Global Sourcing in Oil Markets*

Farid Farrokhi

Purdue

November 2016

Abstract

Trade in oil accounts for a large share of world trade, but occupies a small part of the trade literature. This paper develops a multi-country general equilibrium model that incorporates crude oil purchases by refineries and refined oil demand by end-users. I begin by examining data on the crude oil imports of American refineries, then estimate the model by deriving a new procedure that combines data on refineries' selected suppliers and purchased quantities. Using the estimates to simulate the effects of counterfactual policies on oil trade and prices, 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) By lifting the ban on U.S. crude oil exports, annual revenues of U.S. crude oil producers increase by \$8.4 billion, annual profits of U.S. refineries decrease by \$6.5 billion, while American final consumers face a negligibly higher price of refined oil. (iii) Gains from oil trade are immensely larger than gains from trade in the existing models designed for manufacturing.

*I am grateful to my advisors Jonathan Eaton and Stephen Yeaple for their substantial guidance and support, and to Paul Grieco and James Tybout for various fruitful discussions. I am thankful to Thomas Covert, Ed Green, Cecile Gaubert, Marc Henry, David Hummels, Barry Ickes, David Jinkins, Amir Khosrashahi, Kala Krishna, Vijay Krishna, Ahmad Lashkaripour, Heitor Pellegrina, Anson Soderbery, Tim Schmidt-Eisenlohr, Chong Xiang and seminar participants at the University of British Columbia, University of California Berkeley (ARE), Brown University, Pennsylvania State University, Purdue University (Econ and Ag Econ), 2015 Empirical Investigations in International Trade conference, 2016 New Faces in International Economics conference, and 2016 NBER meetings on economics of commodity markets 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. Correspondence: ffarrokh@purdue.edu.

1 Introduction

Trade in natural resources occupies a small part of international trade literature. Oil alone, as the most traded natural resource, has accounted for 12.5% of world trade in the recent decade. 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.[1] Both the specifics of this industry and a worldwide equilibrium analysis must come together to address trade-related questions on oil markets. I seek to further this objective.

This paper develops a general equilibrium framework to study how local 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 countries. 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 end-users as consumers of refined oil? Lastly, I study the welfare implications of ceasing international oil trade between countries or regions of the world. This counterfactual provides a benchmark to compare gains from oil trade and gains from trade in the existing models that are designed for manufacturing.

To address these questions, I first model and estimate costs that refineries face in their international crude oil sourcing, including transport costs, contract enforcement costs, and technological costs of refining. Then, I embed my estimated model of refineries' sourcing into a multi-country general equilibrium framework that also incorporates refined oil demand by downstream end-users. Global trade in crude oil is the endogenous outcome of the aggregation of refineries' sourcing. Trade in refined oil is modeled in a similar fashion to Eaton and Kortum (2002). The down-

[1] For the former, for example, Caliendo and Parro (2015) include refined oil trade in their sectoral analysis of gains from tariff reductions. For the latter, for instance, Sweeney (2014) studies the effect of environmental regulations on refineries' costs and product prices within the U.S. economy.

stream sector uses refined oil and labor to produce final goods. The framework is designed for a medium run in which production flows of crude oil, incumbent refineries, and labor productivity are given. The equilibrium determines prices and trade flows of crude and refined oil as well as the price indicies of final goods.

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 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 24/7– lower their input costs when they diversify their suppliers. Offsetting this benefit, sourcing from each supplier creates fixed costs associated with writing and enforcing contracts. The trade-off between diversification gains and fixed costs explains fact (i).

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

Based on this specification, I develop a new procedure for estimating refineries' sourcing. The task has proved challenging because a refiner's buying decisions are interdependent. In particular, 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 a Tobit formulation. In dealing with these interdependencies, the literature on firms' import behavior usually makes an extreme timing assumption by which a firm learns about its unobserved components of variable trade costs only after selecting its suppliers. Under this assumption, quantities of trade can be estimated independently from selection decisions, e.g.

