fig. 1 shows a simplified flow chart of a refinery. Crude oil is first pumped into the distillation unit. Refinery *capacity* is the maximum amount of crude oil (in barrels per day) that can be processed in the distillation unit. The process of boiling crude oil in the distillation unit separates the crude input into a variety of intermediate fuels based on differences in boiling points. *Upgrading units* further break, reshape, and recombine the heavier lower-value fuels into higher-value products.[3]

2.1.2. Types of crude oil and complexity of refineries

Crude oil comes in different types. The quality of crude oil varies mainly in two dimensions: density and sulfur content. Along the dimension of density, crude oil is classified between light and heavy. Along the dimension of sulfur content, it is classified between sweet and sour.

The *complexity index* measures refineries' capability for refining low quality crude inputs. This index, developed by Nelson (1960a, 1960b, 1961), is the standard way of measuring complexity in both the academic literature and the industry. This index is a weighted size of upgrading units relative to capacity size.[4] To produce the same value of output, refining heavy and sour crude involves more upgrading process. For this reason, a more complex refinery has a cost advantage for refining lower quality crude oil.

2.1.3. Crude oil procurement, contracts, and logistics

For the most part of oil markets, production and refining are not integrated and refiners engage in arm's length trade to secure supplies for their refineries (Platts, 2010). 90–95% of all crude and refined oil are sold under term contracts, usually *annual* contracts that may get renewed after a year (Platts, 2010). A typical term contract covers multiple transactions within a year (Senate Report, 2003). These contracts specify the quantity of trade, dates of delivery, and fix the method of calculating the price (Senate Report, 2003; Platts, 2010).[5]

[2] See Section 6.5 for a detailed discussion on sources of gains from oil trade in this paper.

[3] A refinery produces a range of products that are largely joint. In the short run where refinery technology and downstream production technology are given, the composition of refined oil products tend to remain unchanged. These products include gasoline, kerosene and jet fuels, diesel, oil fuels, and residuals. Typically, heavier fuels are the byproduct of lighter ones.

[4] Let b_k be the size of upgrading unit $k = 1, \ldots, K$. To every upgrading unit k, weight w_k is assigned based on costs of investment in unit k. The complexity index is defined as $(\sum_{k=1}^{K} w_k b_k)/r$ where r is refinery capacity (i.e. size of distillation unit). See the Online Appendix for details.

[5] The remaining 5–10% is the share of spot transactions. By definition, a spot transaction is a one-off deal between willing counterparties. They are surpluses or amounts that a producer has not committed to sell on a term basis or amounts that do not fit scheduled sales (Platts, 2010). Independent companies such as Platts and Argus post prices for these spot transactions, particularly for three crude oil streams, West Texas Intermediate, Brent, and Dubai. The price specified in a term contract is typically tied to the price of one of these three benchmark crude oil streams. West Texas Intermediate is produced in the U.S. Gulf Coast, Brent in the North Sea of the United Kingdom, and Dubai in the UAE. The benchmark chosen for a term contract depends on the type of crude and the location of trade. The price in a contract is typically a function of a benchmark price plus or minus adjustments for quality and factors related to market conditions, based on a formula specified in the contract. For details on the relation between posted prices of benchmark crude oil streams and term contracts, see (Fattouh, 2011, Chapter 3). The amount of discounts or premiums in contracts are often unobservable as they are kept confidential by oil companies.

Crude oil transportation costs not only vary across space but also fluctuate over time. Specifically, for a fixed source-destination route these costs fluctuate within a year and their monthly growth rates differ across routes. For example, according to Oil & Tanker Trades Outlook (2015, Februrary) in January 2015 transport costs of small-size crude oil tankers increased by 27% between the Caribbean and the U.S. Gulf Coast whereas those costs fell by 9% between the Persian Gulf and East Asia. These variations reflect seasonal congestion in transportation resources, maintenance at the location of supply, and fluctuations in the daily availability of tankers and pipeline capacity.

Refineries heavily rely on a constant supply of crude oil as they operate 24/7 over the entire year. In particular, the costs of shutting down and restarting are large.[6] As a result, careful scheduling for procurement of crude oil is important. The optimal logistical arrangements in crude oil sourcing are subject to the availability of tankers, ports storage tanks, inland pipeline slots, and other logistical frictions in oil procurement. Available information from inside the business confirms the importance of input-saving decisions by refiners. For example, the CEO of Phillips 66 asserts that "the single biggest lever we have to improve value in our refining business is through lowering our feedstock costs. A saving of $1 per barrel across our refining system is worth about $450 million of net income to us."[7] A number of academic studies have developed mathematical programming techniques to address the problem. A notable paper is Shah (1996) which formulates a refinery's optimal scheduling of multiple crude oil grades of different quality and origin.

2.1.4. Market structure

An overview of interviews with representatives of the refining industry conducted by RAND, writes: "Although refining operations share many technologies and processes, the industry is *highly competitive* and diverse" (Peterson and Mahnovski, 2003, p. 7). Textbooks on engineering and economics of refineries assume that refineries take prices of refined products and prices of crude oil as given, e.g. as a widely used reference see Gary et al. (2007). Such a description is also in line with reports by governments. For example, according to Economics of Petroleum Refining by Canadian Fuels Association (2013, p. 3), "refiners are price takers: in setting their individual prices, they adapt to market prices".[8]

2.2. Empirical patterns

I document several key patterns in oil trade data at the level of countries and refineries in Sections 2.2.1 and 2.2.2. I then discuss how these patterns motivate the model in Section 2.3. I present details on the construction of data and further description of empirical patterns in the Online Appendix.[9]

2.2.1. Country-level observations

Data. Table 1 lists countries and reports their crude oil production, refining capacity, average complexity index, average utilization rate, and refined oil consumption. Table A.1 lists country-level variables and their sources that I have put together. The sample is a cross-section in 2010 consisting of 39 countries and regions covering the entire world geography. There are 359 nonzero trade flows of crude oil plus 31 own-purchases, summing up to 390 nonzero entries in the crude oil trade matrix. I have conducted an accounting of oil flows, described in the Online Appendix, to deal with potential mismeasurement in reported oil trade. See the Online Appendix also for more detailed patterns in the country-level data.

Macro Pattern 1. International trade in crude oil features a gravity equation conditional on nonzero flows.

The following reports an OLS regression of international trade in crude oil against characteristics of exporters and importers for the sample of nonzero trade flows,

$$\log Q^0_{ni} = constant + \underset{(0.116)}{0.99}\ \log TQP^0_i + \underset{(0.117)}{1.05}\ \log TRC^0_n - \underset{(0.105)}{0.96}\ \log dist_{ni} + error^0_{ni},$$

where Q^0_{ni} is quantity of crude oil trade from i to n, TQP^0_i is total crude production of i, and TRC^0_n is total refining capacity in n, all in units of barrels per day. Inside parentheses are standard errors. This relation resembles a gravity equation. The absolute value of estimated coefficients on the mass of exporter and importer, and on the distance between the two, are all nearly one. Note, however, that variables are measured in quantities rather than dollar values.

Macro Pattern 2. (a) International trade in refined oil features a gravity equation. (b) Refined oil trade contains significantly more nonzero flows, and it is much more two-way than crude oil trade.

A regression for refined oil trade delivers

$$\log X^1_{ni} = constant + \underset{(0.114)}{1.25}\ \log TRC^1_i + \underset{(0.091)}{1.23}\ \log GDP_n - \underset{(0.121)}{1.63}\ \log dist_{ni} + error^1_{ni},$$

where X^1_{ni} is value of refined oil trade from i to n, TRP^1_i is total refined oil production of i, and GDP_n is a proxy for the size of refined oil demand in the importer country, as GDP alone explains 85% of variations in total refined oil consumption (in log terms) across countries.

For crude oil, nonzero flows (including domestic purchases) account for 32% of all entries when defined between producers (countries with nonzero production) and all destinations. Compared with crude, international flows of refined oil contain 2.5 times more nonzero entries. In addition, 89.5% of refined oil trade (in terms of value) is two-way compared to 26.4% for crude.

2.2.2. Refinery-level observations

Data. I have used three *refinery-level* datasets collected by the U.S. Energy Information Administration (henceforth, EIA): (i) capacity of distillation unit and upgrading units, (ii) imports of crude oil, (iii) domestic purchases of crude oil.[10]The merged dataset contains 110 refineries in 2010 importing from 33 countries. The sample consists of volume of imports (by origin and type of crude), volume of domestic purchases, capacity of distillation unit, capacity of upgrading units, and refinery location. Volumes and capacities are measured in units of barrels per day. Using the data on upgrading units, I construct Nelson complexity index of refineries.

Since EIA does not assign id to refineries, I have manually matched the three above mentioned pieces of data. Not all refineries in one of the three datasets can be found in the other two. To match these data I have checked the entires of each one with the other two, often using online information on refineries to make sure of their correct geographic location. The merged sample accounts for 95% of total capacity and 90% of total imports of the U.S. refining industry in 2010.

[6] Moreover, refineries keep inventories of crude, but since inventory costs are large, the inventory levels are significantly smaller than refinery capacity. In 2010, the total refinery stock of crude was less than 1.7% of total use of crude oil in the U.S., that is, the inventories suffice for less than a week of usual need of crude. Moreover, the change in these inventories from Dec. 2009 to Dec. 2010 was only 2.5% which translates to only one-fifth of a day of the crude oil used in the entire year.

[7] See the report entitled "Phillips66 Delivers on Advantaged Crude Strategy" at the website of Phillips 66.

[8] The assumption, however, remains a simplification particularly for studying product prices across regions within a country. For a study that addresses imperfect competition in the sale side of refineries across US regions, see Sweeney (2015).

[9] Download the Online Appendix by clicking http://web.ics.purdue.edu/~ffarrokh/Farrokhi_jmp_OnlineAppendix.pdf.

[10] While (i) and (ii) are publicly available, I obtained (iii) through a data-sharing agreement with EIA that does not allow me to reveal refinery-level domestic purchases.

Table 1
List of countries & selected oil-related variables, year 2010.

Country	Crude oil production (1000 b/d)	Crude oil exports (1000 b/d)	Crude oil imports (1000 b/d)	Total refining capacity (1000 b/d)	Average utilization rate	Average complexity index
Algeria	1540	1116	0	476	0.89	1.34
Angola	1899	1869	0	41	0.72	1.79
Azerbaijan	1035	907	0	422	0.30	3.89
Brazil	2055	619	321	2019	0.87	4.28
Canada	2741	1449	521	2158	0.84	8.14
China	4078	61	3515	8589	0.88	2.73
Colombia	786	577	0	303	0.69	4.67
France	0	0	1637	2099	0.78	6.96
Germany	0	0	2296	2551	0.90	7.90
India	751	0	2099	3001	0.95	3.20
Indonesia	953	338	177	1071	0.74	3.75
Iran	4080	2622	0	1536	0.95	3.91
Iraq	2399	2021	0	675	0.56	4.05
Italy	0	0	1707	2473	0.69	6.87
Japan	0	0	3670	4893	0.75	7.84
Kazakhstan	1525	1406	118	365	0.65	5.25
Korea	0	0	2316	2859	0.81	4.98
Kuwait	2300	1389	0	991	0.92	5.02
Libya	1650	1306	0	400	0.86	1.57
Mexico	2621	1220	0	1630	0.86	7.62
Netherlands	0	0	1034	1276	0.81	7.52
Nigeria	2455	2263	0	534	0.36	4.43
Norway	1869	1602	12	338	0.83	4.39
Oman	865	788	0	90	0.85	2.56
Qatar	1129	824	0	358	0.85	4.25
Russia	9694	4524	0	5745	0.90	4.38
Saudi Arabia	8900	6897	0	2201	0.91	3.79
Singapore	0	0	905	1436	0.63	5.29
Spain	0	0	969	1346	0.72	7.04
UAE	2415	1883	0	818	0.65	2.44
United Kingdom	1233	740	1008	1975	0.76	8.41
United States	5471	42	10,388	18,608	0.85	9.77
Venezuela	2216	1130	0	1357	0.80	5.41
RO_America	1408	551	1516	3198	0.74	4.79
RO_Europe	662	190	4049	5989	0.75	7.01
RO_Eurasia	324	46	749	2151	0.48	4.51
RO_Middle East	1028	245	66	999	0.85	3.72
RO_Africa	2257	1627	875	2017	0.75	3.11
RO_Asia & Oceania	2047	1138	1439	3050	0.77	3.84
World	74,386	41,389	41,389	92,039	0.81	4.93

In addition, I link refinery-level imports to crude oil prices. Specifically, I have constructed a concordance between worldwide crude oil streams collected by Bloomberg and a classification of crude oil based on origin country and type. Using this concordance and the free on board prices reported by Bloomberg, I compiled the prices of crude oil at each origin country for each type. In addition, using EIA data on before-tax price at the wholesale market of refinery products, I construct the price of the composite of refinery output.

To clarify my data limitations, I do not observe: sales or production of individual refineries, from which domestic suppliers a refinery purchases, and crude oil pipelines within the United States. I make a number of assumptions in model specification to deal with these data limitations which I will discuss throughout the paper.

Table 2 reports a summary of statistics. Fig. A.1 shows the location of refineries within the U.S. geography. Figs. A.2–A.4 show the distribution of refineries' capacity, distance to coast, and complexity. Appendix A contains supporting regressions for the patterns reported below.

Table 2
Summary of statistics of American refineries' characteristics, year 2010.

Percentile	P10	P25	P50	P75	P95
Capacity (1000 b/d)	8.9	61.1	116.7	236.8	452.1
Distance to coast (km)	2.8	9.3	127.2	837.3	1437.6
Complexity index	4.7	7.8	9.1	10.9	14.2
Imports/Capacity (%)	0.0	2.7	30.4	61.4	88.1
Utilization rate (%)	69.2	77.0	83.7	89.5	96.8
Number of foreign suppliers	0	1	2	7	12

Micro Pattern 1. Refineries typically buy crude oil from multiple sources, and buy multiple types.

Table A.2 reports the number of refineries importing from none, one, and more than one origin. More than half of American refineries, accounting for 77.2% of U.S. refining capacity, import from more than one origin. The last row of Table 2 reports the distribution of the number of import origins. The median refiner imports from two countries. The distribution is skewed to the right, and the maximum is 16 (compared to 33 origins in the aggregate).

In Table A.3, types are classified into four groups as (light, heavy) × (sweet, sour).[11] The table shows that 88.4% of refineries import more than one type of crude oil, and 36.1% of refineries import all types. I emphasize that the data invalidate a popular prior that a typical refinery purchases only one type of crude oil.

Micro Pattern 2. Observed heterogeneity. Refineries' capacity, geographic location, and complexity correlate with their imports: (1) Larger refineries import from a greater number of sources. (2) Distance to a source decreases refineries' imports from that source. (3) More complex refineries import relatively more low-quality crude oil.

[11] Specifically, crude oil is light when its API gravity is higher than 32, and is sweet when its sulfur content is less than 0.5%.

Pattern 2.1 is supported by regressions reported in Table A.4. The likelihood that a refinery imports from a higher number of sources strongly correlates with its capacity size. The strong statistical significance of this correlation is robust to controlling for the location and complexity of refineries.[12]

Table A.5 reports how refineries' capacity, location, and complexity correlate with their quantities of imports from international supplier. Each observation is the volume of imports of a refinery from a source of crude oil including zero import flows. The distance coefficient is highly significant and equals −1.4, where distance is defined between the exact location of a refinery and the capital city of source country. A refinery whose state shares a border with a source imports more from that source —partly reflecting the effect of pipelines from Canada and Mexico. In the table, Type τ is a dummy variable equals one when the traded crude is of type $\tau \in \{L, H\}$, where low-quality type L includes heavy and sour crude, and high-quality H includes the rest. *CI* is complexity index. All else equal, more complex refineries import more low-quality inputs, but the correlation between complexity and imports of high-quality crude is not statistically significant. In other words, holding capacity and location fixed, a more complex refinery imports more, and its larger imports are mainly due to its purchases of low-quality inputs. This evidence confirms that complex refineries have a cost-advantage in refining low-quality crude.

Micro Pattern 3. Unobserved heterogeneity. Capacity, location, and complexity explain "which suppliers refineries purchase from and by how much" to a limited extent.

I compare imports of refineries with similar observable characteristics including location, capacity, complexity. For example, consider a group of refineries that are large and complex, and located in the Gulf Coast. The average number of import origins in this group equals 10.1. I count the number of common origins for every pair of refineries in this group. The average of this number across all pairs in this group equals 5.1; meaning that only half of the trading relationships could be explained by the observables. Appendix A.3.1 reports a set of detailed patterns on differences in the import behavior of observably similar refineries. The above example is representative.

Micro Pattern 4. Capacity and complexity of refineries change slowly, if at all.

I look into annual data between 2008 and 2013. Fig. A.6 shows the distribution of the annual changes of refineries' capacity and complexity. Both distributions have a large mass at zero. There are zero annual changes of capacity in 79.1%, and of complexity in 40.3% of observations (each observation is a refinery-year). Moreover, the annual growth is in the range of (−0.05,0.05) for 90.2% and 85.5% of observations for capacity and complexity, respectively. The average annual growth rates of capacity and complexity across all refineries equal 1.1% and 0.8%, respectively.

Some of the features which I documented here for refineries' sourcing are comparable to those in the manufacturing input sourcing literature. I provide a detailed comparison in Appendix A.3.2.

2.3. From the empirical patterns to the model

As for country-level patterns, to fit in with pattern 1, *a conditional gravity holds in international oil trade data*, the model is designed to generate a gravity equation conditional on refineries' selections. Since key features of refined oil trade is in line with trade in manufactured products (macro pattern 2), I model refined oil trade using standard assumptions in the literature.

As for refinery-level observations, motivated by patterns 1 and 2.1, *refineries typically import from more than one origin, larger refineries import from a higher number of origins*, I model the refiner's problem as a trade-off between gains from purchasing a greater number of suppliers against fixed costs per supplier. The model parsimoniously incorporates variations of trade costs (Section 2.1) as a shorthand for complex logistical frictions in crude oil sourcing. These logistical frictions make room for gains from having access to a broader set of suppliers. Given that the majority of oil trade is based on contracts (Section 2.1), I interpret fixed costs as costs associated with writing contracts and maintaining business relationships with suppliers.

To accommodate patterns 2.2 and 2.3, *distance correlates with trade, complexity correlates with trade of low-quality crude*, the model incorporates transport costs as well as a cost advantage for complex refineries in refining low quality crude.

To explain pattern 3, *differences in the import behavior of refineries after controlling for observables*, I introduce unobserved heterogeneity to variable costs that a refiner faces in importing crude oil from a supplier. This heterogeneity reflects unobserved costs of an input user with respect to a supplier.

Since crude oil is purchased by and large based on annual term contracts (Section 2.1), I take annual observations as the period in which a refinery chooses its suppliers. Motivated by pattern 4, *capacity and complexity change slowly*, I design my framework for a medium run in which refineries' capacity and complexity remain unchanged.

This model is the simplest I could think of that can accommodate the observed patterns. It is designed to be estimated using the available data, to maintain computational tractability in general equilibrium, and to address the questions spelled out in the Introduction. I make no pretense of incorporating all forces at work in oil markets.

3. The model

The world economy consists of N countries indexed by n, and $K+1$ industries: crude oil ($k=0$), refined oil ($k=1$), and non-oil industries ($k=2,\ldots,K$). Each country n is endowed by L_n units of labor, $Q^0_{n\tau}$ units of crude oil of type τ, and a given measure of heterogeneous refineries. A refinery is indexed by φ, and the measure of refineries below φ in country n is denoted by $\Phi_n(\varphi)$. The mass of refineries, $|\Phi_n|$, is the total size of refining industry in n, and so, $\Phi_n(\varphi)/|\Phi_n|$ is the distribution of φ in country n.

In the upstream part of the model, refineries decide which suppliers to select and how much crude oil to buy from each. An individual refinery takes the prices of crude oil inputs and of composite output as given. Aggregating refineries' production determines refined oil supply. The downstream incorporates multiple industries with input-output linkages as well as final consumption. The downstream generates demand for refined oil. In general equilibrium, prices of crude and refined oil as well as other goods are endogenously determined.

3.1. Upstream

3.1.1. Refineries' sourcing and production

I classify suppliers of crude oil by source country and type. Supplier $j=(i,\tau)$ supplies crude oil of type τ from source i. A menu that lists suppliers $j=1,\ldots,J$ is available to all refineries. Let p^0_j denote the price at the location of supplier j. Every refinery φ has a technology that converts crude oil input to a composite refined output. Capacity of refinery φ is denoted by $r(\varphi)$. Utilization rate, $u(\varphi)$, equals the ratio of total crude input use to capacity. The wholesale price index of the composite refinery output in country n is denoted by $\tilde{P}^0_n$.

[12] For location I specifically control for distance to coast and geographic regions defined by the EIA. I do not have access to data on the output mix of individual refineries to directly control for the sale side variables. However, geographic variables and the complexity index serve as reasonable proxies for the output mix. In addition, look at Tables A.11–A.13 in the Appendix A for further evidence on the strong, robust correlation between capacity size and the number of import origins. For instance, within the sample of refineries located at the coastlines, average number of import origins for those above the 66th percentile of capacity is between 8 and 11 depending on complexity, while that is between 3 and 6 for those between the 33th and 66th percentile.