Halpern et al. (2015), and Antràs et al. (2016).[2] 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. The likelihood function lets a refiner not only buy less from its higher-cost suppliers but also select them with lower probability (from an econometrician's point of view). As a result, my estimation procedure allows the parameters that affect trade quantities to change the selections.

This methodological departure is crucial to my estimates. In particular, either large diversification gains with large fixed costs or small diversification gains with small fixed costs could explain the sparse patterns of trade. Compared to independent estimations of quantities and selections, my all-in-one estimation generates smaller gains and smaller fixed costs. The reason is that there is information in quantities about fixed costs. Observing small quantities of trade rather than zeros implies that fixed costs should be small.

I embed my model of refineries' sourcing, with the parameter estimates, into a general equilibrium framework that features downstream refined oil trade and consumption. To complete my empirical analysis, I estimate refined oil trade costs, and calibrate 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.

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 tightly predicts the change to average prices of crude oil in the US relative to the rest of the world. In addition, the model closely predicts the pass-through to the price of refined oil, as well as the volume of 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

[2] While these papers 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.

at refinery drops by 13.2% in the U.S., 12.2% in other countries of Americas, 11.6% in African countries, and on average 10.5% elsewhere. The results indicate that a shock to U.S. production modestly changes the relative prices of crude oil across countries. 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 4.6%, profits of U.S. refineries would have decreased by 6.3%, and American end-users would have faced 0.1% higher prices of refined oil. These changes translate to $8.4 billion increase in annual revenues of U.S. crude oil producers, and $6.5 billion decrease in annual profits of U.S. refineries.

Lastly, I study gains from oil trade. Specifically, I consider gains to U.S. consumers, as the change to their real wages when oil trade between the U.S. and the rest of the world is prohibitive. I compare my results to benchmarks in the literature. US gains from oil trade are at least ten times larger than its gains from trade in models that are typically designed for manufacturing.

This study relates to a literature that aims to identify causes and consequences of oil shocks using time-series oil price data, e.g. Kilian (2009). My paper complements this literature by studying oil prices across space rather than over time. Moreover, since I model economic decisions that underlie oil trade, I can address a wide set of counterfactual policies.

The paper fits into the literature on international trade in two broad ways. First, the trade literature has studied manufacturing more than natural resources or agriculture. One well-known result that holds across workhorse trade models is that gains from trade are often small (Costinot and Rodriguez-Clare, 2014). These small gains are at odds with the critical role of trade in natural resources. Results in this paper confirm that gains from trade in oil are notably larger than those predicted by standard models typically designed for manufactured products.

In addition, the trade literature has focused more on the behavior of firms' exports than imports. In contrast to canonical models of export participation, models of firms' sourcing feature interdependent decisions for selecting suppliers. In explaining selections into import markets, Antràs et al. (2016) 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 quantities at the level of individual buyers.

The next section provides background and facts on crude oil trade. Section 3 models refineries' sourcing. Section 4 concerns the estimation. Section 5 closes the equilibrium. Section 6 explores quantitative implications of the equilibrium framework. Section 7 concludes.

2 Background & Facts

My purpose in this section is to motivate the main features of my model based on evidence. I first provide background on the refining industry, and document the main features of refineries' import behavior. Then, I explain how the facts motivate the model.

2.1 Background

The Structure of a Refinery. A refinery is an industrial facility for converting crude oil into refined oil products. Figure 1 shows the 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 flow into the distillation unit. The process of boiling crude oil in the distillation unit separates the crude 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]

Figure 1: Refinery Process Flow Chart. Source: Simplified illustration based on Gary et al (2007).

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

[3] A refinery produces a range of products that are largely joint. At a point in time, since the technology of a refinery remains unchanged, the refiner has little flexibility in changing the composition of its products. These products include gasoline, kerosene and jet fuels, diesel, oil fuels, and residuals. Typically, the heavier fuels are the byproduct of the lighter ones.