The entire period in which a refinery decides on sourcing and production consists of a continuum of infinitesimal periods $t \in [0,1]$. Let $p_{nj}(\varphi)$ denote the *average cost* per unit of crude oil from supplier j faced by refiner φ in country n. $p_{nj}(\varphi)$ depends on the price of supplier j at origin, p_j^0, as well as transport costs, cost-advantage due to complexity, and some unobserved term. I specify this relationship in Section 4.1. The unit cost of supplier j at t equals $p_{nj}(\varphi)\epsilon_{njt}$, where ϵ reflects variations in transport costs reflecting the daily availability of tankers and limited pipeline capacity. ϵ's are iid, and correlate neither over time nor across space. $1/\epsilon$ follows a Fréchet distribution with dispersion parameter η and location parameter s_ε. Variance of ϵ is governed by η. The higher η, the smaller the variance.[13]

The refiner knows $p_{nj}(\varphi)$ and ϵ_{njt}, and orders crude oil for all t by making contracts with set $S \in \mathcal{S}$ of suppliers (with $\mathcal{S}$ as the power set). The refiner selects supplier $j \in S$ for t, if supplier j is the lowest-cost supplier at t within the selected set S. For making and maintaining a contract with each supplier, the refiner incurs a *fixed cost* $f_n(\varphi)$ which varies across refineries and it is the same across suppliers. In addition, utilizing capacity requires costly refining activity. Refinery φ which operates at utilization rate $u \in [0,1)$, incurs a *utilization cost* of $r(\varphi) \times C_n(u;\varphi)$, where $C_n(u;\varphi)$ is refiner-specific utilization cost per unit of capacity. $C_n(u;\varphi)$ is increasing and convex in u, which I fully specify in Section 4.1. Here, the convexity reflects the capacity constraints.

On the sale side, the refiner enters into a contract with wholesale distributors.[14] The refiner commits to supply its composite output and the distributor commits to pay the price $\tilde{P}_n^0$.

3.1.2. The refiner's problem

Let $P_n^0(S;\varphi)$ denote the average input price of refiner φ if set S of suppliers is selected,

$$\begin{aligned} P_n^0(S;\varphi) &= \int_{\mathbb{R}_+} \dots \int_{\mathbb{R}_+} \left(\min_{j \in S} \left\{ p_{nj}(\varphi)\epsilon_{nj} \right\} \right) dG_\epsilon(\epsilon_{n1}) \dots dG_\epsilon(\epsilon_{nJ}) \\ &= \varsigma^0 \left[\sum_{j \in S} p_{nj}(\varphi)^{-\eta} \right]^{-1/\eta} \end{aligned} \quad (1)$$

where ς^0 is a constant. A larger set S broadens a refiner's access to a wider range of low-cost suppliers, so it lowers the average input costs. This input-saving channel brings about gains from selecting a greater number of suppliers.[15] Total profits equal variable profits net of fixed costs, and variable profits integrate profit flows over the entire period,

$$\begin{aligned} &\tilde{\pi}_n(S,u;\varphi) = \pi_n(S,u;\varphi) - |S|\, f_n(\varphi) \\ &\text{where } \pi_n(S,u;\varphi) = \left(\tilde{P}_n^0 - P_n^0(S;\varphi)\right) r(\varphi) u - r(\varphi) C_n(u;\varphi) \end{aligned} \quad (2)$$

where $|S|$ is the number of elements in S. Refiner φ in country n maximizes total profits by choosing a set $S_n(\varphi)$ of suppliers and utilization rate $u_n(\varphi)$,

$$\max_{S_n(\varphi) \in \mathcal{S}, u_n(\varphi) \in [0,1)} \tilde{\pi}_n(S,u;\varphi)$$

3.1.3. Solution to the refiner's problem

Conditional on sourcing set S and utilization rate u, quantity of crude oil imported from j by refiner φ in n, denoted by $q_{nj}(\varphi)$, is zero if $j \notin S$; and,

$$q_{nj}(\varphi) = k_{nj}(\varphi) r(\varphi) u \quad \text{with} \quad k_{nj}(\varphi) = \frac{p_{nj}(\varphi)^{-\eta}}{\sum_{j \in S} p_{nj}(\varphi)^{-\eta}} \quad \text{if } j = (i,\tau) \in S. \quad (3)$$

Here, $k_{nj}(\varphi)$ is the import share of supplier j, and the dispersion parameter η appears as the elasticity of substitution across suppliers. Conditional on set S of suppliers, the maximization of variable profits delivers the optimal utilization rate,[16]

$$u_n(S;\varphi) = (C_n')^{-1}\left(\tilde{P}_n^0 - P_n^0(S;\varphi);\varphi\right). \quad (4)$$

The variable profit, $\pi_n(S;\varphi)$, equals $r(\varphi)[uC_n'(u;\varphi) - C_n(u;\varphi)]$ evaluated at $u = u_n(S;\varphi)$. In the eyes of refiner φ, two suppliers differ only through their average costs $p_{nj}(\varphi)$. Refiner φ ranks suppliers based on $p_{nj}(\varphi)$, then finds the optimal cut-point on the ladder of suppliers — where adding a new supplier does not cover fixed costs anymore. Hence, the solution to the refiner's problem reduces to finding the number of suppliers rather than searching among all possible combinations. Let $\pi_n^\star(L;\varphi)$ denote the variable profit of φ evaluated at optimal input purchases and utilization rate when selecting the L suppliers with the lowest $p_{nj}(\varphi)$. Optimal total profits and set of suppliers are the solution to

$$\tilde{\pi}_n(\varphi) = \max_{1 \le L \le J} \left[\pi_n^\star(L;\varphi) - Lf_n(\varphi)\right]. \quad (5)$$

3.1.4. Upstream aggregates

For every country n, crude oil revenues, R_n^0, trade quantity of crude oil of type τ from i, $Q_{ni\tau}^0$, aggregate crude oil purchases in quantity Q_n^0 and in value X_n^0, and average crude oil price at the location of refineries, P_n^0, are given by

$$R_n^0 = \sum_\tau p_{n\tau}^0 Q_{n\tau}^0, \quad (6)$$

$$Q_n^0 = \sum_{i,\tau} Q_{ni\tau}^0, \quad \text{with} \quad Q_{ni\tau}^0 = \int q_{ni\tau}(\varphi)\, d\Phi_n(\varphi) \quad (7)$$

$$X_n^0 = \int P_n^0(\varphi) u_n(\varphi) r(\varphi)\, d\Phi_n(\varphi) \quad (8)$$

$$P_n^0 = X_n^0 / Q_n^0 \quad (9)$$

where $q_{ni\tau}(\varphi)$ is given by Eq. (3). Aggregate composite output, $\tilde{Q}_n^0$, is given by

$$\tilde{Q}_n^0 = \int u_n(\varphi) r(\varphi)\, d\Phi_n(\varphi) \quad (10)$$

Similarly, I define aggregate fixed and utilization costs, $\tilde{F}_n^0$ and $\tilde{C}_n^0$. Let $\tilde{R}_n^0 \equiv \tilde{P}_n^0 \tilde{Q}_n^0 - \tilde{F}_n^0 - \tilde{C}_n^0$ be aggregate sales net of fixed and utilization costs. Then, aggregate profits, $\tilde{\Pi}_n^0 = \int \tilde{\pi}_n(\varphi)\, d\Phi_n(\varphi)$, plus total crude oil input costs must equal total supply of refineries' composite output,

$$\tilde{R}_n^0 = \tilde{\Pi}_n^0 + X_n^0 \quad (11)$$

[13] Specifically, $Pr(1/\epsilon \le 1/\epsilon_0) = \exp(-s_\epsilon \epsilon_0^{-\eta})$. The distribution of ϵ under the independence assumption is observationally equivalent to a more general distribution that allows ϵ's to correlate across suppliers, $Pr\left(\frac{1}{\epsilon_1} \le \frac{1}{\epsilon_{01}}, \dots, \frac{1}{\epsilon_J} \le \frac{1}{\epsilon_{0J}}\right) = \exp\left\{-\left[\sum_{j=1}^J (s_\epsilon \epsilon_j^{-\eta})^{1/\rho}\right]^\rho\right\}$, where $\rho \in (0,1]$ is the parameter of correlation. The equivalence holds by reinterpreting η as η/ρ.

[14] Sweeney (2015) provides evidence that 87% of gasoline sales and 83% of distillate sales are at the wholesale market.

[15] In an extreme case where $\eta = \infty$, the cost of each supplier does not vary with ϵ, and so, sourcing collapses to a discrete choice problem. In general, the smaller η, the larger increase in the variable profit from adding a new supplier.

[16] For completeness, there is a corner solution $u(S) = 0$ and $\pi(S) = 0$, when $C'(0) > \tilde{P}^0 - P^0(S)$.

Shipping crude oil between countries is carried out by global shippers who own crude oil transportation resources. By the accounting of crude oil flows, global returns to these transportation resources equal $\sum_n X_n^0 - \sum_n R_n^0$. Every country n receives a fixed share, ω_n, of these returns, with $\sum_n \omega_n = 1$.[17]

3.2. Downstream

Production: In every country i, refinery output is sold domestically at a competitive wholesale market at price $\tilde{P}_i^0$ to a continuum of refined oil wholesalers (indexed by $k = 1$) who supply ready-to-use refined oil products $\omega^1 \in [0,1]$ in international markets to non-oil industries (indexed by $k = 2, \ldots, K$). The production technology of variety $\omega^k \in [0,1]$ in industry $k = 2, \ldots, K$ is a CES combination of refined oil and material-augmented labor —which is a Cobb-Douglas combination of labor and material inputs from all non-oil industries.

The unit cost of ω^k for $k = 1, 2, \ldots, K$ in country i is $c_i^k / v_i^k(\omega^k)$ where $v_i^k(\omega^k)$ is drawn independently from a Fréchet distribution with dispersion parameter θ^k and location parameter A_i^k. For $k = 1$, $c_i^1 = \tilde{P}_i^0$, and for $k = 2, \ldots, K$,

$$c_i^k = \left[\mu_i^{k,o}\left(P_i^1\right)^{1-\rho} + \mu_i^{k,g}\left(\tilde{c}_i^k\right)^{1-\rho}\right]^{\frac{1}{1-\rho}}, \quad \tilde{c}_i^k = (w_i)^{\gamma_i^k}\left(\prod_{k'=2}^{K}\left(P_i^{k'}\right)^{\alpha_i^{kk'}}\right)^{1-\gamma_i^k} \tag{12}$$

Here, P_i^1 is price index of refined oil at the location of consumption in country i. $\tilde{c}_i^k$ is the unit cost of material-augmented labor with γ_i^k as the share of spending on labor, and $\alpha_i^{kk'}(1-\gamma_i^k)$ as the share of spending on industry $k' = 2, \ldots, K$, where $\sum_{k'=2}^{K}\alpha_i^{kk'} = 1$. $\mu_i^{k,o}$ and $\mu_i^{k,g}$ are demand shifters for refined oil and material-augmented labor, respectively. ρ is the elasticity of substitution between refined oil and material-augmented labor. The production technology is Leontief if $\rho \to 0$, and it collapses to Cobb-Douglas if $\rho \to 1$. When the share of refined oil use is small relative to total costs, the price elasticity of expenditure on petroleum is approximately $-\rho$.

Share of spending of industry $k = 2, \ldots, K$ on material-augmented labor and on refined oil are β_n^k and $1 - \beta_n^k$, respectively, where

$$\beta_n^k = \mu_n^{k,g}\left(\frac{\tilde{c}_n^k}{c_n^k}\right)^{1-\rho}, \quad k = 2, \ldots, K \tag{13}$$

Final Consumption: Preferences are CES over refined oil and a Cobb-Douglas bundle of consumption across all non-oil industries. The resulting final price index, P_n^F, is then given by

$$P_n^F = \left[\mu_n^{F,o}\left(P_n^1\right)^{1-\rho} + \mu_n^{F,g}\left(\prod_{k=2}^{K}\left(P_n^k\right)^{\alpha_n^{F,k}}\right)^{1-\rho}\right]^{\frac{1}{1-\rho}} \tag{14}$$

where $\sum_{k=2}^{K}\alpha_n^{F,k} = 1$. The price elasticity of demand for refined oil in final consumption is assumed to be the same as that in production. Share of expenditure of final consumers on oil and all non-oil industries are respectively $1 - \beta_n^F$ and β_n^F, where

$$\beta_n^F = \mu_n^{F,g}\left(\frac{\prod_{k=2}^{K}\left(P_n^k\right)^{\alpha_n^{F,k}}}{P_n^F}\right)^{1-\rho} \tag{15}$$

Trade and Price Indices: The aggregate output of industry k is a CES combination of all varieties within industry k purchased from the international lowest-cost supplier. Trade costs, denoted by d_{ni}^k, are of standard iceberg type, with $d_{ni}^k \geq 1$ and $d_{nn}^k = 1$. If industry k is nontradeable, $d_{ni}^k = \infty$. The price of variety ω^k in destination market n is $c_i^k d_{ni}^k / v_i^k(\omega^k)$ if ω^k is supplied to n from i. Since input users select the lowest-cost supplier, $p_n^k(\omega^k) = \min_i\{c_i^k d_{ni}^k / v_i^k(\omega^k)\}$. Share of trade from i to n in industry k, π_{ni}^k, and price index of industry k in n, P_n^k, are given by

$$\pi_{ni}^k = \frac{A_i^k\left(d_{ni}^k c_i^k\right)^{-\theta^k}}{\sum_{i=1}^{N} A_i^k\left(d_{ni}^k c_i^k\right)^{-\theta^k}}, \quad k = 1, 2, \ldots, K \tag{16}$$

$$P_n^k = \varsigma^k\left[\sum_{i=1}^{N} A_i^k\left(d_{ni}^k c_i^k\right)^{-\theta^k}\right]^{-\frac{1}{\theta^k}}, \quad k = 1, 2, \ldots, K \tag{17}$$

where ς^k is a constant.[18]

Equilibrium: Let R_n^k denote country n's revenues of industry k, and X_n^k denote n's total expenditure on industry k. R_n^k equals the sum of sales across destinations,

$$R_n^k = \sum_{n'=1}^{N} \pi_{n'n}^k X_{n'}^k, \quad k = 1, 2, \ldots, K \tag{18}$$

Wages equal the sum of payments to workers across industries,

$$w_n L_n = \sum_{k=2}^{K} \gamma_n^k \beta_n^k R_n^k \tag{19}$$

Market clearing for crude oil requires global demand for each source-type pair (i,τ) to be equal its supply

$$\sum_{n=1}^{N} Q_{ni\tau}^0 = Q_{i\tau}^0 \tag{20}$$

where $Q_{ni\tau}^0$ is given by (7). Trade deficits associated to crude oil markets are given by

$$D_n^0 = X_n^0 - R_n^0 - \omega_n\left(\sum_n X_n^0 - \sum_n R_n^0\right) \tag{21}$$

In addition, total rents from oil markets in country n is given by the sum of oil revenues, returns received from shipping crude oil, and refineries' profits,

$$\Pi_n = R_n^0 + \omega_n\left(\sum_n X_n^0 - \sum_n R_n^0\right) + \tilde{\Pi}_n^0 \tag{22}$$

At the wholesale market at the gate of refineries in every country n, market clearing requires demand to be equal supply,

[17] Modeling crude oil trade costs as iceberg type is somewhat unsettling given this paper's focus on this market, because that implies the amount of crude oil that is globally processed in refineries will be less than global crude oil supply. This alternative specification preserves the global balance of crude oil flows in terms of quantity, and also guarantees that no dollar spent on crude oil is lost in general equilibrium.

[18] $\varsigma^k = \left[\Gamma\left(\frac{\theta^k + 1 - \eta}{\theta^k}\right)\right]^{\frac{1}{1-\eta}}$

$$\tilde{R}_n^0 = R_n^1 \tag{23}$$

Eq. (23) connects upstream supply to downstream demand. The LHS, $\tilde{R}_n^0$, given by Eq. (11), is the endogenous outcome of aggregation of refineries' production, which it turn depends on their global sourcing in crude oil markets. The RHS, R_n^1, given by Eq. (18) when $k=1$, is total sales of the refining industry, which in turn depends on international demand for refined oil from downstream industries and final consumption.

Market clearing in industries $k = 1, 2, \ldots, K$ requires total expenditure X_n^k to be equal the sum across intermediate use and final consumption of k. For refined oil industry $k=1$,

$$X_n^1 = \sum_{k'=2}^{K} \left(1-\beta_n^{k'}\right) R_n^{k'} + \left(1-\beta_n^F\right) E_n \tag{24}$$

For non-oil industries $k = 2, \ldots, K$,

$$X_n^k = \sum_{k'=2}^{K} \beta_n^{k'} \alpha_n^{k'k} \left(1-\gamma_n^{k'}\right) R_n^{k'} + \alpha_n^{F,k} \beta_n^F E_n \tag{25}$$

Trade deficits, D_n^k, are given by expenditures X_n^k net of sales R_n^k,

$$D_n^k = X_n^k - R_n^k, \quad k = 1, 2, \ldots, K \tag{26}$$

In the Online Appendix I show that Eqs. (18)–(25) guarantee that total expenditure spent by final consumers, E_n, equals the sum of wages, total rents from oil markets, Π_n, and national debt $D_n \equiv \sum_{k=0}^{K} D_n^k$,

$$E_n = w_n L_n + \Pi_n + D_n$$

Eqs. (18)–(26) guarantee that the global trade deficit is zero per industry, $\sum_{n=1}^{N} D_n^k = 0$ for $k = 0, 1, \ldots, K$, which in turn implies that the sum of national trade deficits is zero, $\sum_{n=1}^{N} D_n = 0$.

Definition. Given parameters of endowments $\{L_n, Q_{n\tau}, \Phi_n\}$, production $\{\alpha_i^{kk'}, \mu_i^{k,o}, \mu_i^{k,g}, \gamma_i^k\}$, demand $\{\omega_n, \alpha_n^{F,k}, \mu_n^{F,o}, \mu_n^{F,g}\}$, and trade costs $\{d_{ni}^k\}$, a **general equilibrium** consists of crude oil prices $p_{i\tau}^0$ and wages w_i such that Eqs. (1)–(5) solve the refiner's problem, and Eqs. (6)–(26) hold.

4. Estimation of fixed and variable costs of refineries

I complete the specification of refineries' sourcing in Section 4.1, then present a mapping between observed refinery-level selections and import quantities to unobserved model variables in Section 4.2, which I then use to derive a new estimation procedure that allows me to jointly identify fixed and variable costs of refineries' sourcing in Sections 4.3, 4.4, 4.5.

The estimation procedure that is developed in this section allows me to jointly identify variable and fixed costs of sourcing. Estimating firm-level imports based on a gravity equation requires a timing assumption by which a firm learns about its unobserved components of variable costs, z, only after selecting its suppliers, e.g. Antràs et al. (2017). Under this assumption, one can estimate the trade elasticity using data on quantities independently from selections, and estimate fixed costs without taking into account the heterogeneity in variable costs. In contrast, the estimation procedure derived in this section illustrates that there is information in selections that matters for the identification of the trade elasticity, and information in quantities that matters for the identification of fixed costs.[19]

[19] Alternatively, a number of papers such as Blaum et al. (2018) estimate the trade elasticity using firm-level data on input expenditure shares and revenue. In this approach, unobserved heterogeneity in costs of sourcing is taken into account in the estimation of the trade elasticity since the variable cost heterogeneity is absorbed in input expenditure shares. However, that approach alone does not provide tools to estimate fixed costs.

4.1. Specification

Price of supplier (i, τ) at the location of refiner φ, $p_{ni\tau}(\varphi)$, depends on price at origin $p_{i\tau}^0$, observable trade cost $d_{ni}(\varphi)$, cost-advantage due to complexity $\zeta_{n\tau}(\varphi)$, and an unobserved component $z_{ni\tau}(\varphi)$. Specifically, for refiner φ, for supplier j as a pair of source-type $i\tau$,

$$p_{ni\tau}(\varphi) = \underbrace{p_{i\tau}^0 (1 + d_{ni}(\varphi) + \zeta_{n\tau}(\varphi))}_{\text{observable}} \times \underbrace{z_{ni\tau}(\varphi)}_{\text{unobs.}} \tag{27}$$

By introducing z, the model allows for heterogeneity in variable costs that individual refineries face in importing from suppliers. This heterogeneity embodies different degrees of integration between refineries and suppliers, geopolitical forces, unobserved technology of a refiner to process crude oil of a supplier, and unobserved location of infrastructure such as pipelines.

Observable trade costs are specified as $d_{ni}(\varphi) = (\gamma_i + \gamma_d\, distance_{ni}(\varphi))(\gamma_b)^{border_{ni}(\varphi)}$. Here, γ_i is a source-specific parameter, γ_d is distance coefficient, and γ_b is border coefficient. $distance_{ni}(\varphi)$ is the shortest distance between the capital city of country i and refiner φ. The dummy variable $border_{ni}(\varphi) = 1$ if only if the state in which refiner φ is located shares a common border with country i. Let $j = 0$ refer to the domestic supplier. I normalize the cost of the domestic supplier to its origin price.

Since the majority of heavy crude oil grades are also sour, I use a parsimonious specification with two types of crude oil: low-quality type that includes heavy and sour crude, and high-quality type that includes the rest. The cost-advantage due to complexity $\zeta_{n\tau}(\varphi)$ equals $\beta_0 + \beta_{CI} CI(\varphi)$ if τ is low-quality, and $-\beta_0$ if τ is high-quality. Here, $CI(\varphi)$ is the complexity index of refiner φ.[20]

The unobserved term z, is a realization of random variable Z drawn independently (across pairs of refiner-supplier) from probability distribution G_Z, specified as Fréchet,

$$G_Z(z) = \exp\left(-s_z \times z^{-\theta}\right),$$

with $s_z = [\Gamma(1 - 1/\theta)]^{-\theta}$, where Γ is the gamma function. The normalization ensures that the mean of z equals one. In addition, for the domestic supplier $j = 0$, by normalization $z_0 = 1$.

I write fixed cost $f_n(\varphi) = \tilde{P}_n^0 f(\varphi)$ to report refiner's total profit in dollar values. Here, $\ln f$ is a random variable drawn independently across refineries from a normal distribution G_f with mean μ_f and standard deviation σ_f.[21] [22]

Refineries consume a mix of refined oil products as their energy requirement and as to facilitate refining crude oil. Since refined oil is itself an input needed to refine crude oil, the unit cost of refining is the price of refinery output. Utilization costs are specified by

$$C_n(u; \varphi) = \tilde{P}_n^0 \frac{u}{\lambda(\varphi)(1-u)}, \tag{28}$$

where $\lambda(\varphi)$ is refinery-specific efficiency. This functional form is a

[20] A negative β_{CI} implies that more complex refineries have a cost-advantage with respect to low-quality crude. I specify ζ_H to be the same across refineries because, as shown in Table A.5, there is no statistical correlation between imports of high-quality crude and complexity of refineries. Since I do not observe which type of domestic crude oil refiners buy, I assume that they buy a composite domestic input with a neutral complexity effect, $\zeta = 0$. Lastly, I normalize β_0 such that for the most complex refinery, refining the high-quality crude is as costly as the low-quality crude. $1 - \beta_0 = 1 + \beta_0 + \beta_{CI} CI^{\max} \Rightarrow \beta_0 = -\beta_{CI} CI^{\max}/2$.

[21] Since all refineries in the sample buy domestic crude, I assume a refiners does not pay a fixed cost for its domestic purchase.

[22] The previous literature has typically assumed that fixed costs do not vary across suppliers due to dimensionality issues. The main exception is Antràs et al. (2017) who incorporate fixed cost heterogeneity. Their methodology deals with heterogeneous fixed costs in the case that variable profit rises by increasing margins from adding new suppliers. Since my model features decreasing margins (see the Online Appendix), that methodology is not applicable here.

simple adaptation of increasing and convex utilization costs estimated elsewhere.[23] $\ln\lambda$ is a realization of a random variable drawn independently across refineries from a normal distribution $G_{\lambda,\,n}$ with mean $\mu_{\lambda,\,n}$ and standard deviation $\sigma_{\lambda,\,n}$.

To summarize, a refiner φ is characterized by capacity r, cost-advantage of complexity ζ, and transport costs $d=(d_j)_{j=1}^J$ that are *observables*; and unobserved variable costs $z=(z_j)_{j=1}^J$, efficiency λ, and fixed costs f that are *unobservables* to an econometrician.

4.2. Mapping between observed trade and unobservables

Handling interdependent decisions for selecting among suppliers in firm-level models has proved challenging. This interdependency arises as selected suppliers jointly contribute to the marginal cost of a firm (here, refiner). For example, if the price of a supplier rises, not only an input-user may drop that supplier but also may add new suppliers or purchase more from its existing suppliers.

In this subsection, I connect the model predictions on selections and quantities to the observed trade data. Holding a refinery fixed, I drop subscripts φ and n. For a refinery, I map the *observed* vector of purchased quantities q to *unobserved* cost shocks z, efficiency λ, and fixed cost f. I then use this mapping to derive a tractable likelihood function that allows for an all-in-one estimation of refineries' purchased quantities and selection decisions.

The mapping between q and (z,λ,f) has two parts. Holding the refinery fixed, let set A consist of the observed selected suppliers, and B consist of the unselected ones. The first part maps purchased quantities from selected suppliers q_A to cost shocks of selected suppliers z_A and efficiency λ. The second part of the mapping determines thresholds on cost shocks of unselected suppliers z_B and fixed cost f to ensure that the observed set S of suppliers is optimal. I summarize the mapping in Proposition 1.

I recover z_A as the solution to a system of $|S|-1$ equations and $|S|-1$ unknowns, where the equations are the model predictions of nonzero trade shares and the unknowns are $|S|-1$ entries in z_A. While trade shares pin down z_A, the refiners' total crude oil demand pins down efficiency λ. The second part of the mapping provides necessary and sufficient conditions for the optimality of observed set S of suppliers. The optimal number of suppliers is where adding new suppliers and dropping existing ones decrease profits. The former implies that the costs of unselected suppliers are large enough, and the latter implies that fixed costs are small enough. If the cost of a supplier was too low, the refiner would add that supplier. If the fixed costs were too large, the refiner would drop its existing suppliers.

Proposition 1. The mapping between observed quantities, q, and unobservables (cost shocks z, efficiency λ, and fixed cost f) is as follows.

- Observed purchased quantities from selected suppliers, q_A, map to cost shocks of selected suppliers and efficiency, $[z_A,\lambda]$, according to a one-to-one function $h(q_A)=[z_A,\lambda]$.
- Observed selections are optimal if and only if cost shocks of unselected suppliers, z_B, are larger than a lower bound $\underline{z}_B$, and the draw of fixed cost, f, is smaller than an upper bound $\overline{f}$.

In Appendix B.1.2 I provide a guideline to construct h, $\underline{z}_B$, and $\overline{f}$ in closed form.

[23] Sweeney (2015) estimates utilization costs using a piecewise linear specification. He finds that these costs are much less steep at low utilization rates, and much steeper near the capacity bottleneck. The functional form that I use features the same shape.

4.3. The likelihood function

Summary of parameters and data. I classify the vector of parameters, Ω, into six groups: (i) trade elasticity η; (ii) observed part of trade costs, $\gamma=[\{\gamma_i\}_{i=1}^I,\gamma_d,\gamma_b]$; (iii) dispersion parameter of Fréchet distribution for trade cost shocks, θ; (iv) complexity coefficient, β_{CI}; (v) parameters of log-normal distribution for efficiency, $(\mu_\lambda,\sigma_\lambda)$; and (vi) parameters of log-normal distribution for fixed costs, (μ_f,σ_f). Let $\mathbf{D}=[\mathbf{D}(\varphi)]$ for all φ denote the observed data, where

$$\mathbf{D}(\varphi)=\left[\left(p_j^0\right)_{j=1}^J,\tilde{P}^0,r(\varphi),I^d(\varphi),CI(\varphi)\right].$$

where I^d is data on distance and common border.[24]

Let $L_\varphi(\Omega|\mathbf{D}(\varphi),q(\varphi))$ denote the likelihood contribution of refiner φ, as a function of the vector of parameters Ω, given exogenous data $\mathbf{D}(\varphi)$ and dependent variable $q(\varphi)$.[25] As there is no strategic competition, the whole likelihood, is given by:

$$\prod_\varphi L_\varphi(\Omega|\mathbf{D}(\varphi),q(\varphi)).$$

Proposition 2. The contribution of a refiner to the likelihood function equals

$$\underbrace{J(\lambda,z_A)g_\lambda(\lambda)\prod_{j\in S}g_Z(z_j)}_{L_A,\text{purchased quantities}}\times\underbrace{\int_0^{\overline{f}(\lambda,z_A)}\ell_B(\lambda,z_A,f)\,dG_F(f)}_{L_B,\text{selections}} \tag{29}$$

where $\ell_B=Pr\{z_B\ge\underline{z}_B(\lambda,z_A,f)\}$, and $[\lambda,z_A]$, $\underline{z}_B$, $\overline{f}$ are given by Proposition 1. The Jacobian, $J(\lambda,z_A)$, is the absolute value of the determinant of the $|S|\times|S|$ matrix of partial derivatives of the elements of $[\lambda,z_A]$ with respect to the elements of q_A.

Here, L_A is the probability density of purchased quantities from selected suppliers, as the density of efficiency λ times that of cost shocks of selected suppliers z_A, corrected by a Jacobian for the nonlinear relation between q_A and $[\lambda,z_A]$. L_B is the probability of selecting set S of suppliers among all other possibilities.[26]

This likelihood function highlights the interdependency of the identification of quantities and selections. It takes into account that a refiner not only buys less from its higher-cost suppliers, but also selects them with a lower probability from an econometrician's point of view. In addition, it is an easy-to-compute integral with respect to the draw of f where the integrand $\ell_B(\lambda,z_A,f)$ has a closed-form solution. Calculating the likelihood function without using Proposition 1 involves high-dimensional integrals that demand nearly infeasible computational burden.

I estimate the model, in what I call the "all-in-one" estimation, by minimizing the whole likelihood function with respect to $\Omega\equiv(\eta,\theta,\gamma,\beta_{CI},$

[24] I observe no domestic purchases of crude oil for twelve refineries possibly due to imperfect data gathering. To avoid potential measurement errors, I restrict the sample to the remaining ones which I use to estimate my model.

[25] Refer to a random variable by a capital letter, such as Q; its realization by the same letter in lowercase, such as q; and its p.d.f. by g_Q. The likelihood contribution of refiner φ, L_φ, is given by

$$\begin{aligned}L_\varphi(\Omega|\mathbf{D}(\varphi),q(\varphi))&\equiv g_{Q_A}(q_A(\varphi)\mid S\text{ is selected};\Omega,\mathbf{D}(\varphi))\times Pr(S\text{ is selected}\mid\Omega,\mathbf{D}(\varphi))\\&=g_{Q_A}(q_A(\varphi)\mid Q_A(\varphi)>0,Q_B(\varphi)=0;\Omega,\mathbf{D}(\varphi))\times Pr(Q_A(\varphi)>0,Q_B=0\mid\Omega,\mathbf{D}(\varphi)),\end{aligned}$$

where by construction, $q(\varphi)\equiv[q_A(\varphi),q_B(\varphi)]$ with $q_A(\varphi)>0$ and $q_B(\varphi)=0$.

[26] Three points are worth-mentioning. (i) In the data, a refiner never buys from all suppliers. However, for the sake of completeness, in a corner case where a refiner buys from all, define $L_B=1$. (ii) Since in the data, all refiners buy from the domestic supplier, I have assumed no fixed cost with respect to the domestic supplier. So, the likelihood always contains the density probability of λ. (iii) For buyers who buy only domestically, $\overline{f}=\infty$. In this case, we can infer no information from dropping a supplier simply because no foreign supplier is selected.

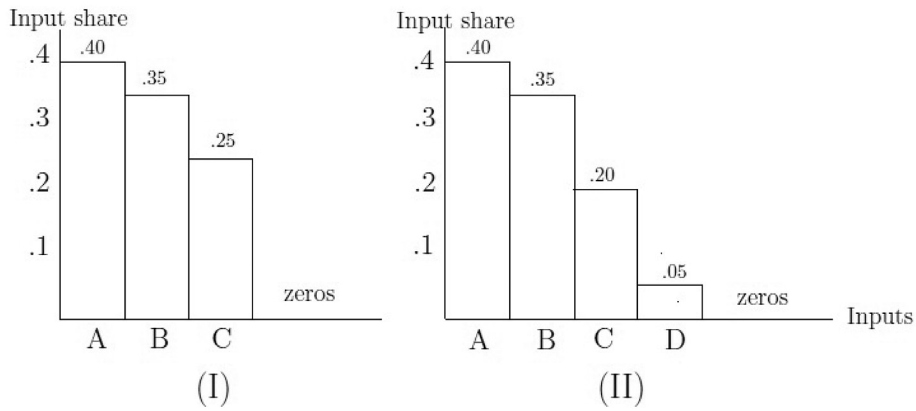

Fig. 2. Dependence of identification of fixed costs on import quantities.

$\mu_\lambda, \sigma_\lambda, \mu_f, \sigma_f$). This is the main estimation whose results I will use in Sections 5–6 for quantitative exercises.

To illustrate the interdependence in the identification of parameters, I estimate the model also in two sequential stages. First, in what I call "quantities only", I estimate $(\eta, \theta, \gamma, \beta_{CI}, \mu_\lambda, \sigma_\lambda)$ by minimizing L_A. The resulting estimates are hence obtained from data on quantities only. These estimates can be thought of as the maximum likelihood counterpart of a firm-level gravity equation using nonzero flows of trade. Then, taking these parameter estimates from "quantities only" as given, in what I call "selections only" I minimize L_B with respect to fixed cost parameters (μ_f, σ_f). The resulting fixed cost estimates will be then obtained from data on zero-one selections only. Comparing "all-in-one" estimation results with those of "quantities only" and "selections only" sheds light on the importance of the joint identification of fixed costs and other parameters such as the trade elasticity η.

I turn into reporting the estimation results after a brief discussion about identification.

4.4. Identification

I briefly discuss the intuition behind the identification of model parameters.

Fixed costs. The sparse patterns of sourcing could be justified by either (large diversification gains, large fixed costs) or (small diversification gains, small fixed costs). These two combinations, however, have different implications. In particular, larger gains from diversification (for example, reflecting by a smaller trade elasticity η) implies more scope for gains from trade. Using an example, I explain what variation in the data identifies the right combination.

Suppose that a refiner ranks suppliers as A, B, C, D, E, etc. with A as the supplier with the lowest cost. Fig. 2 illustrates two cases. In case (I), the refiner buys from suppliers A, B, and C. In case (II), the refiner buys less from supplier C while he adds supplier D. In case (II), the share of D is rather small, equal to 0.05. The larger the share of D, the larger the value it adds to the variable profit. In this example, a relatively small share of D implies that selecting D adds a relatively small value to the variable profit. As D is selected despite its small added gain, the fixed cost of adding D should be also small. So, in case (II) compared with case (I), both the diversification gains and fixed costs are smaller.

Trade elasticity. Holding a refiner fixed, the cost of supplier j can be written as $p_j = p_j^{obs} z_j$, where p_j^{obs} is the observable part of the cost, and z_j is the unobserved part (which is normalized to one for the domestic supplier, $j = 0$). Eq. (3) implies:

$$\ln\frac{q_j}{q_0} = -\eta \ \ln\frac{p_j^{obs}}{p_0^{obs}} - \eta \ \ln z_j, \quad \text{if } \ j \in S.$$

According to this relationship, the slope of $\ln(p_j^{obs}/p_0^{obs})$ identifies η if relative observed price $\ln(p_j^{obs}/p_0^{obs})$ is uncorrelated with unobserved cost $\ln z_j$. This condition is likely to be violated because a refiner is more likely to select supplier j when z_j is smaller. As a result, estimating η according to the above equation creates a sample selection bias. My estimation procedure corrects for this bias by using information on the entire space of cost shocks z's.[27]

Here, I briefly discuss about the direction of the bias when using quantities only. Consider the intuitive case in which unobserved costs tend to be smaller for selected suppliers, and import quantities from selected suppliers tend to be larger than hypothetical import quantities from unselected suppliers if they had been selected. This decreases the variation of quantities in the sample of selected suppliers compared to the variation that is implied for the entire sample (quantities are actual for selected suppliers and latent for unselected suppliers). Provided that variations in prices within the selected sample remain large enough compared to that in the entire sample, the resulting lower variations in quantities within the selected sample will then imply a lower extent to which quantities would react to prices, hence an underestimation of η. Intuitively, a higher η helps explain zeros by implying lower import quantities for unselected suppliers if they had been selected. As a result, excluding zeros is likely to imply less sensitivity of imports with respect to higher-cost suppliers. I provide a more detailed discussion in the Online Appendix.[28]

[27] Two comments come in order regarding the sources of bias other than the selection issue. First, since I observe prices at the location of supplier, I likely avoid the usual endogeneity issue in the related literature. For example, similar to that in the original Eaton and Kortum (2002), consider a regression of log trade flow against gravity variables such as log distance, log exporter wage, and measures of exporter productivity such as human capital and R & D. In such a regression the coefficient on log wage is the negative of trade elasticity. However, because unobserved productivity is expected to be positively correlated with wage, the OLS is likely to underestimate the trade elasticity. Using prices help me avoid this issue. Second, a correlation between relative observed prices $p_j^{obs}/p_{j'}^{obs}$ and relative unobserved costs $z_j/z_{j'}$ requires that observably more costly suppliers be also unobservably more costly. This is the case if, for example, a supplier that is relatively more remote or it has crude oil of higher quality, also ships out its crude oil using unobserved transportation resources that create larger costs in a systematic way. Intuitively, it seems hard to make a non-coincidental case for such correlations.

[28] Generally, the direction of the bias depends on the correlation between relative quantities and relative observed prices within the range of selected suppliers. Although it is intuitive that the inclusion of zeros is likely to imply a higher sensitivity of imports with respect to prices, it is not necessary for the assumptions implicit in my discussion to hold in any other data set. As a result, I cannot rule out the possibility that the direction of the bias in other data sets may go the other way.

Heterogeneity of variable costs. Parameter θ governs the degree of heterogeneity in variable trade costs. In the absence of this heterogeneity, the model predicts the same trade shares for refineries with the same observable characteristics. The more heterogeneity in trade shares conditional on observables, the larger the variance of z, the smaller θ.

To interpret parameter θ, note that besides unobserved location of trading infrastructure, z captures other unobserved buyer-seller relations such as unobserved technological advantage of a refinery in processing crude oil of a supplier, or factors related to geopolitical forces.

Efficiency of utilization costs. Refinery utilization rate governs total use of crude. A higher efficiency λ increases total refinery demand, hence utilization rate. Thus, the distribution of unobserved λ partly reflects the distribution of observed utilization rates. The efficiency λ also captures cross-sectional differences within a country which are not explicitly modeled, such as scheduled maintenance or local demand conditions.[29]

In addition, I conduct a Monte Carlo analysis described in the Online Appendix. A key finding is that my estimation procedure is capable of recovering parameters with standard errors similar to those of the main estimation results which I report below.

Table 3
Estimation results.

Description	Parameter	All-in-one	Quantities only	Selections only
Trade elasticity	η	19.77	10.92	
		(2.74)	(2.20)	
Dispersion in trade costs	θ	3.16	5.10	
		(0.31)	(1.06)	
Distance coefficient	γ_d	0.020	−0.017	
		(0.007)	(0.018)	
Border coefficient	γ_b	0.72	0.60	
		(0.05)	(0.22)	
Complexity coefficient	β_{CI}	−0.028	−0.005	
		(0.004)	(0.009)	
Mean of lnλ	μ_λ	5.45	5.36	
		(0.14)	(0.14)	
Standard deviation of lnλ	σ_λ	1.37	1.38	
		(0.10)	(0.12)	
Mean of ln$f$	μ_f	4.13		5.86
		(0.40)		(0.34)
Standard deviation of ln$f$	σ_f	1.99		2.76
		(0.26)		(0.22)
Log-likelihood		−6513.7	−5419.2	−4216.8

Notes: Bootstrap standard errors in parentheses. See the definition of all-in-one estimation and independent estimations of "quantities only" & "selections only" in Section 4.3.

4.5. Estimation results

Table 3 in column "all-in-one" reports the main estimation results. Standard errors are shown in parenthesis. In addition, I report the results based on independent estimations of "quantities only" and "selection only" as defined in Section 4.3.

The all-in-one estimation delivers a relatively high trade elasticity and small fixed costs. The trade elasticity, $\eta = 19.77$, is greater than the estimates for manufactured products, while it is in the range of oil elasticities in the literature.[30] The ratio of fixed costs paid by a refinery relative to its total profit, on average, equals 3.1%.

The estimates imply large unconditional trade costs but small conditional ones. I begin with *unconditional* trade costs as those for the entire sample of zero and positive trade. If the origin price of crude oil is \$100/bbl, every 1000 km adds on average \$2/bbl to unconditional trade costs. If the state where the refinery is located shares a border with a supplier, either from Canada or Mexico, trade costs reduce by 28%. In addition, the complexity parameter β_{CI} is negative as expected. The source-specific estimates of trade costs range from 0.86 to 1.33 (see Table A.6). Putting these together, *unconditional* prices increase by more than 100% from origins to refineries. In addition, $\theta = 3.16$ implies that the variance of z equals 0.38, which I interpret as the variance of unconditional trade costs if all refineries were observably the same. Conditional trade costs are those for the sample of nonzero import flows. The median of conditional costs equals 0.17 which is less than one sixth of the unconditional size.[31] If I estimate trade quantities independently from selections, then the trade elasticity is half −10.92 compared to 19.77; and fixed costs are 5.6 times larger at the median —exp(5.86) compared to exp(4.13). Moreover, the distance coefficient has the wrong sign and loses its statistical significance (see the 3rd row of Table 3). Besides, the source-specific parameters of variable trade costs are sizably smaller (see Table A.6).[32]

Model fit & implications of the estimates. I simulate the model of refineries' sourcing and production at the parameter estimates. Specifically, I draw (z,λ,f) for each observable (R,ζ,d) for a large number of times. Each (z,λ,f,R,ζ,d) represents a refiner for which I solve its problem. Then I calculate the average outcome of the industry. In addition, I perform the same exercise using the results of the independent estimations of quantities only and selections only. Note that in this exercise, origin prices of crude oil and prices of refinery composite output are kept constant in order to clearly illustrate the role of parameter estimates.

The simulation sets the average cost of crude oil at the location of refineries, P^0_{USA}, described by Eq. (9), at 73.4 \$/bbl, and at 75.5 \$/bbl once the cost advantage due to complexity is netted out. I do not observe these figures in my dataset – I only observe prices at origin and quantities at destinations. However, the EIA reports that the average landed cost of crude oil was 76.7\$/bbl in 2010. Although this figure might be calculated from price data that are not necessarily consistent with prices I use based on Bloomberg, the small gap between 75.5 and 76.7 is reassuring. See Table 4.