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,b, 1961) is the standard way of measuring complexity in both the academic literature and the industry. The index is a weighted size of upgrading units divided by capacity.[4] For producing the same value of output, refining heavy and sour crude involves more upgrading processing. For this reason, a more complex refinery has a cost advantage for refining lower quality crude oil.

Crude Oil Procurement. 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 facilities (Platts, 2010). 90-95% of all crude and refined oil are sold under term contracts, usually *annual* contracts that may get renewed each year (Platts, 2010).[5] A typical contract specifies trade volume, date of trade, and the mechanism for price setting. The price is set when the cargo is loaded at the supplier terminal (or delivered at the delivery port). The price is usually set as a function of posted prices assessed by independent companies.[6]

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.[7] [8] As a result, careful scheduling for procurement of crude oil is important. The complications relate to the logistical arrangements in crude procurement including the variations of arrival of tankers at ports, availability of jetties and storage tanks, and availability of pipeline slots. A number of academic studies have developed mathematical programming techniques to solve the problem. A notable

[4] Let B_k be the size of upgrading unit $k = 1, ..., K$. In the literature on engineering and economics of refineries, a weight w_k is given to each unit k, reflecting the costs of investment in unit k. The complexity index equals to $(\sum_{k=1}^{K} w_k B_k)/R$ where R is refinery capacity (i.e. size of distillation unit). See the online data 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)

[6] The two most important of these agencies are Platts and Argus. For details on the relation between posted prices of crude and term contracts, see (Fattouh, 2011, Chapter 3).

[7] Unlike power plants, refineries operate except during scheduled maintenance every three to five years (Sweeney, 2014). Also, as a rare event, an unplanned shutdown for repairs, for example due to a fire, may occur.

[8] 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.

paper here is Shah (1996) which formulates a refinery's optimal scheduling of multiple crude oil grades of different quality and origin.[9]

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."[10] Textbooks on engineering and economics of refineries assume that refineries take prices of refined products and prices of crude oil as given.[11] Such a description is also in line with reports by governments. For example, according to the Canadian Fuels Association, "refiners are price takers: in setting their individual prices, they adapt to market prices."[12,13]

2.2 Facts

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.[14]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 of refineries.

Since EIA does not assign id to refineries, I have 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 manually 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

[9] The scheduling problems have been studied for short-term (month) and long-term (year) horizons. In the short term, the unloading schedules of suppliers are given, and the problem is defined as optimal scheduling from the port to refinery (Pinto et al., 2000). In the long term, the concerns include multiple orders as well as price and cost variability (Chaovalitwongse et al., 2009, p. 115).

[10] Peterson and Mahnovski (2003, p. 7).

[11] As a widely used reference see Gary et al. (2007, p. 19).

[12] Economics of Petroleum Refining by Canadian Fuels Association (2013, p. 3)

[13] 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 (2014).

[14] 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.

accounts for 95% of total capacity and 90% of total imports of the U.S. refining industry in 2010.

In addition, I link refinery-level imports to crude oil prices. Specifically, I have constructed a concordance between worldwide crude oil grades collected by Bloomberg and a classification of crude oil based on origin country and type. Using this concordance and the f.o.b. prices reported by Bloomberg, I compiled the prices of crude oil at each origin country for each type.[15] 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.

The online data appendix describes the details of how I have constructed this unified dataset.[16]

As for clarifying my data limitations, I do not observe the following: sales and production of a refinery, from which domestic suppliers a refinery purchases, and crude oil pipelines within the United States.

I document the main facts in my data, then explain how these facts motivate my model of refineries' sourcing. Appendix A.1 contains supporting tables and figures.