In addition, by sourcing globally (baseline simulation) compared with buying only domestically (in which a refinery is not allowed to import) the complexity-adjusted input cost for a typical refinery falls by 8.2%, accompanied by 56.3% increase in profits. Here, the increase in profits is associated with 47.1% increase in the profit margin while only 4.1% increase in production. The latter number is smaller due to capacity constraints. See Table 5.

In comparison, model predictions based on independent estimations of "quantities only" and "selections only" are not satisfying. The average input cost in this case stays below 60\$/bbl, which is too low, relative to comparable data. Here, sourcing globally relative to only domestically, for a typical refinery, brings about input costs that are 26.8% higher, and profits that are 183.8% larger.

[29] The model can be extended to incorporate these features, but note that data on maintenance and sales or output prices at the level of individual refineries are not available to me. The most disaggregated data on wholesale prices of refined products are publicly available at twelve refining regions defined by the EIA. In addition to my benchmark estimation, I have estimated the model using these wholesale prices. The result is that the estimates of log mean and log variance of λ slightly change, and other parameter estimates virtually do not change.

[30] For example, Broda and Weinstein (2006) report that the median elasticity of substitution for 10-digit HTS codes is less than four, but they find the elasticity of substitution for crude oil to be 17.1 in 1972–1988 and 22.1 in 1990–2001. Soderbery (2015) estimates elasticity of heavy crude oil to be 16.2. However, the estimations in Broda and Weinstein and Soderbery are different from mine in a number of ways. They directly use c.i.f. unit costs for homogeneous consumers using the sample of nonzero imports. In contrast, I use firm-level data; since I know only f.o.b. prices I estimate trade costs; my sample includes not only imports but domestic purchases; and importantly my estimation uses the sparsity of trade matrix.

[31] Notice that this value is still larger than what refiners pay for trade costs, because a refiner purchases from a selected supplier j only in the fraction of times when j is its lowest-cost supplier.

[32] The only parameters with modest change are μ_λ and σ_λ which govern the scale and variation of total input demand.

Table 4
Predicted average input costs in the industry (dollars per bbl, 2010).

	Average input cost (adjusted for complexity)	Average input cost (not adjusted for complexity)
All-in-one estimation	73.4	75.5
Independent estimations	58.5	59.7

Notes: Reported figures are model predictions of input costs in the US refining industry according to Eqs. (1) and (9), based on all-in-one estimation and independent estimations of "quantities only" & "selections only" as explained in Section 4.3.

Table 5
Predicted percentage changes to costs, profits, and production in the US refining industry in case of only domestic sourcing compared to the baseline of global souring.

	Input costs	Profits	Profit Margin	Production
All-in-one estimation	−8.2	56.3	47.1	4.1
Separate estimation	−26.8	183.8	153.9	10.0

Notes: Reported figures are model predictions of the average percentage change to the listed variables for an individual refinery when it buys inputs only domestically compared to the baseline in which it buys globally. Input cost is given by Eq. (1), profits by Eq. (2), profit margin by $[\tilde{P}^0 - P^0(\varphi)]/P^0(\varphi)$, and production by $u(\varphi)r(\varphi)$. Results are based on all-in-one estimation and independent estimations of "quantities only" & "selections only" as explained in Section 4.3.

The model predictions, at the main estimates, closely fit the actual distribution of the number of import origins of refineries. The median is 2 in the data and 2 according to the model. The 99th percentile is 14 in the data and 12 according to the model. The estimates from the independent estimations result in an over-stating of the number of origins. See Table 6.

In order to see why the all-in-one estimation has a superior performance compared to independent estimation, note that the trade elasticity η is largely underestimated in "quantities only". Sourcing from a greater number of suppliers benefits a refinery by reducing its input costs. This reduction, as Eq. (5) describes, is governed by η, with a smaller η implying larger gains. With $\eta \approx 11$, the model predicts too much benefits from spreading purchases across suppliers, which subsequently leads to too much reduction in input costs and larger gains from global sourcing.

Table 6
Distribution of number of foreign origins.

	P25	P50	P75	P90	P99
Data	1	2	7	10	14
All-in-one estimation	1	2	4	7	12
Independent estimations	1	4	11	21	30

Notes: Reported figures are data and model predictions of the distribution of the number of import origins across US refineries in 2010. Results are based on all-in-one estimation and independent estimations of "quantities only" & "selections only" as explained in Section 4.3.

5. Calibration & simulation

In this section, I complete the quantification of the model, and explain how I conduct quantitative exercises.

In particular, the method of exact hat algebra, popularized by Dekle et al. (2007), cannot be readily implemented here because refineries' trading relationships endogenously change in response to shocks. Instead, I introduce a hybrid method by which I directly compute variables of crude oil markets in initial and new equilibria, combined with the exact hat algebra implemented for the downstream part of the economy.

5.1. Calibration procedure

I explain the entire task of model parametrization, to the extent required for quantitative exercises, in three steps. These parameters are reported in Table A.7.

5.1.1. Step 1. Estimation

I use the estimates in Section 4, reported in Table 3, for the trade elasticity η, distribution of fixed costs $G_F \sim \log N(\mu_f, \sigma_f)$, distribution of cost shocks G_z~ Fréchet distribution with dispersion parameter θ, and complexity coefficient β_{CI}. λ in country n has a log-normal distribution with mean $\mu_{\lambda,n}$ and standard deviation σ_λ. I use the estimated standard deviation σ_λ, and calibrate $\mu_{\lambda,n}$ below.

5.1.2. Step 2. Literature and auxiliary data

A subset of parameters are taken from auxiliary data or related studies. The distribution of capacity r is specified as a truncated Pareto distribution with shape parameter ϕ over $[r_n^{\min}, r_n^{\max}]$. In line with the smallest refinery size in various countries $r_n^{\min}$ is set at 50′000 b/d, and $r_n^{\max}$ is taken from the Oil and Gas Journal. Using the best fit to U.S. refinery capacity distribution, I set $\phi = .11$.[33] I assume that all refineries within a country has the same complexity index equal to its average in that country. I set the share of countries from crude oil transportation returns, ω_n, proportional to their wages.

A wide range of studies have estimated the elasticity of demand for refined oil products. In their meta-analysis on 97 estimates for gasoline demand, Dahl and Sterner (1991) find a range of 0.22 to 0.31 for short- to medium-run, and a range of 0.80 to 1.01 for long-run elasticities. In another meta-analysis on hundreds of gasoline demand studies, Espey (1998) reports a range of 0 to 1.36 as short- to medium-run averaging 0.26 with a median of 0.23, and a range of 0 to 2.72 for long-run elasticities averaging 0.58 with a median of 0.43. In addition, there has been evidence that at least in the United States, price elasticity of refined oil demand has declined. For example, Hughes et al. (2008) estimated that short- to medium-run gasoline demand elasticity was between 0.21 and 0.34 in 1975–1980, and between 0.03 and 0.08 in 2001–2006. Kilian and Murphy (2014) argue that near zero estimates in the literature could be downward biased due to the endogeneity of oil prices. They estimated the oil demand elasticity at 0.24 for the period 1973–2009. As the benchmark, I set the elasticity of demand for refined oil products $\rho = 0.25$. This value is well in line with the above-mentioned estimates and consistent with the short- or medium-run nature of my equilibrium framework.

In addition to crude oil, every economy has 19 industries consisting of refined oil, other tradeable fuels, agriculture, mining excluding crude oil, 12 manufacturing industries, utilities (electricity, gas, water), transport, and services. For these industries I obtain γ_i^k, $\alpha_i^{k'k}$, α_i^{Fk}, β_i^k, and β_i^F from World Input-Output Database (WIOD), and trade elasticities θ^k from Caliendo and Parro (2015). Table A.8 reports the list of industries as well as their ISIC codes and trade elasticities.

5.1.3. Step 3. Calibration

All parameters listed in Table A.7 are set in steps 1–3 except mean log efficiency $\mu_{\lambda,n}$, variable costs of crude oil d_{ni}^0, efficiency in retail sale of refined oil products, A_n^1, and the scalar ς^0 in crude oil price index at the location of refineries. I calibrate these parameters by matching the model predictions to a set of moments. To do so, I draw a set of realizations independently from a uniform distribution $U[0,1]$. I save these draws and keep them fixed through the calibration process. As I search for the parameters, I use these draws to construct artificial refineries φ in every country n according to measure Φ_n. I solve the refiner's problem for every refiner φ in every country n, aggregate refinery-level to

[33] Specifically, $G_{r,n} = [1-(r/r_n^{\min})^{-\phi}]/[1-(r_n^{\max}/r_n^{\min})^{-\phi}]$. I estimate ϕ using maximum likelihood and data on U.S. refinery capacity.

country-level variables, then match the model to a set of moments, as explained below.

The first set of moments, m^1, consists of *total use* of crude oil n, $m_n^1 \equiv \sum_{i=1}^{N}\sum_{\tau=1}^{2} Q_{ni\tau}^0$ for all n. The second set of moments, m^2, consists of crude oil trade shares, $m_{ni}^2 \equiv (\sum_{\tau=1}^{2} Q_{ni\tau}^0)/m_n^1$ for all n and $i \neq n$. The third set of moments m^3 consists of $m_n^3 \equiv \tilde{P}_n^0/P_n^0$ for all n. The fourth moment is global crude oil trade costs relative to global value of crude oil supply, $m^4 = (\sum_n X_n^0 - \sum_n R_n^0)/\sum_n R_n^0$. These moments sum up to $N^2 + N + 1$ known entries.

The parameters to be calibrated also sum up to $N^2 + N + 1$ unknowns: N for $\mu_{\lambda,n}$, $N^2 - N$ for d_{ni}^0 (by normalization $d_{ii}^0 = 1$), N for A_n^1, and 1 for ς^0. Given all other parameters that are known previously, the four sets of moments $[m_n^1]_n$, $[m_{ni}^2]_{n,i\neq n}$, and $[m_n^3]_n$, and m^4 respectively pin down $[\mu_{\lambda,n}]_n$, $[d_{ni}]_{n,i\neq n}$, $[A_n^1]_{nN}$, and ς^0. Refinery efficiency governs *total* crude oil purchases. All else being equal, a higher $\mu_{\lambda,\,n}$ implies a larger m_n^1. Variable trade costs determine the allocation of aggregate purchases of crude oil by every country across exporters. All else being equal, a larger d_{ni}^0 means a smaller m_{ni}^2. Efficiency in retail sales of refined products governs demand for refinery output in the wholesale market. All else being equal, a higher A_n^1 implies a greater m_n^3.[34] Lastly, a higher ς^0 scales up the payments to trade costs, and so, it governs the fit to m^4. Appendix 10.2 describes the calibration algorithm in details.

5.2. Computing baseline and counterfactual equilibria

5.2.1. Baseline equilibrium

Given expenditure shares, $\gamma_n^k, \alpha_n^{kk'}\beta_n^k, \alpha_n^{Fk'}\beta_n^F, \pi_{ni,\,k}$ for all n, i and $k = 1, \ldots, K$, I use Eqs. (18), (24), (25) and linear algebra to compute X_n^k and R_n^k in the baseline equilibrium. This calculation, in turn, determines the demand from downstream for refined oil, R_n^1. Then, I use an iterative method to compute oil prices in the upstream such that the resulting refined oil supply, $\tilde{R}_n^0$ equals its demand, R_n^1. Specifically, I iteratively solve the refiner's problem for every refiner φ in very country n, aggregate to country-level variables, and update the guess of oil prices until upstream supply equals downstream demand. This simulation delivers baseline equilibrium variables that are fully consistent with my data and parameters.

5.2.2. Counterfactual equilibrium

Consider a shock to endowments, trade costs, or parameters of demand or production. The method of exact hat algebra can not be readily used here because because refiners' selections endogenously change in response to these shocks. For this reason, I compute the counterfactual equilibrium using a "hybrid" method in which I simultaneously solve for upstream variables in *levels* and downstream variables in *changes*. For a generic variable x in the baseline, let x' be its value in the new equilibrium, and $\hat{x} \equiv x'/x$. The algorithm is summarized as follows.

Algorithm. Solve for $(p_{i\tau})'$ and $(w_i)'$ —start with a guess for $(p_{i\tau})'$ and $(w_i)'$.

1. Solve for $(\tilde{P}_i^0)'$ —start with a guess for $(\tilde{P}_i^0)'$.
   (a) Upstream. Solve the refiner's problem for every refinery φ in every country i. By aggregation, calculate supply of refined oil $(\tilde{R}_i^0)'$.
   (b) Downstream. Conditional on upstream aggregates calculated in Step (a), use the *exact hat algebra* to compute demand for refined oil in changes, $\hat{R}_i^1$. Given its baseline value, calculate $(R_i^1)'$.
   (c) Check market clearing condition for refined oil, $(\tilde{R}_i^0)' = (R_i^1)'$. Stop upon convergence; otherwise, update the guess of $(\tilde{P}_i^0)'$ and go to Step (a).
2. Check market clearing condition for crude oil, $\sum_n (Q_{ni\tau}^0)' = Q_{i\tau}^0$, and for labor, $w_n' L_n = \sum_k \gamma_n^k (R_n^k)'$. Stop upon convergence; otherwise, update crude oil prices $(p_{i\tau})'$ and wages $(w_i)'$ and go to Step (1).

In Appendix C I describe the details of this algorithm and other numerical methods used in this paper.

6. Quantitative exercises

The framework –developed in Section 3 and quantified in Sections 4,5– allows me to assess gains to suppliers, refineries, downstream industries, and final consumers from policy changes. Section 6.1 tests out-of-sample predictions of the model for factual changes of crude oil production and refinery capacity. Section 6.2 explores how a shock to crude oil production of a source propagates around the world. Section 6.3 examines the implications of lifting the ban on US crude oil exports. Section 6.4 explores the distributional implications of reductions in crude oil trade costs between net exports and net importers of crude oil. Section 6.5 evaluates the implications of ceasing international trade in this model and compare them to existing benchmarks.

6.1. Worldwide changes to crude oil supply and demand

I test out-of-sample predictions of my framework for the factual changes in crude oil supply and demand from 2010 to 2013. Recall that in my framework, flows of crude oil production $Q_{i\tau}$, and total refining capacity $|\Phi_i|$, are exogenously given. In Sections 4 and 5 I quantified the framework using cross sectional data from 2010. Here, I re-calculate the equilibrium when crude oil production and refining capacity of countries are set to their factual values in 2013.

From 2010 to 2013, U.S. oil crude production grew by 36% or equivalently by 1.98 million b/d. While total production in the rest of the world remained stagnant, its composition slightly changed. Production in Europe, Libya, and Iran declined; and in Canada, and part of the Middle East rose. On the demand side also, refining capacity grew to some extent in Asia. See the Online Appendix for all these changes.

I report in Table 7 model predictions as well as data on changes to oil prices and imports of the US refining industry. Regarding prices, two observations are noteworthy. Between 2010 and 2013, average crude oil prices at the location of suppliers in the US relative to the rest of the world decreased by 3.40%.[35,36] In addition, US prices of refined oil did not perfectly track US prices of crude oil. Specifically, US wholesale price of refined products increased by 2.64% relative to US average price of crude oil.

The model predicts the decline in US/ROW crude oil price ratio at 5.46% compared with 3.40% in the data. Further, the model predicts that US wholesale price of refined products relative to US price of crude oil increases by 4.89% compared with 2.64% in the data. These predictions are in the right direction, and given that they are out of sample targets, their magnitudes are reasonably close to the factual changes. In

[34] Specifically, I target A_n^1 indirectly by modifying the raw data on expenditure and trade shares of refined oil in order to make them consistent with the accounting implied by the model. As equations 13, 15, 16 show, expenditure shares on refined oil, β_n^k and β_n^F, as well as trade shares of refined oil, π_{ni}^1, depend on A_n^1. I modify recorded data on β_n^k, β_n^F, and π_{ni}^1 by finding changes in A_n^1 that result in magnitudes of demand for refined oil consistent with m_n^3.

[35] Average crude oil price at the location of suppliers for a country or region is defined as weighted average free on board prices of crude oil grades in that country or region with weights equal suppliers' output.

[36] Note that well-known benchmark prices, such as West Texas Intermediate (WTI) in the US or Brent in the North Sea in Europe are only one of the crude oil streams in a country or region. In particular, in this period the crude oil price of WTI in Cushing, Oklahoma diverged relative to the crude oil price in a number of other locations within the US. For example, between 2010 and 2013, the price of WTI decreased by 8.8% relative to Brent, but the price of Light Louisiana Sweet or Alaska North Slope decreased only 2–3% relative to Brent. This price separation was likely due to congestions in the pipelines in Oklahoma, e.g. see McRae (2015). In any event, my calculation takes a weighted average of prices, as defined in the previous footnote.

Table 7
Model vs Data — percent change of oil trade and prices related to the United States.

	Crude oil import quantity	# of trading relationships	Total use of crude	U.S. refined to crude price ratio	US/ROW crude price ratio
Data	−16.10%	−15.25%	2.31%	2.64%	−3.40%
Model	−16.37%	−13.88%	2.33%	4.89%	−5.46%

Notes: For the listed variables, row Data reports percentage changes from 2010 to 2013, and row Model reports model predictions of percentage changes for the case in which crude oil production and refining capacity of all countries are set to their 2013 values compared to the baseline in 2010.

addition, the model tightly predicts changes to crude oil imports, number of trading relationships, and total use of crude oil in the US refining industry.

6.2. A boom in U.S. crude oil production

The production of crude oil in the United Sates, due to the shale oil revolution, increased by 36% from 2010 to 2013. Implications of this boom for oil prices and trade have been at the center of the conversation on energy policy in the US and at the global scale. Here, I consider a counterfactual world where only US production would change by 36% from 2010 to 2013 (or equivalently by 1.98 million b/d), then study how this boom would propagate across the world geography.

Table 8 reports model predictions of changes to crude oil price index at refineries P_n^0 given by Eq. (9), average utilization rate U_n^0, refining profits $\tilde{\Pi}_n^0$, and price indices of refined oil P_n^1, chemicals P_n^{chem}, electrical equipments P_n^{elect} given by Eq. (17), and of final goods P_n^F based on Eq. (14). Among manufacturing industries, I report prices of chemicals that have the highest average expenditure share on refined oil, and electrical equipments that have the lowest. A number of results stand out.

First, the boom has a regional effect on the prices of crude oil, although these regional effects are modest. Price of crude oil *at source* falls by 13.34% in the U.S. and on average 9.23% in the rest of the world. Price of crude oil *at refinery*, drops by 11.48% in the U.S., 10.8–11.0% in Canada and Mexico, and slightly less in Brazil, Colombia, Venezuela, and countries in the west of Africa; and down to 7.9–8.2% in Singapore, Japan, and parts of Eurasia. Compared with Americas and Africa, countries in Europe, Russia, and parts of Asia are less integrated with the U.S. market.

However, there are no regional effects on refined oil prices. The change in refineries' production depends on the gap between prices of crude and refined oil as well as the initial utilization rate. Refineries' production increases more in countries that initially utilized their capacity at lower rates, because they are not close to the bottleneck of capacity constraints. Since these countries are not necessarily close to the source of the shock (here, the United States) the regional component of the shock disappears in refined oil markets. Azerbaijan, Iraq, and Nigeria whose initial utilization rates are the lowest among all countries (reported in Table 1) exemplify this mechanism.

In addition, prices of refined oil products fall less than prices of crude oil. If global supply of crude oil increases while refineries' capacity remain unchanged, refineries have to operate at higher utilization rates in equilibrium. To have than happen, the price gap between crude and refined oil rises so that refineries afford the higher utilization costs imposed by capacity constraints. Together, profits rise more for refineries that are closer to the location of supply shock (e.g. Canada) or that had lower initial utilization rates (e.g. Nigeria). Note that these important effects are absent in a model in which the refining technology features constant returns to scale, or in which refineries charge a constant markup.

Given that a crude oil price shock transmits to the downstream through refined oil prices, there is little regional variation in the magnitude of the shock itself. However, expenditure shares on refined oil vary both across countries and across downstream industries. This feature, reflected in input-output parameters, matters for the fall in the final price index. In addition, since wages change only negligibly, the effect on real wages is almost exclusively due to prices. Real wages rise for all countries and by 0.17% at the median.

Table 8
Percentage change to selected variables in response to the boom in U.S. crude oil production.