Fact 1. Input diversification. *Refineries typically diversify across sources and across types.*

Table A.1 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. Table A.2 reports the distribution of the number of import origins. The median refiner imports from two countries. The distribution has a fat tail, and the maximum is 16 (compared to 33 origins in total).

In Table A.3, types are classified into four groups as (light, heavy) $\times$ (sweet, sour).[17] The table shows that 88.4% of refineries import more than one type of crude oil, and 36.1% of refineries import all types.

Fact 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 source discourages refineries' imports. (3) More complex refineries import more low-quality crude oil.*

Figures A.1–A.4 show the location of refineries in the U.S., and the distribution of their capac-

[15] F.o.b. stands for "free on board" as the price at source.

[16] Download the online data appendix by clicking here.

[17] 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%.

ity, distance to coast, and complexity. Fact 2.1 is shown by Table A.4: the likelihood that a refinery imports from a higher number of sources strongly correlates with its capacity size. The elasticity of the number of sources with respect to capacity is 0.74.

Table A.5 reports how refineries' capacity, location, and complexity correlate with their imports. Each observation is the volume of imports of a refinery from a source of crude oil including zero import flows.[18] The distance coefficient is highly significant and equals -1.4, where distance is defined between the exact location of a refinery and a source country. A refiner 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.[19] The evidence confirms that complex refineries have a cost-advantage in refining low-quality crude.

Fact 3. Unobserved heterogeneity. *Refineries with similar capacity, location, and complexity allocate their total input demand across suppliers in different ways.*

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 the group equals 5.1; meaning that only half of the trading relationships could be explained by observables. Appendix A.1.3 reports a set of detailed facts on differences in the import behavior of observably similar refineries. The above example is representative.

Fact 4. *Capacity and complexity of refineries change slowly, if at all.*

I look into annual data between 2008 and 2013. Figure A.6 shows the distribution of the annual changes of refineries' capacity and complexity. Both distributions have a large mass at zero.

[18] There are 48 distinct pairs of origin and type of crude oil and 110 refineries importing from these 48 alternatives. Total number of observations equals 5,280.

[19] In other words, all else equal, a more complex refinery imports more, and its larger imports are due to its purchases of low-quality inputs.

There are zero annual changes of capacity in 79.1%, and of complexity in 40.3% of observations (each observation is a refiner-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.

2.3 From the facts to the features of the model

Motivated by facts 1 and 2.1, *refineries diversify, larger refineries diversify more*, I model the refiner's problem as a trade-off between gains from diversification against fixed costs per supplier.

To accommodate facts 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 fact 3, *differences in the import behavior of refineries after controlling for observables*, I introduce unobserved heterogeneity to the variable trade costs between all pairs of refineries and suppliers.[20]

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 fact 4, *capacity and complexity are fairly constant over a year*, I design my framework for a medium run in which refineries' capacity and complexity remain unchanged.

3 A Model of Refineries' Sourcing

I present a model of a refinery's decisions on which suppliers to select and how much crude oil to buy from each supplier. An individual refinery takes the prices of crude oil inputs and of the composite output as given. Section 5 allows these prices to be endogenously determined in a general equilibrium.

[20] It is worth mentioning that trade shares of refineries are typically concentrated on few suppliers that are often not the same across observably similar refineries. This fact is strikingly common in the data of firms' imports in other industries and other countries. Hummels et al. (2014) report that Danish firms concentrate their imports in a narrow but stable set of products which are largely unique to each firm. Blaum et al. (2013, 2015) report the same pattern for French manufacturing firms. This pattern suggests that costs of a foreign supplier vary across importing firms of the same country.

3.1 Environment.

I classify *suppliers of crude oil* by source country and type. Supplier $j = (i, \tau)$ supplies the crude oil from source i of type τ. A menu that lists J suppliers is available to all refineries. Let p_j^{origin} denote the price at the original location of supplier j.