Country	P^0	U^0	P^1	$\tilde{\Pi}^0$	P^{chem}	P^{elect}	P^F
Algeria	−9.83	0.81	−5.94	9.73	−0.39	−0.15	−0.18
Angola	−10.01	2.59	−5.96	13.50	−0.38	−0.14	−0.16
Azerbaijan	−8.94	7.54	−6.03	16.27	−0.42	−0.17	−0.25
Brazil	−10.35	1.17	−5.93	10.96	−0.52	−0.26	−0.31
Canada	−10.80	2.13	−5.94	17.44	−0.57	−0.12	−0.19
China	−9.53	1.11	−5.90	8.98	−0.60	−0.21	−0.24
Colombia	−10.28	3.10	−5.97	15.65	−0.40	−0.15	−0.17
France	−8.40	3.25	−5.98	10.08	−0.38	−0.11	−0.17
Germany	−8.80	2.23	−5.95	6.47	−0.30	−0.12	−0.17
India	−9.64	0.55	−5.92	8.60	−0.62	−0.31	−0.45
Indonesia	−8.97	1.97	−5.93	9.05	−0.18	−0.16	−0.18
Iran	−9.00	0.64	−5.86	6.73	−0.47	−0.17	−0.19
Iraq	−10.03	2.04	−5.89	0.12	−0.41	−0.18	−0.27
Italy	−8.64	4.36	−6.03	8.29	−0.29	−0.12	−0.19
Japan	−8.21	4.45	−6.04	8.05	−0.71	−0.14	−0.21
Kazakhstan	−8.73	3.19	−5.96	8.39	−0.42	−0.16	−0.19
Korea	−9.35	3.42	−5.93	9.06	−1.32	−0.27	−0.45
Kuwait	−9.17	0.49	−5.90	6.89	−0.43	−0.18	−0.28
Libya	−8.94	0.84	−5.95	6.47	−0.39	−0.15	−0.20
Mexico	−10.96	1.77	−5.94	17.55	−0.30	−0.16	−0.33
Netherlands	−9.21	2.54	−5.95	7.93	−0.44	−0.14	−0.17
Nigeria	−10.11	7.84	−6.21	19.47	−0.45	−0.17	−0.20
Norway	−8.73	1.09	−5.94	6.70	−0.35	−0.13	−0.17
Oman	−8.87	0.91	−5.93	6.60	−0.41	−0.14	−0.18
Qatar	−8.47	0.80	−5.90	5.00	−0.38	−0.13	−0.18
Russia	−8.88	0.86	−5.75	6.40	−0.61	−0.34	−0.44
Saudi Arabia	−9.63	0.95	−5.89	8.82	−0.41	−0.16	−0.22
Singapore	−7.91	5.96	−5.94	8.72	−0.46	−0.16	−0.28
Spain	−9.71	3.62	−5.99	10.60	−0.39	−0.12	−0.17
UAE	−8.63	2.36	−5.93	7.93	−0.42	−0.16	−0.19
United Kingdom	−8.73	2.92	−5.96	9.61	−0.21	−0.09	−0.09
United States	−11.48	2.02	−5.94	18.94	−0.46	−0.11	−0.22
Venezuela	−10.14	1.76	−5.91	13.04	−0.41	−0.15	−0.23
RO_America	−9.87	3.18	−5.97	14.93	−0.39	−0.15	−0.18
RO_Europe	−10.21	2.59	−5.96	8.83	−0.32	−0.13	−0.20
RO_Eurasia	−8.02	5.04	−5.93	8.27	−0.47	−0.19	−0.23
RO_Middle East	−8.89	1.00	−5.93	7.09	−0.41	−0.15	−0.18
RO_Africa	−9.88	2.81	−5.97	14.93	−0.40	−0.15	−0.17
RO_Asia & Oceania	−9.85	2.23	−5.94	10.00	−0.23	−0.09	−0.13

Notes: The boom in US crude oil production corresponds to its 36% rise between 2010 and 2013. Crude oil price index at the location of refineries P^0 is given by Eq. (9), price index of refined oil P^1, of chemicals P^{chem}, and of electrical equipment P^{elect} are given by Eq. (17), average utilization rate is $U_n^0 \equiv (\int u_n(\varphi) r(\varphi) d\Phi_n)/(\int r(\varphi) d\Phi_n)$ refining profits are $\tilde{\Pi}_n^0 \equiv \int \tilde{\pi}_n(\varphi) d\Phi_n(\varphi)$, and the final price index is given by Eq. (14).

6.3. Lifting the ban on U.S. crude oil exports

The remarkable boom in U.S. crude oil production stimulated a policy debate about removing the US 40-year-old ban on crude oil exports. As such, there has been much interest in implications of lifting this ban. I specifically ask: Had this ban overturned in 2010, how much would have U.S. oil imports and exports changed from 2010 to 2013? How much would have American suppliers, refineries, and end-users gained?

To perform this experiment, one needs to know the counterfactual trade costs of shipping crude oil from U.S. to every other country. I use the relationship between the crude oil calibrated trade costs and geographic variables to predict the counterfactual costs. See Appendix A.3.3 for details. Let $[d_{n,\,US}^{new}]_{n=1}^{N}$ denote counterfactual trade costs from US to all other countries when the ban is lifted. Consider two cases: (i) when the ban is maintained and U.S. production rises by 36%, and (ii) when the ban is lifted and U.S. production rises by 36%. Table 9 reports changes to selected variables of the US oil industry in case (ii) compared to case (i).

Had the ban been lifted when US production rose from 2010 to 2013, the average US price of crude oil at source would have been higher by 4.05%, US refining industry would have lost 6.97% of its profits. Translating these percentage changes to dollar values, revenues of U.S. crude oil producers would have increased by $7.49 billion and profits of US refineries would have decreased by $7.76 billion. In addition, downstream industries and final consumers would have faced a negligibly higher price of refined oil.[37]

The reason that US crude oil imports increase in case (i) compared to (ii), is that the origin price of crude oil in the US relative to average of that in the rest of the world rises by 3.27%. Given that the elasticity of substitution across suppliers is large, this small increase in the relative domestic price results in a notable increase in imports.[38]

Table 10 reports model predictions for changes to US crude oil exports and imports in cases (i) and (ii). In 2010 as baseline, US exports were negligible (and only to Canada). From 2010 to 2013, US crude oil production rose by 1.98 million b/d. I find that, as a result of the increase in production, exports largely rise when the ban is lifted.

I am given by an opportunity to assess these predictions since in December 2015, the US government removed the 40-year-old ban on US crude oil exports. I compare the model predictions with the most recent available data on US crude oil imports and exports. An important caveat applies here because the experiment is based on a calibration in 2010 where as the ban was lifted in the end of 2015. Between these periods, changes like developments in domestic and global demand conditions are not controlled for in the experiment. In the data, in 2018 compared with 2015, US crude oil production increased by 1.52 million b/d, US crude oil exports and imports respectively increased by 1.54 and 0.40 million b/d, and total use of crude oil rose by 0.37 million b/d. The model prediction of the increase in exports is reasonably close to that in the data. According to factual changes, all the increase in production was channeled to increase exports (Table 11). In the model experiment, around 70% of the increase in production ($1.37/1.98 \approx 0.7$) is shipped out by exports (Table 10).

6.4. Reductions in crude oil trade costs

Lower costs of crude oil trade imply a reallocation of crude oil resources from exporting countries to importing ones. To examine the distributional implications of such reallocation, I consider a counterfactual in which variable trade costs of crude oil decrease by 10% between all crude oil exporters and importers.

For this counterfactual change, I find that crude oil prices at source, on average, rise by 3.7%. One could think of $3.7/10 = 0.37$ as a local average general equilibrium elasticity of crude oil price with respect to variable trade costs of crude oil at the global scale. However, changes to crude oil prices at origin, that is proportional to change in crude oil revenue R^0, vary across countries. Prices of crude oil at source rise relatively more in countries where there is more excess refining capacity nearby. In the extreme case and as an exception, crude oil price at source falls in the United States. Due to the US export ban, American crude oil producers have to sell domestically, but reductions in international trade costs encourage US refineries to import more. Thus, crude oil prices at source fall in the US. In contrast, crude oil prices at source (or crude oil revenues) rise more in Colombia, Mexico, and Venezuela that are close to the US as well as countries in Middle East that are close to India, China, and Europe.

In addition, there are notable distributional impact on refineries across countries. The variation in changes to refined oil prices is modest, and the gap between crude and refined oil prices effectively changes utilization rates and profits of refineries. Refining profits rise on average by 7.0% in countries that do not produce crude oil, and notably fall in countries that are net exporters of crude oil (Table 12).

6.5. Gains from oil trade

I examine gains from trade by simulating counterfactual experiments in which trade in oil and other goods between countries or regions of the world is prohibitive. I then compare my results to the literature on gains from trade.[39]

6.5.1. Gains from trade for the United States

I start with a counterfactual world where trade in all goods between the United States and the rest of the world is prohibitive. Specifically, I raise trade costs of all goods between the U.S. and all other countries to infinity. This autarky is an extreme counterfactual policy, but it provides a benchmark for comparing gains from trade in my framework to typical gains from trade in the literature.

In the U.S. economy, in the autarky compared to the baseline, price index of crude oil increases by 2404.03%, price index of refined oil increases by 1855.79%, profits of refineries drop by 86.73%, and price index of final goods rises by 40.37%. From the baseline to autarky, U.S. real wage (wage divided by price index of final goods) drops by 28.76%. See Table 13.

The price change is notably heterogeneous across industries. In one extreme, the price index of land, air, and water transport increases by 243.20% followed by agriculture, chemicals, and fuels (excluding refined oil) in which the price indices rise by around 100%. In the other extreme, the price of electrical equipment rises by only 16.75%, and in particular, the price of services that account for a large share of final expenditure rises by 25.40%. See Table 14.

I compare my results on gains from trade to the existing literature. As a benchmark to compare, I consider a GE model of trade, with input-output linkages, in which crude oil is part of mining, and refined oil is part of fuels. This is a proper benchmark because it precisely highlights what this paper adds to the literature —crude oil global sourcing by refineries in upstream and its connection to demand for refined oil in downstream. For this benchmark, using WIOD data and trade elasticities from Caliendo and Parro (2015), Costinot and Rodriguez-Clare (2014) report that US gains from trade are 8.0%. In terms of changes to real wage from baseline to autarky, US gains from oil trade are 20.76 percentage point larger in this model than that in the benchmark.

In addition, I find that US crude oil revenues increase by 2151%, which substantially increases the nominal national income. As a result, real GDP per capita falls by 11.22% rather than 28.76%. Note that the model is silent about who receives crude oil revenues. However, since oil revenues are rents on natural resources rather than rewards to labor, the results suggest important distributional gains from trade

[37] These findings are in line with the views by some of the experts on oil markets. For example, see (Kilian, 2016, p. 20): "[…] gasoline and diesel markets have remained integrated with the global economy, even as the global market for crude oil has fragmented. This observation has far-reaching implications for the U.S. economy.".

[38] Note that due to convex costs of capacity utilization and due to fixed costs of imports, the trade elasticity is not the only parameter that governs the pattern of substitution between domestic and foreign suppliers.

[39] If trade in crude and refined oil is prohibitive for a non-producer of crude oil like Germany, the model predicts that the price of crude oil in Germany must be infinity. An infinite price of crude oil results in an infinite price index of final goods. Hence, Germany's gains from oil trade is trivially unbounded. In this section, I focus on less extreme counterfactuals for which my model delivers more informative results.

Table 9
The effects of removing crude oil export restrictions on the U.S. oil industry (percentage change).

Crude oil import quantity	# of trading relationships	Total use of crude oil	Refineries' profits	US refined oil price	US crude oil price
14.30%	9.28%	−0.44%	−6.97%	0.23%	4.05%

Notes: Reported numbers are model predictions of the listed variables in case that (i) U.S. crude oil export ban is lifted and U.S. production rises by 1.98 million b/d corresponding to its growth between 2010 and 2013, compared to the case in which (ii) the ban is maintained and U.S. production rises by the same amount.

Table 10
Model predictions of changes to US exports, imports and use of crude oil, with and without the export ban (million b/d).

	Production	Exports	Imports	Total use
Baseline (2010)	5.47	0.04	10.44	15.88
Change in counterfactual 2013 compared to baseline 2010:				
Ban in place	1.98	0.01	−1.58	0.39
Ban is lifted	1.98	1.37	−0.30	0.33

Notes: Reported numbers are model predictions of U.S. crude oil exports, imports, and total use in case that (i) U.S. crude oil export ban is lifted and U.S. production rises by 1.98 million b/d corresponding to its growth between 2010 and 2013, compared to the case in which (ii) the ban is maintained and U.S. production rises by the same amount.

Table 11
Data on changes to US exports, imports and use of crude oil (million b/d).

	Production	Exports	Imports	Total use
Baseline (2015)	9.43	0.46	7.36	16.33
Change to data in 2018 compared to baseline 2015:				
	1.52	1.54	0.40	0.37

Notes: Reported numbers are factual changes to U.S. crude oil exports, imports, and total use between 2015 and 2018 when U.S. crude oil production rose by 1.52 million b/d. Year 2015 is set as the baseline since U.S. crude oil export ban was lifted in December 2015.

once we depart from a labor theory of trade into one that incorporates natural resources.

6.5.2. Gains from trade for Europe

Consider a counterfactual world where trade between European countries and the rest of the world is prohibitive. Specifically, while I do not change the trade costs between any two European countries, I raise trade costs of all goods between European countries and all non-European countries to infinity.

Across European countries the price index of crude oil at the location of refineries increases by 7279–8592%, price index of refined oil increases by 6597–7200%, and price index of final goods rises by 40.68–92.05%. Price of crude oil at refinery increases more in Italy and Spain because in the baseline these two countries import relatively more from non-European suppliers. In addition, real wages across these countries decrease between 28.92% in the United Kingdom and 51.40% in Italy (Table A.9). This counterfactual is notably less extreme than a complete autarky at the level of individual countries. However, changes to real wages are still large compared to the benchmark. According to Costinot and Rodriguez-Clare, in a standard trade model with multiple industries and input-output linkages, gains from trade for the European countries in my sample range between 16.2% in Italy to 39.80% in the Netherlands.

In addition, I examine industry-level price changes. Differences in price changes across countries within an industry partly reflects parameters of input-output table. For example, although the price of refined oil rises more in Italy than elsewhere, the price index of chemicals and transports rises modesty in Italy relative to elsewhere, or price of agriculture in Norway rises considerably less than elsewhere (Table A.10).

Table 12
Percentage change to selected variables of 10% reduction in variable trade costs of crude oil.

	R^0	P^0	U^0	P^1	$\tilde{\Pi}^0$
Algeria	5.02	5.02	−1.07	0.45	−16.86
Angola	4.67	4.67	−2.86	0.51	−17.37
Azerbaijan	4.89	4.89	−8.96	0.82	−21.97
Brazil	3.16	1.76	−0.61	0.47	−6.91
Canada	3.27	2.50	−0.75	0.45	−7.82
China	0.63	−1.16	0.05	0.44	1.15
Colombia	5.13	5.13	−3.47	0.50	−19.09
France	–	−0.35	1.56	0.40	7.83
Germany	–	−1.05	1.48	0.43	6.78
India	0.22	−0.36	0.05	0.45	1.23
Indonesia	3.82	1.88	−1.55	0.46	−9.05
Iran	4.95	5.20	−1.99	0.51	−31.27
Iraq	4.95	5.18	−11.55	0.82	−40.62
Italy	–	−0.17	1.86	0.40	5.26
Japan	–	−0.58	2.14	0.32	6.53
Kazakhstan	4.35	2.54	−3.66	0.65	−9.95
Korea	–	−1.81	1.85	0.43	7.84
Kuwait	4.30	4.30	−0.65	0.45	−14.26
Libya	4.74	4.74	−1.31	0.46	−16.16
Mexico	4.70	5.01	−1.75	0.45	−17.77
Netherlands	–	−0.33	1.41	0.45	7.27
Nigeria	4.52	4.52	−7.34	0.78	−19.95
Norway	4.31	3.57	−1.47	0.48	−14.27
Oman	5.54	5.54	−1.79	0.47	−19.98
Qatar	5.02	5.02	−1.58	0.46	−17.93
Russia	3.89	4.09	−1.72	0.71	−21.26
Saudi Arabia	4.42	4.65	−2.07	0.49	−26.69
Singapore	–	−0.70	2.94	0.45	7.48
Spain	–	−1.08	1.58	0.40	6.12
UAE	4.55	4.55	−3.70	0.56	−18.35
United Kingdom	3.72	2.50	−2.00	0.49	−6.93
United States	−0.53	−1.20	0.47	0.42	5.82
Venezuela	5.36	5.36	−2.26	0.48	−19.30
RO_America	2.64	1.81	−0.75	0.46	−2.51
RO_Europe	2.13	0.61	−0.46	0.50	−1.34
RO_Eurasia	2.18	1.07	−2.81	0.67	−5.68
RO_Middle East	4.34	3.38	−1.29	0.46	−13.89
RO_Africa	4.29	1.99	−1.15	0.47	−4.73
RO_Asia & Oceania	3.11	0.58	−0.82	0.44	−3.48

Notes: Reported numbers are model predictions of percentage changes to the listed variables from 10% reduction in all international variable trade costs of crude oil. Crude oil revenue R^0 is given by Eq. (6), crude oil price index P^0 by Eq. (9), price index of refined oil P^1 by Eq. (17), average utilization rate is $U_n^0 \equiv (\int u_n(\varphi) r(\varphi) d\Phi_n)/(\int r(\varphi) d\Phi_n)$ refining profits are $\tilde{\Pi}_n^0 \equiv \int \tilde{\pi}_n(\varphi)\, d\Phi_n(\varphi)$.

Discussion. The features in my model that matter for gains from oil trade could be distinguished in connection with the literature that illustrates sources of gains from trade as studied in details in Costinot and Rodriguez-Clare (2014). It is important to note that the elasticity of substitution across oil suppliers is *not* a source of large gains. This elasticity is rather high, meaning that oil from one supplier is highly substitutable for oil from other ones. In contrast, sector-level elasticity of substitution between oil and other inputs is small, meaning that input-users can substitute oil with other inputs only to a small extent. This small elasticity is a major source of large gains in this setting compared to models that

Table 13
percentage change to variables in upstream and final demand in the US due to moving to autarky.

Price index of crude oil, P^0	Price index of refined oil, P^1	Refining profits, $\tilde{\Pi}^0$	Price index of final goods, P^F	Real wage w/P^F
2404.03%	1855.79%	−86.73%	40.37%	−28.76%

Notes: Reported numbers are model predictions of percentage changes to the listed variables in the U.S. when U.S. moves from baseline to auatarky.

Table 14
percentage change to industry-level price index in the US due to moving to autarky.

Fuels	Agriculture	Mining	Food	Textiles	Wood	Paper	Chemicals	Plastics
97.52%	105.29%	39.87%	53.18%	45.54%	45.74%	36.98%	96.48%	69.71%
Mineral	Metals	Machinery	Electrical	Transp. Equip	Other Manuf	Utilities	Transport	Services
46.92%	31.93%	42.48%	16.75%	35.48%	27.24%	23.70%	243.20%	25.40%

Notes: Reported numbers are model predictions of percentage changes to the industry-level price indices, given by Eq. (17), in the U.S. when U.S. moves from baseline to auatarky.

assume Cobb-Douglas production function across oil and other factors of production or intermediate inputs.[40]

For this reason, net importers of crude oil benefit from oil trade as they would face significantly higher oil prices and so higher costs of production in autarky. In addition, countries that are net exporters of crude oil would face a lower oil price in autarky compared to the baseline. Trade in oil can also benefit these countries as oil revenues they receive would fall in autarky. Large oil revenues received by large net exporters in the baseline is mainly due to the global rather than their own domestic demand.[41] This channel, in turn, highlights that prices of oil are largely rents on natural resources rather than marginal costs that are indirectly paid as wages to labor. While labor theories of trade assume that all factor rewards are wages, the introduction of natural resources in the analysis of gains from trade makes room for a discussion about the importance of within-country distribution of rents from exports of natural resources.

7. Conclusion

This paper develops a multi-country general equilibrium framework that incorporates crude oil global sourcing by refineries and refined oil demand by downstream end-users that are connected through input-output linkages and international trade. I derive an estimation procedure that allows for joint identification of fixed and variable costs that refineries face in their procurement of crude oil. I use the estimates to conduct counterfactual exercises. A shock to U.S. crude oil production changes the relative prices of crude oil across countries to a modest degree. As markets of crude oil are not entirely integrated, trade-related policies such as lifting the ban on U.S. crude oil exports can generate important distributional effects between crude oil producers and refineries. Reductions of trade costs in crude oil markets also generate distributional impact between countries that are net importers and net exporters of crude oil. Lastly, gains from oil trade in this model is larger than gains from trade in benchmark models.

The model of refineries' sourcing developed in this paper can be used in other applications in which input users select among available alternatives and purchase nonzero amounts from the selected ones. An important direction for future research is modeling dynamic decisions of crude oil producers to explore and to extract, and of refiners to invest in refinery capacity and complexity. While my framework is designed for a medium-run analysis, these dynamic considerations are key to study long-run outcomes.

[40] To the extent that oil and non-oil energy are substitutable more than what I specify here, results in this section might be thought of as an upper bound on the gains from trade. Accordingly, we could extend this model by incorporating a production structure that allows for more flexible elasticities of substitution between oil and other forms of energy such as natural gas.

[41] The imbalance between demand and supply contributes to gains from trade through two other channels in multi-sector models: scale economies e.g.. Kucheryavyy et al. (2019), or imperfect resource mobility e.g. Galle et al. (2018), or both e.g. Farrokhi and Soderberyi (2020). These two channels are not operative here.

Appendix A. Empirics

A.1. Tables

Table A.1
Country-level variables and their data sources.

Variable	Source
Crude oil production	Energy Information Administration & Oil and Gas Journal
Refining capacity	Energy Information Administration & Oil and Gas Journal
Free on board prices of crude oil	Bloomberg
Refined oil consumption	Energy Information Administration
Capacity of upgrading units	Oil and Gas Journal
Utilization rate	World Oil and Gas Review published by Eni
Input-Output parameters	World Input-Output Table (WIOD)
International trade flows	UN Comtrade, BACI
Gravity variables	CEPII
Population and GDP	Penn World

Note: The WIOD data are reported based on ISIC classification whereas the trade data are based on HS codes. I use a crosswalk between HS and ISIC to obtain international trade flows for the ISIC classification.

Table A.2
Capacity and number of refineries importing from none, one, and more than one origin.