I index refineries by x. Each refiner has a technology that converts crude input to a composite refined output. Capacity of refiner x is denoted by $R(x)$, and its *utilization rate*, denoted by $u(x)$, equals the ratio of the volume of input to capacity. The wholesale price index of the composite refinery output in country n is $\tilde{P}_n$.

The model is designed for a time period that I call a year. The year consists of a continuum of infinitesimal periods $t \in [0, 1]$ that I call days. Let $p_{nj}(x)$ denote the *average cost* per unit of crude oil from supplier j for refiner x in country n. The average cost, $p_{nj}(x)$, depends on the price of supplier j at origin, p_j^{origin}, as well as transport costs, cost-advantage due to complexity, and one unobserved term. I will specify this relation in Section 3.4. The per unit cost of supplier j at t equals

$$p_{nj}(x)\epsilon_{njt}(x),$$

where ϵ is the daily 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. ϵ has *mean one*. $1/\epsilon$ follows a Fréchet distribution with dispersion parameter η. Variance of ϵ is governed by η. The higher η, the smaller the variance.[21]

I now focus on refiner x in country n. Henceforth, I also drop x and n to economize on notation. For example, read p_j as $p_{nj}(x)$. The refiner knows p_j's and ϵ_{jt}'s. In the beginning of

[21] Specifically, $Pr(1/\epsilon \leq 1/\epsilon_0) = \exp(-s_\epsilon\, \epsilon_0^{-\eta})$. Three points come in order: (i) I normalize $s_\epsilon = \left[\Gamma\left(1 + 1/\eta\right)\right]^\eta$ where Γ is the gamma function. This normalization ensures that the mean of ϵ equals one. (ii) Variance of ϵ equals $\frac{\Gamma(2/\eta+1)}{\Gamma(1/\eta+1)^2} - 1$, which is decreasing in η. In a special case where $\eta = \infty$, $Var(\varepsilon) = 0$. In this case, a supplier's daily cost equals its average cost. (iii) The distribution of ϵ under my independence assumption is observationally equivalent to a more general distribution that allows ϵ's to correlate across suppliers,

$$Pr(\frac{1}{\epsilon_1} \leq \frac{1}{\epsilon_{01}}, ..., \frac{1}{\epsilon_J} \leq \frac{1}{\epsilon_{0J}}) = \exp\Big\{ - \Big[\sum_{j=1}^{J} (s_\epsilon \epsilon_j^{-\eta})^{1/\rho} \Big]^\rho \Big\},$$

where $\rho \in (0, 1]$ is the parameter of correlation. The equivalence holds by reinterpreting η as η/ρ.

the year, he orders crude oil for all days of the year by making contracts with set S of suppliers ($S \in \mathbf{S}$, with $\mathbf{S}$ as the power set). The refiner orders crude from supplier $j \in S$ for day t, if supplier j is his lowest-cost supplier at day t, $j = \arg\min_{k \in S}\{p_k \epsilon_{kt}\}$. For making and enforcing a contract with each supplier, the refiner incurs a fixed cost F. The fixed cost is the same across suppliers.

Utilizing capacity requires costly refining activity. For this activity, refineries consume a mix of refined oil products. Since refined oil is also an input needed to refine oil, the unit cost of refining is the price of refinery output, $\tilde{P}$. A refiner that operates at utilization rate $u \in [0, 1)$, incurs a *utilization cost* equal to $R \times C(u)$, where

$$C(u) = \tilde{P} \frac{u}{\lambda(1-u)}. \tag{1}$$

Here, $1/[\lambda(1-u)]$ is the refining activity per unit of utilized capacity. $uR \times (1/[\lambda(1-u)])$ is total refining activity, and the whole term times $\tilde{P}$ is total refining cost. $\lambda > 0$ is the *efficiency* of utilization cost and is refiner-specific. $C(u)$ is increasing and convex in u. The convexity embodies the capacity constraints, and has been estimated and emphasized in the literature on refining industry.[22]

On the sale side, the refiner enters into a contract with wholesale distributors.[23] The refiner commits to supply $\tilde{q} = uR$, and the distributor commits to pay $\tilde{P}uR$. The value of u is held constant over the year, and $\tilde{P}$ is the average value of the price of composite output over the year.