		# of foreign origins		
	Total	0	1	2+
# of refineries	110	25	26	59
capacity share (%)	100	5.6	17.2	77.2

Table A.3
Share of refineries importing types of crude oil, 2010.

Share of importing refiners from			
One type	Two types	Three types	Four types
21.6%	12.4%	29.9%	36.1%

Note: Types are classified to four groups as (light, heavy) × (sweet, sour). A crude oil is light when its API gravity is higher than 32, and is sweet when its sulfur content is less than 0.5%.

Table A.4
Number of import origins vs capacity, geography, and complexity. Results from Poisson maximum likelihood estimation.

Dependent variable: number of import origins, 2010		
	(1)	(2)
Log(capacity)	0.740	0.764
	(0.074)	(0.085)
Distance to coast	−1.424	−1.907
	(0.184)	(0.407)
Complexity index	0.034	0.0410
	(0.017)	(0.019)
PADD-effects	no	yes
# of observations	110	110
Log-likelihood	−189.399	−183.760
Pseudo-R^2	0.498	0.513

Notes: Standard errors are in parenthesis. The results are robust to inclusion of the five Petroleum Administration Defense Districts (PADDs) defined by EIA. For the map of PADDs, see Fig. A.5.

Table A.5
Imports vs capacity, geography, and complexity. Results from Poisson pseudo maximum likelihood estimation.

Dependent variable: Imports of crude oil (bbl/day) from country i of type $\tau \in \{L, H\}$ to refinery φ, possibly zero		
	(1)	(2)
log distance$_{\varphi i}$	−1.389	−2.168
	(0.245)	(0.342)
border$_{\varphi i}$	0.788	0.717
	(0.404)	(0.422)
log (origin price)$_{i\tau}$	−4.681	−4.427
	(2.449)	(2.395)
Type L	−4.514	−4.413
	(1.448)	(1.866)
Type L×log CI$_{\varphi}$	1.449	1.827
	(0.401)	(0.826)

(continued on next page)

Table A.5 (*continued*)

Dependent variable: Imports of crude oil (bbl/day) from country i of type $\tau \in \{L,H\}$ to refinery φ, possibly zero		
	(1)	(2)
Type H×log CI_{φ}	−0.408	–
	(0.501)	
log $capacity_{\varphi}$	1.415	–
	(0.111)	
source FE	yes	yes
refinery FE	no	yes
# of observations	5280	4080
# of nonzero observations	514	514
R^2	0.178	0.239

Notes: Standard errors are in parenthesis. Each observation is a trade flow (possibly zero) from a foreign supplier to an American refiner in year 2010. In column (2), the regression is feasible by keeping observations for only importing refineries, and dropping capacity and either *TypeH* × *log* (*CI*) or *TypeL* × *log* (*CI*).
A Tobit regression delivers similar results in terms of signs and significance of all coefficients. I have reported results from Poisson pseudo maximum likelihood because the trade literature favors it in estimating a gravity-like equation. See Santos Silva and Tenreyro (2006).

Table A.6
Estimation Results —Estimates of γ_i, source-specific parameters of variable trade costs.

Country	All-in-one	Quantities only
Canada	1.08	0.58
	(0.11)	(0.14)
Mexico	1.27	0.24
	(0.14)	(0.15)
Saudi Arabia	0.86	0.58
	(0.12)	(0.21)
Nigeria	0.99	0.32
	(0.15)	(0.26)
Venezuela	1.24	0.27
	(0.18)	(0.17)
Iraq	0.95	0.59
	(0.13)	(0.24)
Colombia	1.11	0.39
	(0.15)	(0.16)
Angola	0.95	0.63
	(0.15)	(0.31)
Russia	0.91	0.51
	(0.14)	(0.19)
Brazil	1.04	0.54
	(0.14)	(0.17)
Every other source	1.33	0.57
	(0.18)	(0.21)

Note: standard errors in parentheses.

Table A.7
List of parameters.

Upstream Parameters —related to industry $k = 0$		
η	Trade elasticity for crude oil	estimation
G_F	distribution of fixed costs, log-normal (μ_f,σ_f)	estimation
$G_{\lambda,\,n}$	distribution of efficiency, log-normal ($\mu_{\lambda,\,n}$,σ_{λ})	estimation & calibration
G_z	distribution of cost shock, Fréchet with disp. Parameter θ	estimation
$G_{r,\,n}$	distribution of capacity r, truncated Pareto with parameter ϕ	data on refineries
β_{CI}	coefficient of complexity index	estimation
d_{ni}^{0}	trade costs of crude oil	calibration
Downstream Parameters —related to industries $k = 1, \ldots, K$		
ρ	elast. of subst. Between refined oil and other inputs	literature
θ^k	trade elasticity in industry k	Caliendo and Parro
$\gamma_i^k, \alpha_i^{k'k}, \alpha_i^{Fk}, \beta_i^k, \beta_i^F$	spending shares by industries and final consumers	WIOD

Table A.8
Industry-level trade elasticities.

ID	Industry	ISIC rev. 3	Trade elasticity, θ^k
1	Refined Oil	section 2320	51.08
2	Tradeable Fuels — Excluding Refined Oil	section 23 excluding 2320	51.08
3	Agriculture, Hunting, Forestry and Fishing	sections A-B	8.11
4	Mining and Quarrying — Excluding Crude Oil	section C excluding C11	15.72
5	Food, Beverages and Tobacco	section 15-16	2.55
6	Textiles and Textile Products, Leather, Leather and Footwear	sections 17-18-19	5.56
7	Wood and Products of Wood and Cork	section 20	10.83
8	Pulp, Paper, Paper, Printing and Publishing	sections 21-22	9.07
9	Chemicals and Chemical Products	section 24	4.75
10	Rubber and Plastics	section 25	1.66
11	Other Non-Metallic Mineral	section 26	2.76
12	Basic Metals and Fabricated Metal	sections 27-28	7.99
13	Machinery, Nec	section 29	1.52
14	Electrical and Optical Equipment	section 30-33	10.60
15	Transport Equipment	section 34-35	5.00
16	Manufacturing, Nec; Recycling	section 36-37	5.00
17	Electricity, Gas and Water Supply	section E	5.00
18	Land, Water, Air Transport	sections 60-61-62-62	5.00
19	Services	section FC-HH	5.00

Notes: Trade elasticities are taken from Caliendo and Parro (2015). They report an elasticity of 1.01 for transport equipment, and set the elasticity to 5 for the industries they do not estimate. I set the elasticity for these industries to 5.

Table A.9
From baseline to autarky between Europe and the rest of the world (percentage change).

	Price index of crude oil, P^0	Price index of refined oil, P^1	Price index of final goods, P^F	Real wage w/P^F
France	7758.29%	6828.43%	55.96%	−48.03%
Germany	7481.74%	6681.73%	68.24%	−47.44%
Italy	8592.30%	7200.51%	61.72%	−51.40%
Netherlands	7496.98%	6720.12%	71.84%	−42.20%
Norway	7641.80%	6597.06%	92.05%	−37.60%
Spain	8499.43%	7031.74%	49.13%	−48.50%
UK	7279.69%	6636.86%	40.68%	−28.92%

Notes: Reported numbers are model predictions of percentage changes to the listed variables in European countries when Europe as a whole moves from baseline to auatarky.

Table A.10
Percentage change to industry-level prices in European countries due to moving Europe as a whole to autarky.

	Fuels	Agriculture	Mining	Food	Textiles	Wood	Paper	Chemicals	Plastics
France	75.14%	120.05%	75.80%	49.42%	35.12%	47.95%	30.25%	162.92%	70.61%
Germany	71.70%	116.56%	46.10%	63.28%	56.69%	65.39%	35.85%	123.71%	109.07%
Italy	71.37%	101.25%	65.16%	43.32%	27.73%	27.74%	23.40%	105.61%	53.39%
Netherlands	82.06%	91.66%	28.11%	62.64%	57.03%	42.03%	36.14%	157.30%	101.79%
Norway	73.76%	81.78%	53.35%	78.47%	83.28%	79.60%	60.45%	165.59%	138.27%
Spain	63.96%	66.58%	90.13%	33.79%	26.07%	51.70%	19.67%	163.96%	65.82%
UK	64.62%	98.77%	28.50%	40.32%	34.61%	39.06%	21.20%	66.39%	55.13%

	Mineral	Metals	Machinery	Electrical	Transp. Equip	Other Manuf	Utilities	Transport	Services
France	61.76%	15.01%	25.90%	22.15%	25.98%	38.19%	10.44%	216.75%	5.58%
Germany	69.41%	28.19%	43.14%	29.26%	36.50%	67.15%	12.67%	209.39%	19.09%
Italy	76.38%	34.28%	26.46%	19.64%	26.92%	25.31%	22.52%	172.99%	22.78%
Netherlands	56.84%	39.35%	59.27%	46.17%	45.66%	48.41%	21.46%	235.28%	34.10%
Norway	113.93%	53.91%	81.94%	44.69%	61.65%	79.01%	62.24%	278.67%	60.71%
Spain	62.30%	32.39%	25.55%	21.91%	24.90%	22.38%	35.26%	260.58%	3.79%
UK	44.99%	37.73%	44.88%	26.95%	31.21%	42.76%	16.27%	86.06%	18.15%

Notes: Reported numbers are model predictions of percentage changes to the industry-level price indices, given by Eq. (17), in European countries when Europe as a whole moves from baseline to auatarky.

A.2. Figures

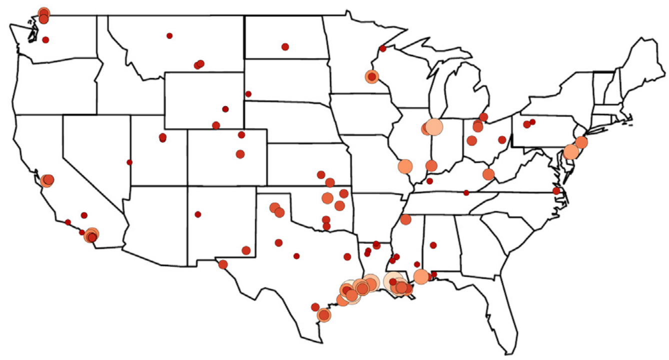

Fig. A.1. U.S. refineries and capacity, 2010. Diameter of circles is proportional to capacity size. For visibility of smaller refineries, the smaller capacity size, the darker it is.

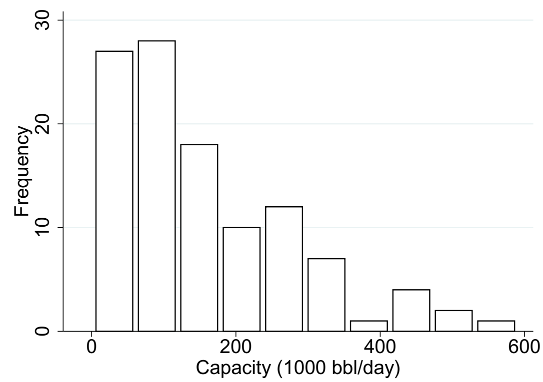

Fig. A.2. Distribution of refinery capacity in the U.S. refining industry, 2010.

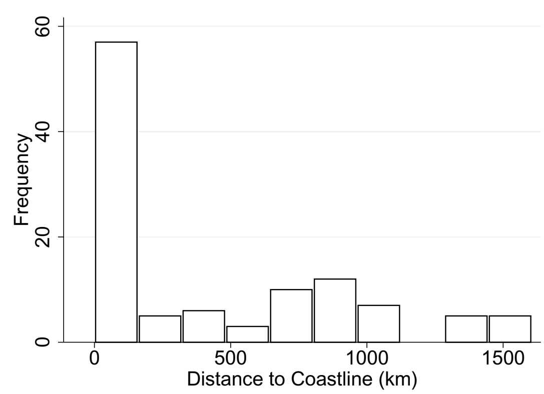

Fig. A.3. Distribution of refinery distance to coastline for the U.S refineries, 2010. Distance to coastline is defined as the distance between location of a refinery to the closest port in the U.S.

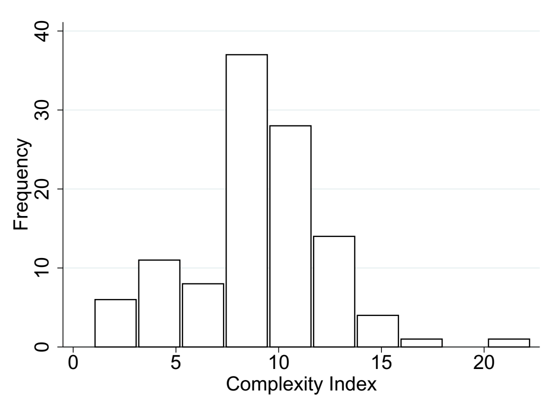

Fig. A.4. Distribution of complexity index in the U.S. refining industry, 2010.

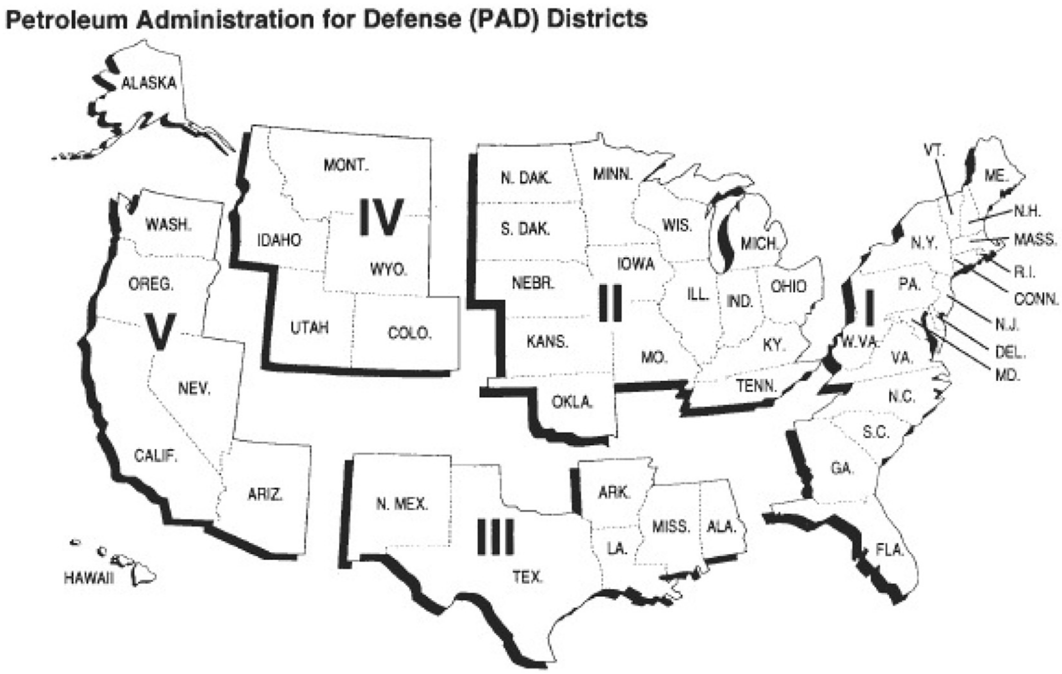

Fig. A.5. Petroleum Administration for Defense Districts (PADD).

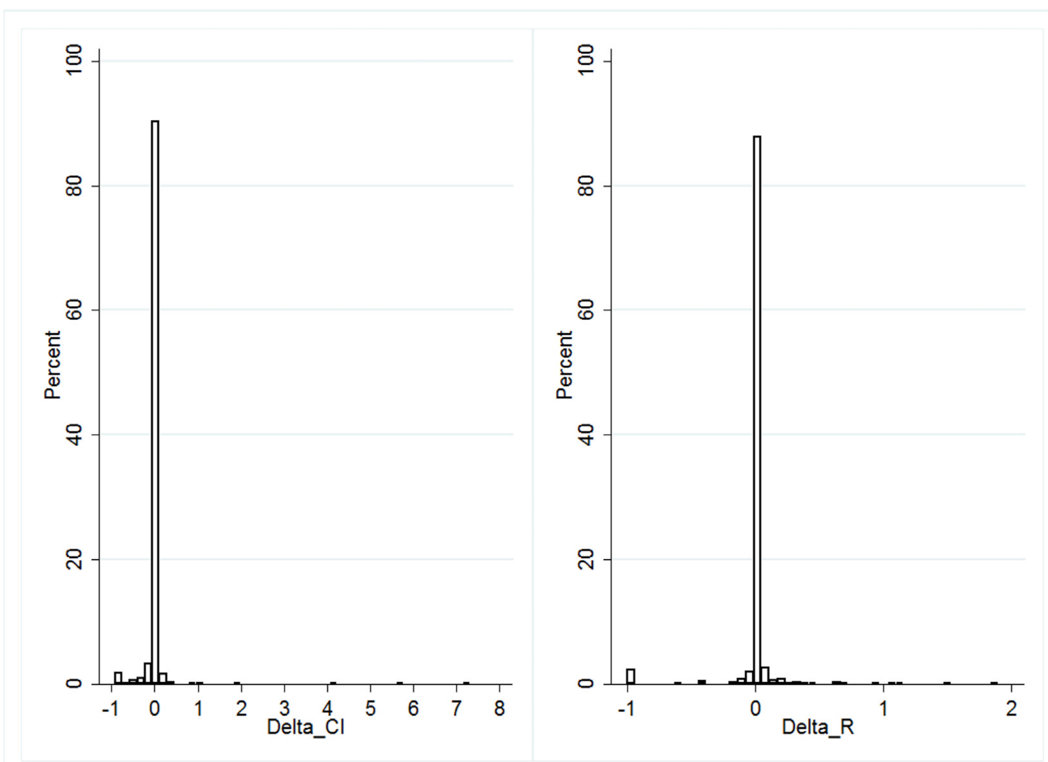

Fig. A.6. Distribution of annual percentage change of complexity (left) and capacity (right) in the U.S. refining industry, 2008–2013.

A.3. Notes

A.3.1. Details of Micro-level Pattern 3.b.

I consider three samples of refineries: (i) all refineries, (ii) refineries located near the Gulf Coast,[42] (iii) refineries that are located within 40 km (or 25 miles) to coastline.

I divide each of these samples into nine groups, as (small capacity, medium capacity, large capacity) × (low complexity, medium complexity, high complexity). I have divided the space of capacity and complexity at their 33.3 and 66.6 percentiles. Holding each of the above samples fixed, I label a group as $g_{(R,C)}$, for example $g_{(3,2)}$ refers to (large capacity, medium complexity).

For each refinery x, I consider a vector $S(x) = [S_i(x)]_{i=1}^{I}$, where i is an import origin, and $I = 33$. $S_i(x) = 1$ if refiner x imports from i, otherwise $S_i(x) = 0$. For each pair of refiners x_1 and x_2, I define an index of *common selections*,

$$common_S(x_1, x_2) = \sum_i [S_i(x_1) = S_i(x_2) = 1]$$

I define $common_S(g)$ for group g

$$common_S(g) = \frac{\sum_{x_1, x_2 \in g} common_S(x_1, x_2)}{N_g(N_g - 1)/2}$$

where N_g is the number of refineries in group g. Table A.11–A.13 report the results for each of the three samples. For example, take Table A.11 which itself contains three sub-tables. According to cell (C3, R3) in these three sub-tables: (i) there are 18 refineries with large capacity and high complexity; (ii) these 18 refineries on average import from 8.3 origins; and (iii) the average number of common origins across all pairs of these 18 refineries is 3.6. That is, among all large and complex refineries, a typical refinery imports from 8.3 origins; and out of these 8.3 it shares only 3.6 origins with another typical refinery.

The ratio of 0.43 = 3.6/8.3 means that on average 57% of trading relationships remains unexplained for the sample of large and complex refineries. The other two tables show similar results for the sample of refineries in the Gulf Coast and in the coastlines. A basic observation is that the selection behavior of observably similar refineries differ to a fairly large extent.

Table A.11
Common Selections, sample (i): All.

Sample size				avg # of origins				Common origins			
	R1	R2	R3		R1	R2	R3		R1	R2	R3
C1	22	10	5	C1	0.6	2.1	6.8	C1	0.1	0.4	2.0
C2	11	11	14	C2	0.6	2.8	7.1	C2	0.2	0.6	2.6
C3	3	16	18	C3	0.3	3.9	8.3	C3	0.0	1.1	3.6

Table A.12
Common selections, sample (ii): Gulf.

Sample size				avg # of origins				Common origins			
	R1	R2	R3		R1	R2	R3		R1	R2	R3
C1	8	4	4	C1	0.5	1.5	8.0	C1	0	0.2	3.0
C2	3	4	5	C2	0	3.0	8.0	C2	0	0.5	3.4
C3	1	4	12	C3	0	4.7	10.1	C3	0	1.3	5.1

Table A.13
Common selections, sample (iii): Coastlines.

Sample size				avg # of origins				Common origins			
	R1	R2	R3		R1	R2	R3		R1	R2	R3
C1	4	5	3	C1	1.2	3.0	10.7	C1	0.2	0.3	6.0
C2	0	4	10	C2	0	5.2	8.1	C2	0	1.7	3.5
C3	0	9	15	C3	0	5.6	9.5	C3	0	1.8	4.5

Regarding the import shares, for each refinery x, I consider a vector $T(x) = [T_i(x)]_{i=1}^{I}$ where $T_i(x)$ is the import share of refiner x from i. For each pair of refiners x_1 and x_2, I define *distance in imports*,

$$distance_T(x_1, x_2) = \left[\sum_{\{i \mid S_i(x_1) = S_i(x_2) = 1\}} (T_i(x_1) - T_i(x_2))^2 \right]^{1/2},$$

as the average distance between import shares of those origins from which both x_1 and x_2 import. Define $distance_T(g)$ for group g,

[42] Sample (ii) consists of Alabama, Arkansas, Louisiana, Mississippi, New Mexico, and Texas. See PADD 3 in Figure A.5.

$$distance_T(g) = \frac{\sum_{x_1,x_2 \in g} distance_T(x_1,x_2)}{N_g(N_g-1)/2}$$

$distance_T(x_1,x_2)$ equals zero if x_1 and x_2 allocate the same share of their demand across their suppliers. The maximum value of $distance_T(x_1,x_2)$ is two. Consider group $(C3,R3)$ in sample (i). The average distance in this group equals 0.65 which is far above zero. It is remarkable that this number does not exceed 0.70 in any cell in any sample.

Table A.14
Distance in import shares.

Sample (i): All				Sample: (ii) Gulf				Sample: (iii) Coastline			
	R1	R2	R3		R1	R2	R3		R1	R2	R3
C1	0.04	0.05	0.31	C1	0	0.1	0.34	C1	0.16	0.13	0.67
C2	0	0.13	0.58	C2	0	0.05	0.69	C2	0	0.27	0.54
C3	0	0.34	0.65	C3	0	0.35	0.69	C3	0	0.41	0.66

A.3.2. Comparing micro patterns to manufacturing input sourcing

I compare the patterns I documented for refineries' imports to those in the manufacturing input sourcing literature. Specifically, I report patterns for manufacturing firms in the United States based on Antràs et al. (2017) and Bernard et al. (2018), Hungary based on Halpern et al. (2015), France based on Blaum et al. (2018), and Argentine based on Gopinath and Neiman (2014).

I begin with the patterns on firms' number of supplier countries. While refineries' imports are almost exclusively concentrated in oil, manufacturing firms typically import a range of products. Therefore, a firm's motivation to buy from different countries partly reflects comparative advantage of supplier countries in differentiated intermediate products. With this consideration, for *all* inputs that a firm imports, the median number of supplier countries is 3 among importing firms in French manufacturing in 2001–2006, 2 among US importing manufacturing firms in 2007, and 2 among Argentine importing firms in 2000. In comparison, the median of the number of foreign supplier countries is 3 for importing American refineries in 2010.[43] The 95th percentile of the number of supplier countries in US manufacturing firms is 11. The median number of supplier countries from which a US manufacturing firm imports the same HS-10 product is 1, and the 95th percentile of that across inputs has a maximum of 9. For importing refineries, the 95th percentile is 12. All in all, the distribution of supplier countries for US refineries lines up well with that of manufacturing firms. However, for manufacturing firm, imports do not typically account for a large share of input costs. Among French manufacturing firms, a larger share of importers spend more than 90% of their material spending on domestic inputs. In the US, firms' foreign input share is 0.03 at the median, 0.15 at the 75th percentile, and 0.47 at the 90th percentile.[44] In Hungry, half of firms do not import at all, while import share in materials is 0.10 for the full sample and 0.27 for importing firms. In manufacturing, import activity is highly concentrated in large firms. For example, the largest 5% of importing firms in Argentina contribute 85% of imports there. In contrast, importing is a more common activity among US refineries. For example, the median ratio of imports to capacity is 30.4% among all US refineries and 44.7% among importing ones. Hence, imports account for a larger share of input costs for US refineries than a typical firm in manufacturing.

However, the figures I report for refineries reflect features that are specific to the US. In some large oil producer countries, such as Russia or Saudi Arabia, refineries purchase crude only domestically. In some other countries that are endowed with nearly zero crude oil reserves, like Germany or Japan, all crude oil purchases are imports.

Lastly, I turn to refinery-level pattern 2.1 which states that larger refineries import from a greater number of sources. Likewise, manufacturing firms that are larger (in terms of employment or output) tend to import from a greater number of countries or a greater number of differentiated products. For example, in Hungry "doubling firm size is associated with a 25% increase in the number of imported products" (Halpern, Koren, and Szeidl, 2015, p. 3666). That larger firms import from a greater number of suppliers is a robust pattern documented for other countries as well. In my data, using the estimates of Table A.4, adding one supplier country, at the median, is associated with 67% increase in capacity size.[45]

A.3.3. Counterfactual trade costs from U.S. to elsewhere

I first regress calibrated trade costs of crude oil, d^0_{ni}, against distance and border. Specifically, this regression gives

$$\log d^0_{ni} = imp_n + \; exp_i + \underset{(0.03)}{0.22} \; \log(distance_{ni}) - \underset{(0.08)}{0.03} \, border_{ni} + error_{ni},$$

where imp_n and exp_i are importer and exporter fixed effects. Standard errors are in parenthesis, number of observations are 359, and $R^2 = 0.68$. As expected, distance is highly and positively correlated with trade costs.

Then I use this relationship to predict trade costs of crude oil from the US to elsewhere when the export ban is lifted. In 2010 data, U.S. exports crude oil only to Canada in 2010, so $d^0_{n,\,USA} = \infty$ for all $n \neq$ Canada. Note that distance and border coefficients as well as importer fixed effects are exogenous to a change in U.S. export barriers. However, lifting the ban changes the U.S. exporter fixed effect. I also have estimated refined oil trade costs using a gravity equation (not reported here) where I find that among all countries, export fixed effect is the smallest for the United States. In addition, similar estimations in other papers show that U.S. has the smallest exporter fixed effect in manufactured products, e.g. see Table 3 in Waugh (2010). In the absence of the export ban, it is then reasonable to suppose that, relative to the other suppliers, U.S. faces small barriers for exporting crude oil

[43] The difference with what reported in Table 2 is that the sample in that table includes all refineries rather than only importing ones.

[44] Antràs, Fort, and Tintelnot (2017) report these figures only when they include wage bills in the domestic purchases.

[45] Related to these comparisons, note that capacity size of a typical refinery changes slowly, if at all, from year to year. Thus, we can plausibly suppose that in a given year a refiner faces a constraint in how much input to purchase and how much output to produce. Envisioning firms in manufacturing, in some industries they may face similar constraints and in some others they may not. The literature on manufacturing firms' sourcing, however, typically assumes that firms can grow or shrink flexibly in response to changes in prices and wages. This conceptual remark is another important difference between the specification of refineries in this paper and that of a typical manufacturing firm elsewhere.

(controlling for distance and other gravity variables). Accordingly, I let U.S. exporter fixed effect be equal the minimum of the exporter fixed effects in the sample. Accordingly I calculate the after-ban counterfactual trade costs of crude from the U.S. to elsewhere. I let trade costs from the U.S. to the 16 countries that do not import at all remain at infinity.

Appendix B. Theory

B.1 Proofs and mathematical derivations

B.1.1 Derivation of Eq. (1)

Given sourcing set S and utilization rate u, at each $t \in [0,1]$ the input cost of a refiner at t is a random variable $V(t) = \min_j \{p_j^0/\tilde{\epsilon}_j(t); j \in S\}$, where by a change of variable, $\tilde{\epsilon} \equiv 1/\epsilon$. Also, $Pr(\tilde{\epsilon}_j(t) \le \tilde{\epsilon}) = \exp(-s_\epsilon \tilde{\epsilon}^{-\eta})$. The probability distribution of random variable V is given by

$$G_V(v) = Pr(V \le v) = 1 - Pr(V > v) = 1 - \prod_{j \in S} Pr\left(\tilde{\epsilon}_j < \frac{p_j}{v}\right) = 1 - \exp(-\Phi v^\eta),$$

where $\Phi = s \sum_{j \in S} p_j^{-\eta}$. P aggregates flows of input costs over the entire period. Thus,

$$P(S) = \int_0^\infty v \, dG_V(v) = \int_0^\infty v \, \Phi \eta v^{\eta-1} \exp(-\Phi v^\eta) \, dv = \Gamma\left(1 + \frac{1}{\eta}\right)\Phi^{-1/\eta} = \varsigma^0 \left(\sum_{j \in S} \left(p_j^0\right)^{-\eta}\right)^{-1/\eta}$$

with $\varsigma^0 \equiv s_\varepsilon^{-1/\eta}\Gamma(1 + \frac{1}{\eta})$. □.

B.1.2 Proof of Proposition 1

I provide a guideline to construct the mapping described in Proposition 1. For more details, see the Online Appendix.

Step 1. *One-to-One Mapping.* $h(q_A) = [z_A, \lambda]$. Note that for the domestic supplier, indexed by j_0, $z_{j_0} \equiv 1$. The mapping is defined between $|S|$ observed nonzero quantities q_j for $j \in S$, and $|S|$ unknowns consisting of $|S| - 1$ unknown z_j for $j \in S/\{j_0\}$ and one unknown λ. Using Eqs. (3), (27), with $k_j = q_j/(ur)$,

$$z_j = \frac{\tilde{k}_j}{1 + \zeta_j + d_j}\left(\frac{p_j^0}{p_{j_0}^0}\right)^{-1} \quad \text{where} \quad \tilde{k}_j \equiv \left(\frac{k_j}{k_{j_0}}\right)^{-1/\eta}, \quad \text{for } j \in S \tag{B.1}$$

Using Eqs. (1), (5), (28),

$$\lambda = \frac{\tilde{P}^0}{\left(\tilde{P}^0 - \tilde{K} p_{j_0}^0\right)(1-u)^2} \tag{B.2}$$

where $u = (1/r)\sum_j q_j$ and $h(q_A) = [z_A, \lambda]$ is given by (B.1) that returns z_A and (B.2) that delivers λ.

Step 2. *Lower bound.* $\underline{z}_B$. The observed set S of suppliers is optimal when the total profit falls by adding unselected suppliers. Holding a refiner fixed, re-index suppliers according to their cost, p_j, from 1 as the lowest-cost supplier to J as the highest-cost supplier. In the Online Appendix, I show that the variable profit rises by diminishing margins from adding new suppliers, which in turn implies that it will be not optimal to add the $k+1$st supplier if the kth supplier is not selected. This feature appears because refineries are capacity constrained; when they add suppliers they face increasing costs of capacity utilization.[46] As a result, the gains from adding the kth supplier to a sourcing set that contains suppliers $1, 2, \ldots, k-1$ is *more* than the gains from adding the $k+1$st supplier to a sourcing set that contains suppliers $1, 2, \ldots, k$. This feature implies that if adding one supplier is not profitable, adding two or more suppliers will not be profitable either. Let S^+ be the counterfactual sourcing set obtained by adding the lowest-cost unselected supplier; p^+ be the cost of this added supplier; and $\pi(S^+; p^+)$ be the associated variable profit. Then, the optimality of S implies that,

$$\underbrace{\pi(S^+; p^+) - (|S| + 1).f}_{\text{lowest-cost unselected supplier with price } p^+ \text{ is added}} \quad \le \quad \underbrace{\pi(S) - |S|.f}_{\text{current set of suppliers}} \quad \Leftrightarrow \pi(S^+; p^+) \le \pi(S) + f.$$

Conditional on (z_A, λ, f), the RHS $(\pi(S) + f)$ is known. The LHS $\pi(S^+; p^+)$ is a decreasing function of p^+. Therefore, S is optimal when for each draw of f, p^+ is higher than a threshold which I call $\underline{p}_B$. The threshold $\underline{p}_B$ is the solution to $\pi(S^+; \underline{p}_B) = \pi(S) + f$. Solving for $\underline{p}_B$, I calculate the threshold on cost shocks $\underline{z}_B$. For $j \notin S$, $\underline{z}_B(j) = \dfrac{\underline{p}_B}{p_j^0(1 + d_j + \zeta_j)}$. Note that $\underline{p}_B \in \mathbb{R}$, but $\underline{z}_B \in \mathbb{R}^{J-|S|}$.

---

[46] In comparison, in the model developed by Antràs et al. (2017), variable profits can rise either by decreasing or increasing differences depending on parameter values. They find increasing differences to be the case in their data. In contrast to theirs where firm can become larger by global sourcing, here refineries face a constraint on their scale of production. This difference in turn reflects the medium-run horizon of my setup compared to the long-run horizon of theirs.

Step 3. *Upper bound.* $\overline{f}$. The observed S is optimal when the total profit falls by dropping selected suppliers. Since the variable profit rises by diminishing margins from adding new suppliers, it suffices to check that dropping only the highest-cost selected supplier is not profitable. Suppose S^{-} is obtained from dropping the highest-cost existing supplier in S. Then, the observed S is optimal if

$$\underbrace{\pi(S^{-})-(|S|-1).f}_{\text{highest-cost existing supplier is dropped}} \leq \underbrace{\pi(S)-|S|.f}_{\text{current set of suppliers}} \Leftrightarrow f \leq \pi(S)-\pi(S^{-}) \equiv \overline{f}.$$

Conditional on (z_A,λ), I directly calculate $\pi(S)$ and $\pi(S^{-})$. Then the upper bound on fixed costs, $\overline{f}$, simply equals $\pi(S) - \pi(S^{-})$.

B.1.3 Proof of Proposition 2

I restate Proposition 1 with a notation that can be readily used to prove Proposition 2. Refer to a random variable by a capital letter, such as Q; and its realization by the same letter in lowercase, such as q. Let $x_A \equiv [\lambda,z_A]$ stack efficiency λ and prices of selected suppliers z_A, with corresponding random variable $X_A \equiv [\Lambda,Z_A]$. Then, Proposition 1 can be written as follows:

$$(R.1) \quad \{Q_A = q_A \mid Q_A{>}0, Q_B = 0\} \leftrightarrow \{X_A = h(q_A) \mid Q_A{>}0, Q_B = 0\}$$
$$(R.2) \quad \{Q_A{>}0, Q_B = 0 \mid X_A = x_A, F = f\} \leftrightarrow \{Z_B \geq \underline{z}_B(x_A, f)\} \text{ and } \left\{F \leq \overline{f}(x_A)\right\}$$

The proof uses $(R.1)$ and $(R.2)$ as described above, and requires two steps as I explain below. As a notation, for a generic random variable Q, let its c.d.f. and p.d.f. be denoted by G_Q and g_Q.

Step 1. The likelihood contribution of the refiner is given by

$$\begin{aligned} L &= g_{Q_A}(q_A \mid S \text{ is selected}) \times Pr\{S \text{ is selected}\} \\ &= |\,\partial x_A/\partial q_A\,| \times g_{X_A}(x_A) \times Pr\{Q_A{>}0, Q_B = 0 \mid X_A = x_A\}. \\ &= g_{Q_A}(q_A \mid Q_A{>}0, Q_B = 0) \times Pr\{Q_A{>}0, Q_B = 0\} \\ &= |\,\partial h(q_A)/\partial q_A\,| \times g_{X_A}(h(q_A) \mid Q_A{>}0, Q_B = 0) \times Pr\{Q_A{>}0, Q_B = 0\} \\ &= |\,\partial h(q_A)/\partial q_A\,| \times g_{X_A}(h(q_A)) \times Pr\{Q_A{>}0, Q_B = 0 \mid X_A = h(q_A)\} \end{aligned} \tag{B.3}$$

Here, $x_A \equiv [\lambda,z_A] = h(q_A)$, and $|\partial x_A/\partial q_A|$ is the absolute value of the determinant of the $|S| \times |S|$ matrix of partial derivatives of the elements of $h(q_A)$ with respect to the elements of q_A. (Recall that the price of the domestic supplier is normalized to its origin price, and $|S|$ is the number of suppliers in S. So, the size of x_A equals $|S| = 1 + (|S|-1)$; one for λ and $|S| - 1$ for z_A.) To derive the third line from the second line, I use the first relation in Proposition 1, (R.1). Suppose that w.l.o.g. h is strictly increasing.[47] Then,

$$Pr(Q_A \leq q_A \mid Q_A{>}0, Q_B = 0) = Pr(X_A \leq h(q_A) \mid Q_A{>}0, Q_B = 0).$$

Taking derivatives with respect to q_A delivers the result:

$$g_{Q_A}(q_A \mid Q_A{>}0, Q_B = 0) = |\partial h(q_A)/\partial q_A| \times g_{X_A}(h(q_A) \mid Q_A{>}0, Q_B = 0).$$

The fourth line is derived from the third line thanks to the Bayes' rule. The fifth line simply rewrites the fourth line in a more compact way.

Step 2. Using the second relation in Proposition 1, (R.2), we can write the last term in Eq. (B.3) as follows:

$$\begin{aligned} Pr(Q_A{>}0, Q_B = 0 \mid X_A = x_A) &= \int_0^{\infty} Pr(Q_A{>}0, Q_B = 0 \mid X_A = x_A, F = f)\, dG_F\left(f|\mu_f, \sigma_f\right) \\ &= \int_0^{\infty} Pr(Z_B \geq \underline{z}_B(x_A, f)) \times I\left(f \leq \overline{f}(x_A)\right) dG_F\left(f|\mu_f, \sigma_f\right) \\ &= \int_0^{\overline{f}(x_A)} \ell_B(x_A, f)\, dG_F\left(f|\mu_f, \sigma_f\right), \end{aligned} \tag{B.4}$$

where $I(f \leq \overline{f}(x_A))$ is an indicator function to be equal one only if $f \leq \overline{f}(x_A)$; and by definition, $\ell_B(x_A, f) = Pr\{Z_B \geq \underline{z}_B(x_A, f)\}$. Plugging (B.4) into Eq. (B.3),

$$L = |\partial x_A/\partial q_A| \times g_{X_A}(x_A) \times \int_0^{\overline{f}(x_A)} \ell_B(x_A, f)\, dG_F\left(f|\mu_f, \sigma_f\right). \tag{B.5}$$

Since $x_A \equiv [\lambda, z_A]$, $g_{X_A}(x_A)$ could be written as:

$$g_{X_A}(x_A) = |\,\partial[\lambda, z_A]/\partial q_A\,| \times g_{\Lambda}(\lambda) \prod_{j \in S} g_Z(z_j), \tag{B.6}$$

where $z_A = [z_j]_{j \in S}$, and $|\partial[\lambda,z_A]/\partial q_A|$ is the absolute value of the determinant of the Jacobian of $[\lambda,z_A]$ with respect to q_A. (Recall that for the domestic supplier z_0 is normalized to one, so $[\lambda,z_A]$ is a vector with $|S|$ random variables). It follows that

[47] The argument holds more generally since h is a one-to-one mapping.

$$L = |\,\partial[\lambda, z_A]/\partial q_A\,| \times g_\lambda(\lambda) \prod_{j\in S} g_Z(z_j) \times \int_0^{\bar{f}(\lambda,z_A)} \ell_B(\lambda, z_A, f)\, dG\left(f|\mu_f, \sigma_f\right). \tag{B.7}$$

The above completes the proof. In addition, I calculate ℓ_B as follows:

$$\ell_B = Pr\{Z_B \geq \underline{z}_B\} = 1 - \prod_{j\notin S} Pr\{z_j < \underline{z}_B(j)\} = 1 - \prod_{j\notin S} G_Z(\underline{z}_B(j))$$

where G_Z is the c.d.f. of Z.

B.2 Model in changes

Definition. A **policy** consists of *new levels* of crude oil endowments $\{(Q^0_{n\tau})'\}$, refining capacity $\{(|\Phi_n|)'\}$, crude oil trade costs $\{(d^0_{ni})'\}$ and *changes* to productivity $\{\hat{A}^k_n\}$ and trade costs $\{\hat{d}^k_n\}$ for $k = 1, 2.., K$.

Definition. For a policy, given $\{\gamma^k_n, \alpha^{kk'}_n, \alpha^k_n, \theta_k, \omega_n, \rho\}$ and variables of baseline equilibrium, an **equilibrium in changes** consists of $(p_{i\tau})'$, $(\tilde{P}^0_n)'$, and $(w_i)'$ such that.

1. Eqs. (1)–(5) (with prime superscript) solve the refiner's problem, and Eqs. (7), (8), (11), (21), (22) (with prime superscript) give $(Q^0_{ni\tau})'$, $(X^0_n)'$, $(\tilde{R}^0_n)'$, $(D^0_n)'$, $(\Pi_n)'$.

2. $\hat{\pi}^k_{ni}$, $\hat{P}^k_n$, $\hat{\beta}^k_n$, $\hat{P}^F_n$, $\hat{\beta}^F_n$ as well as $(X^k_n)'$, $(R^k_n)'$, $(D^k_n)'$, $(E_n)'$ are given by Eqs. (B.8)–(B.19).
3. Market clearing holds for the wholesale refined oil according to Eq. (B.20), for crude oil according to Eq. (B.21), for labor according to Eq. (B.22).

$$\hat{\pi}^k_{ni} = \frac{\hat{A}^k_i\left(\hat{d}^k_{ni}\hat{c}^k_i\right)^{-\theta^k}}{\sum_{i=1}^N \pi^k_{ni}\hat{A}^k_i\left(\hat{d}^k_{ni}\hat{c}^k_i\right)^{-\theta^k}}, \quad k = 1, 2, \ldots, K \tag{B.8}$$

$$\hat{P}^k_n = \left[\sum_{i=1}^N \pi^k_{ni}\hat{A}^k_i\left(\hat{d}^k_{ni}\hat{c}^k_i\right)^{-\theta^k}\right]^{-\frac{1}{\theta^k}}, \quad k = 1, 2, \ldots, K \tag{B.9}$$

$$\hat{c}^1_i = \left(\tilde{P}^0_i\right)'/\tilde{P}^0_i, \qquad \hat{c}^k_i = \left[\left(1-\beta^k_i\right)\left(\hat{P}^1_i\right)^{1-\rho} + \beta^k_i\left(\hat{\tilde{c}}^k_i\right)^{1-\rho}\right]^{\frac{1}{1-\rho}}, \quad k = 2, \ldots, K \tag{B.10}$$

$$\hat{\tilde{c}}^k_i = (\hat{w}_i)^{\gamma^k_i}\left(\prod_{k'=2}^K\left(\hat{P}^{k'}_i\right)^{\alpha^{kk'}_i}\right)^{1-\gamma^k_i}, \quad k = 2, \ldots, K \tag{B.11}$$

$$\hat{\beta}^k_n = \left(\frac{\hat{\tilde{c}}^k_n}{\hat{c}^k_n}\right)^{1-\rho}, \quad k = 2, \ldots, K \tag{B.12}$$

$$\hat{P}^F_n = \left[\left(1-\beta^F_n\right)\left(\hat{P}^1_n\right)^{1-\rho} + \beta^F_n\left(\sum_{k=2}^K\left(\hat{P}^k_n\right)^{\alpha^{F,k}_n}\right)^{1-\rho}\right]^{\frac{1}{1-\rho}} \tag{B.13}$$

$$\hat{\beta}^F_n = \left(\frac{\sum_{k=2}^K\left(\hat{P}^k_n\right)^{\alpha^{F,k}_n}}{\hat{P}^F_n}\right)^{1-\rho} \tag{B.14}$$

$$\left(X^1_n\right)' = \sum_{\ell=2}^K\left(1-\left(\beta^\ell_n\right)'\right)\sum_{i=1}^N\left(\pi^\ell_{in}\right)'\left(X^\ell_i\right)' + \left(1-\left(\beta^F_n\right)'\right)(E_n)' \tag{B.15}$$

$$\left(X^k_n\right)' = \sum_{\ell=2}^K\left(\beta^\ell_n\right)'\alpha^{\ell k}_n\left(1-\gamma^\ell_n\right)\sum_{i=1}^N\left(\pi^\ell_{in}\right)'\left(X^\ell_i\right)' + \alpha^{F,k}_n\left(\beta^F_n\right)'(E_n)', \quad k = 2, \ldots, K \tag{B.16}$$

$$\left(R^k_n\right)' = \sum_{i=1}^N\left(\pi^k_{in}\right)'\left(X^k_i\right)', \quad k = 1, 2, \ldots, K \tag{B.17}$$

$$\left(D^k_n\right)' = \left(X^k_n\right)' - \left(R^k_n\right)', \quad k = 1, 2, \ldots, K \tag{B.18}$$

$$(E_n)' = (w_n)'L_n + \sum_{k=1}^{K}\left(D_n^k\right)' + \left(D_n^0\right)' + (\Pi_n)' \tag{B.19}$$

$$\left(\tilde{R}_n^0\right)' = R_n^1 \tag{B.20}$$

$$\sum_{n=1}^{N}\left(Q_{ni\tau}^0\right)' = \left(Q_{i\tau}^0\right)' \tag{B.21}$$

$$(w_n)'L_n = \sum_{k=2}^{K}\gamma_n^k\left(\beta_n^k\right)'\left(R_n^k\right)' \tag{B.22}$$

Appendix C. Computation

C.1 Numerical integration

Refineries within a country are heterogeneous in five dimensions: The vector of trade cost shocks z, the efficiency in utilization costs λ, the fixed cost shock f, the refinery capacity r.[48]

For numerical integration, I use the method of Quasi Monte Carlo (See Miranda and Fackler, Chapter. 5). I generate Neiderreiter equidistributed sequence of nodes $U^z = (U_j^z)_{j=1}^J \in [0,1]^J$ for the vector of variable cost shocks $z = (z_j)_{j=1}^J$, $U^\lambda \in [0,1]$ for efficiency of utilization costs λ, $U^f \in [0,1]$ for fixed cost shock f, and $U^r \in [0,1]$ for refinery capacity r. For every country, I draw T vectors $U \equiv (U^z, U^\lambda, U^f, U^r)$, save all U's, and keep them fixed through the simulation.

Variable cost shock with respect to supplier j, z_j has a Fréchet distribution with dispersion parameter θ and a location parameter equal to $s_z = \Gamma(1 - 1/\theta)^{-\theta}$ that guarantees $E[z_j] = 1$. For every node U_j^z, the realization of trade cost shock z_j is given by the inverse of Fréchet distribution, $z_j = (-\log(U_j^z)/s_z)^{-1/\theta}$. Efficiency draw λ has a log-Normal distribution with $E[\log f] = \mu_\lambda$ and $var[\log f] = \sigma_\lambda^2$. I use U^λ and the inverse c.d.f of the log-Normal to construct realizations of λ. The fixed cost draw f has a log-Normal distribution with $E[\log\lambda] = \mu_f$ and $var[\log\lambda] = \sigma_f^2$. I use U^f and the inverse c.d.f of the log-Normal to construct realizations of f. The draw of capacity r has a truncated Pareto distribution with shape parameter ϕ. I use U^r and the inverse c.d.f of truncated Pareto to construct realizations of r.

C.2 Calibration algorithm

In this calibration procedure, crude oil prices, p_j^0, are given by data. The algorithm calibrates crude oil trade costs, and modifies the data on the expenditure shares on refined oil, β_n^k, β_n^F, π_{ni}^1, such that the accounting of refined oil flows are consistent with the model structure. The following steps describe the calibration algorithm:

1. Guess trade costs d_{ni}^0, efficiency of utilization costs $\mu_{\lambda,n}$, change to efficiency in retail sale of refined oil products $\hat{A}_n^1$ with respect to the one implied by the raw data, and ς^0.
2. Inner loop. Solve for $(\tilde{P}_n^0)'$ —start with some guess for $(\tilde{P}_n^0)'$ and $(E_n)'$.
   (a) Downstream. Demand for refined oil.
      i. Based on the modified change to efficiency in retail sale of refined oil, $\hat{A}_n^1$, using Eqs. (B.8), (B.9), (B.12), (B.14) calculate modified values of β_n^k, β_n^F, π_{ni}^1 with hat superscript. Then, calculate the modified values of these variables.
      ii. Using Eq. (25) and linear algebra compute X_n^k. Using Eq. (18), calculate R_n^k for $k = 2, \ldots, K$. Using Eq. (24), calculate X_n^1. Using Eq. (18), calculate R_n^1.

   (b) Upstream. Supply of refined oil.
      i. Solve the refiner's problem for every refinery φ in every country n. By aggregation of refineries' sourcing and production, calculate $Q_{ni\tau}^0$, $\tilde{\Pi}_n^0$, Π_n, and $\tilde{R}_n^0$.

   (c) Check market clearing condition (20), $\tilde{R}_n^0 = R_n^1$. Stop upon convergence; otherwise update $\tilde{P}_n^0$ and go to Step (a).
3. Using the output of step 2, compute the aggregate crude oil demand by refineries in n for j, Q_{nj}. Construct the first set of moments, m^1, as *total use* of crude oil n, $m_n^1 \equiv \sum_{i=1}^N \sum_{\tau=1}^2 Q_{ni\tau}^0$ for all n; the second set of moments, m^2, as crude oil trade shares, $m_{ni}^2 \equiv (\sum_{\tau=1}^2 Q_{ni\tau}^0)/m_n^1$ for all n and $i \neq n$; the third set of moments m^3 as $m_n^3 \equiv \tilde{P}_n^0/P_n^0$ for all n, and the fourth moment as global crude oil trade costs relative to global value of crude oil supply, $m^4 = (\sum_n X_n^0 - \sum_n R_n^0)/\sum_n R_n^0$.
4. Stop if $\left|\frac{m_{ni}^1}{m_{ni}^{1,data}} - 1\right| < \varepsilon$, $\left|\frac{m_n^2}{m_n^{2,data}} - 1\right| < \varepsilon$, $\left|\frac{m_n^3}{m_n^{3,data}} - 1\right| < \varepsilon$, $\left|\frac{m^4}{m^{4,data}} - 1\right| < \varepsilon$; otherwise update d_{ni}, $\mu_{\lambda,n}$, $\hat{A}_n^1$, ς^0 according to

$$d_{ni} \leftarrow d_{ni} \times \left(\frac{m_{ni}^1}{m_{ni}^{1,data}}\right)^{\kappa_1}, \mu_{\lambda,n} \leftarrow \mu_{\lambda,n} \times \left(\frac{m_n^2}{m_n^{2,data}}\right)^{-\kappa_2}, A_n^1 \leftarrow A_n^1 \times \left(\frac{m_n^3}{m_n^{3,data}}\right)^{-\kappa_3}, \varsigma^0 \leftarrow \varsigma^0 \times \left(\frac{m^4}{m^{4,data}}\right)^{\kappa_4}$$

where $\kappa_1 > 0$, $\kappa_2 > 0$, $\kappa_3 > 0$, $\kappa_4 > 0$. Then, go to step 2.

I obtain data for the first and second moments based on an accounting of crude oil trade flows described in details in the Online Appendix, and third and fourth moment based on US data. As for the third moments, I modify recored data on refined oil trade shares, π_{ni}^1 and expenditure shares β_n^k and β_n^1 by making them consistent with the accounting of refined oil flows implied by the model structure. Specifically, I suppose that

[48] In the multi-country framework, I assume that the observed part of variable trade cost, d, and the refinery complexity, ζ, are the same for refineries within a country.

the data on π^1_{ni}, β^k_n, and β^1_n are recorded by a measurement error, and in order to obtain their true values I allow appropriate changes to efficiency in retails sale of refined oil, A^1_i, letting every other variable be unchanged. This implies that $\hat{P}^1_n = [\sum_{i=1}^{N} \pi^k_{ni} \hat{A}^1_i]^{-\frac{1}{\theta^1}}$, and

$$(\pi^1_{ni})^\star = \pi^1_{ni}\hat{\pi}^1_{ni} \quad \text{where} \quad \hat{\pi}^1_{ni} = \frac{\hat{A}^1_i}{\left(\hat{P}^1_n\right)^{-\theta^1}} \tag{C.1}$$

$$\left(\beta^k_n\right)^\star = \beta^k_n\hat{\beta}^k_n, \quad \text{where} \quad \hat{\beta}^k_n = \left[\frac{1}{\beta^k_n + \left(1-\beta^k_n\right)\left(\hat{P}^1_n\right)^{1-\rho}}\right], \quad k = 2, \ldots, K \tag{C.2}$$

$$\left(\beta^F_n\right)^\star = \beta^F_n\hat{\beta}^F_n, \quad \text{where} \quad \hat{\beta}^F_n = \left[\frac{1}{\beta^F_n + \left(1-\beta^F_n\right)\left(\hat{P}^1_n\right)^{1-\rho}}\right] \tag{C.3}$$

Within every iteration in the calibration, using these modified shares, $(\pi^1_{ni})^\star$, $(\beta^k_n)^\star$, $(\beta^F_n)^\star$ I calculate demand for refined oil and check whether the model prediction of the third moment matches that in the data.

C.3 Algorithm to compute baseline equilibrium

In the following, we take β^k_n, γ^k_n, $\alpha^{kk'}_n$, π^k_{ni}, β^F_n, $\alpha^{F,k}_n$, and E_n as given.

1. Using Eq. (25) and linear algebra to compute X^k_n for $n = 1, \ldots, N$ and $k = 2, \ldots, K$. Using Eq. (18), calculate R^k_n for $n = 1, \ldots, N$ and $k = 2, \ldots, K$. Using Eq. (24), calculate X^1_n for $n = 1, \ldots, N$. Using Eq. (18), calculate R^1_n for $n = 1, \ldots, N$.
2. *Outer Loop.* Solve for $p_{i\tau}$. Start with a guess for $p_{i\tau}$.
3. *Inner Loop.* For the current guess of $p_{i\tau}$, solve for $\tilde{P}^0_n$. Start with a guess for $\tilde{P}^0_n$.
   (a) Solve the refiner's problem for every refinery φ in every country n. By aggregation of refineries' sourcing and production, calculate $Q^0_{ni\tau}$, $\tilde{\Pi}^0_n$, Π_n, and $\tilde{R}^0_n$.
   (b) Check market clearing condition (20), $\tilde{R}^0_n = R^1_n$. Stop upon convergence; otherwise, update $\tilde{P}^0_n$ and go to Step (a).
4. Check market clearing condition, $\sum_n Q^0_{ni\tau} = Q^0_{i\tau}$. Go to the next step upon convergence; otherwise, update crude oil prices $p_{i\tau}$ and on go Step (3).
5. Calculate baseline wages w_n for $n = 1, \ldots, N$ based on Eq. (19), and baseline trade deficits for $n = 1, \ldots, N$ and $k = 0,1,2, \ldots, K$ based on Eqs. (21) and (26)

To update prices based on market clearing condition, I derive Jacobian matrices analytically and do not use approximations. Let $\tilde{H}_n$ be the excess demand at the wholesale refined oil market n. I construct a Jacobian matrix of the excess demand function $\tilde{\mathbf{J}} = [\tilde{H}'_{ni}]$ where $\tilde{H}'_{ni} = \partial\tilde{H}_n/\partial\tilde{P}^0_i$, and update the guess for $\tilde{\mathbf{P}}^0 = [\tilde{P}^0_n]^N_{n=1}$,

$$\tilde{\mathbf{P}}^0 \leftarrow \tilde{\mathbf{P}}^0 - \tilde{\mathbf{J}}\tilde{\mathbf{H}}^{-1}$$

Similarly, let H_j be excess demand for crude oil j. I Construct a Jacobian matrix for excess crude oil demand function, $\mathbf{J} = [H_{jk}']$, where $H_{jk}' = \partial H_j/\partial p^0_k$, and update the guess for $\mathbf{p}^0 = [p^0_j]^J_{j=1}$

$$\mathbf{p}^0 \leftarrow \mathbf{p}^0 - \mathbf{J}\mathbf{H}^{-1},$$

In addition, I use dampening for updating the guesses of prices in every iteration.

C.4 Algorithm to compute counterfactual equilibrium

1. Outer loop. Solve w_i' and $p_{i\tau}'$ —start with some guess for w_i' and $p_{i\tau}'$.
2. Inner loop. Solve for $(\tilde{P}^0_n)'$ —start with some guess for $(\tilde{P}^0_n)'$ and $(E_n)'$.
   (a) Downstream. Demand for refined oil.
      i. Use Eqs. (B.9), (B.10), (B.11) to solve for $\hat{c}^k_i$ and $\hat{P}^k_i$.
      ii. Compute $\hat{\pi}^k_{ni}$ based on Eq. (B.8), $\hat{\beta}^k_n$ based on B.12, $\hat{P}^F_n$ and $\hat{\beta}^F_n$ based on Eqs. (B.13)–(B.14). Using hat and baseline values of trade shares calculate $(\pi^k_{ni})'$.
      iii. Using Eq. (B.16) and linear algebra, compute $(X^k_n)'$ for $k = 2, \ldots, K$.

iv. Calculate $(X_n^1)'$ based on Eq. (B.15).
v. Using Eq. (B.17) calculate $(R_n^k)'$ for $k = 1, 2, \ldots, K$.

(b) Upstream. Supply of refined oil.

i. Solve the refiner's problem for every refinery φ in every country n. By aggregation of refineries' sourcing and production, calculate $(Q_{ni\tau}^0)'$, $\left(\tilde{R}_n^0\right)'$, $(D_n^0)'$, $(\Pi_n)'$

(c) Calculate $(D_n^k)'$ for $k = 1, \ldots, K$ based on Eq. (B.18) and $(E_n)'$ based on Eq. (B.19).

(d) Check market clearing condition B.20. Stop upon convergence; otherwise, update $\left(\tilde{P}_n^0\right)'$ and go to step (a)

3. Update wages and crude oil prices. Using market clearing condition for crude oil, Eq. (B.21), update crude oil prices $(p_{i\tau})'$. Using market clearing condition for labor, Eq. (B.22), update wages w_n. Stop upon convergence, otherwise go step (1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jinteco.2020.103323.

References

Antràs, P., Fort, T.C., Tintelnot, F., 2017. The margins of global sourcing: theory and evidence from U.S. firms. Am. Econ. Rev. 107 (9), 2514–2564.

Arkolakis, C., Costinot, A., Rodriguez-Clare, A., 2012. New trade models, same old gains? Am. Econ. Rev. 102 (1), 94–130.

Bernard, A.B., Jensen, J.B., Redding, S.J., Schott, P.K., 2018. Global firms. J. Econ. Lit. 56 (2), 565–619.

Blaum, J., Lelarge, C., Peters, M., 2018. The gains from input trade with heterogeneous importers. Am. Econ. J. Macroecon. 10 (4), 77–127.

Broda, C., Weinstein, D.E., 2006. Globalization and the gains from variety. Q. J. Econ. 121 (2), 541–585.

Çakir Melek, N., Plante, M., Yücel, M.K., 2017. The U.S. Shale Oil Boom, the Oil Export Ban, and the Economy: A General Equilibrium Analysis NBER Working Paper 23818.

Caliendo, L., Parro, F., 2015. Estimates of the trade and welfare effects of NAFTA. Rev. Econ. Stud. 82 (1), 1–44.

Canadian Fuels Association, 2013. The Economics of Petroleum Refining: Understanding the Business of Processing Crude Oil into Fuels and Other Value Added Products Discussion paper.

Chesnes, M., 2015. The impact of outages on prices and investment in the U.S. oil refining industry. Energy Econ. 50, 324–336.

Costinot, A., Rodriguez-Clare, A., 2014. Trade theory with numbers: quantifying the consequences of globalization. Handbook of International Economics. 4. Elsevier B.V, pp. 97–261.

Dahl, C., Sterner, T., 1991. Analysing gasoline demand elasticities: a survey. Energy Econ. 13 (3), 203–210.

Dekle, B.R., Eaton, J., Kortum, S., 2007. Unbalanced trade. Am. Econ. Rev. 97 (2), 351–355.

Eaton, J., Kortum, S., 2002. Technology, geography, and trade. Econometrica 70 (5), 1741–1779.

Espey, M., 1998. Gasoline demand revisited: an international meta-analysis of elasticities. Energy Econ. 20 (3), 273–295.

Fally, T., Sayre, J., 2018. Commodity Trade Matters Working Paper.

Farrokhi, F., Pellegrina, H.S., 2019. Global Trade and Margins of Productivity in Agriculture Working Paper.

Farrokhi, F., Soderberyi, A., 2020. Trade Elasticities in General Equilibrium Working Paper.

Fattouh, B., 2011. An Anatomy of the Crude Oil Pricing System. Oxford Institute for Energy Studies.

Galle, S., Rodríguez-Clare, A., Yi, M., 2018. Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade Working Paper.

Gary, J., Handwerk, G., Kaiser, M., 2007. Petroleum Refining: Technology and Economics. fifth edn. CRC Press.

Gopinath, G., Neiman, B., 2014. Trade adjustment and productivity in large crises. Am. Econ. Rev. 104 (3), 793–831.

Halpern, L., Koren, M., Szeidl, A., 2015. Imported inputs and productivity. Am. Econ. Rev. 105 (12), 3660–3703.

Hamilton, J.D., 2011. Nonlinearities and the macroeconomic effects of oil prices. Macroecon. Dyn. 15 (Suppl. 3), 364–378.

Hughes, J.E., Knittel, C.R., Sperling, D., 2008. Evidence of a shift in the short-run Price elasticity of gasoline demand. Energy J. 29 (1), 113–134.

Kilian, L., 2009. Not all oil Price shocks are alike : disentangling demand and supply shocks in the crude oil market. Am. Econ. Rev. 99 (3), 1053–1069.

Kilian, L., 2016. The impact of the shale oil revolution on U.S. oil and gasoline prices. Rev. Environ. Econ. Policy 10 (2), 185–205.

Kilian, L., Murphy, D., 2014. The role of inventories and speculative trading in the global market for cude oil. J. Appl. Econ. 29, 454–478.

Kucheryavyy, K., Lyn, G., Rodriguez-Clare, A., 2019. Grounded by Gravity: A Well-Behaved Trade Model with Industry-Level Working Paper.

Mcrae, S., 2015. Vertical Integration and Price Differentials in the U.S. Crude Oil Market. Working Paper. pp. 1–39.

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

Miranda, M.J., Fackler, P.L., 2004. Applied Computational Economics and Finance. MIT Press.

Nelson, W.L., 1960a. How complexity of a refinery affects costs of processing? Oil & Gas J. 58 (Sept. 26) p. 216.

Nelson, W.L., 1960b. How to describe refining complexity? Oil & Gas J. 58 (March 14) p. 189.

Nelson, W.L., 1961. How to compute refinery complexity? Oil & Gas J. 59 (June 19), p. 109.

Oil & Tanker Trades Outlook, 2015. Oil & Tanker Trades Outlook. 20. Clarkson Research Services, p. 2.

Peterson, D.J., Mahnovski, S., 2003. New Forces at Work in Refining: Industry Views of Critical Business and Operations Trend. RAND Corporation.

Platts, 2010. The Structure of Global Oil Markets Discussion Paper June.

Santos Silva, J., Tenreyro, S., 2006. The log of gravity. Rev. Econ. Stat. 88 (November), 641–658.

Senate Report, 2003. U. S. strategic petroleum reserve : recent policy has increased costs to consumers but not overall permanent subcommittee on committee on governmental affairs united states senate. Committee on Governmental Affairs 1–288 pp. S. PRT. 108–18 U.S.

Shah, N., 1996. Mathematical programming techniques for crude oil scheduling. Comput. Chem. Eng. 20 (2), S1227–S1232.