3.2 The Refiner's Problem

The refiner is price-taker in both the procurement and sale sides. Let $P(S)$ denote the average input price if set S of suppliers is selected,

$$P(S) = \int_{\epsilon} \left(\min_{j \in S} \{p_j \epsilon_j\} \right) dG_{\epsilon}(\epsilon). \tag{2}$$

[22] Sweeney (2014) 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.

[23] Sweeney (2014) provides evidence that 87% of gasoline sales and 83% of distillate sales are at the wholesale market.

The variable profit integrates profit flows over the entire period. It equals

$$\pi(S,u) = (\tilde{P} - P(S))uR - C(u)R. \tag{3}$$

Refinery's total profit equals its variable profit net of fixed costs,

$$\Pi(S,u) = \pi(S,u) - |S|F,$$

where $|S|$ is the number of suppliers in S. The refiner maximizes its total profit by choosing a set S of suppliers and utilization rate u,

$$\max_{S\in\mathbf{S},\ u\in[0,1)} \Pi(S,u).$$

A larger S broadens a refiner's access to a wider range of lowest-cost suppliers over the year, so lowers the annual input costs. This mechanism provides a scope for gains from diversification. This scope depends on the variability of suppliers' costs, hence the variance of ϵ, hence η. 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. This relation delivers η as the *trade elasticity*, defined as the elasticity of demanded quantity from a supplier with respect to suppliers' costs conditional on the refiner's selection decisions. See below.

3.3 Solution to the Refiner's Problem

3.3.1 Demand Conditional on Sourcing and Utilization

Since the distribution of prices over the continuum of infinitesimal periods follows a Fréchet distribution, I can closely use the Eaton and Kortum (2002) analysis to calculate trade shares and price indices. Conditional on selecting S, the optimal volume of crude j, denoted by q_j, is zero if

$j \notin S$; and,

$$q_j = k_j uR \quad \text{with} \quad k_j = \frac{p_j^{-\eta}}{\sum\limits_{j \in S} p_j^{-\eta}} \quad \text{for } j \in S. \tag{4}$$

Here, k_j is the demanded share of crude oil j, that is the fraction of times that supplier j is the lowest-cost supplier among the selected suppliers. uR is the utilized capacity, and q_j is the volume of trade. As equation (4) shows, trade elasticity equals η.

It follows from equation (2) that refinery's average input cost equals

$$P(S) = \Big[\sum_{j \in S} p_j^{-\eta} \Big]^{-1/\eta}. \tag{5}$$

Equation (5) measures the extent to which adding a new supplier lowers the input cost. To clarify, suppose a special case where $p_j = p$ for all j. Then $P(S)$ equals $|S|^{-1/\eta} p$. The smaller η, the larger the gains from adding a supplier.

3.3.2 Production and Sourcing

Suppose set S of suppliers is selected. Using equation (3), the F.O.C. delivers the optimal utilization rate,[24]

$$u(S) = (C')^{-1}(\tilde{P} - P(S)). \tag{6}$$

Evaluated at $u(S)$, refinery's variable profit equals

$$\pi(S) = R[uC'(u) - C(u)]\Big|_{u=u(S)}$$

[24] For the sake of completeness, I should add that there is a corner solution $u(S) = 0$ and $\pi(S) = 0$, when $C'(0) > \tilde{P} - P(S)$.

Using the utilization cost given by (1),

$$\begin{aligned}
\pi(S) &= [u(S)]^2 C'(u(S))R \\
&= \underbrace{\tilde{P}u(S)R}_{\text{revenue}} \times \underbrace{\frac{\tilde{P}-P(S)}{\tilde{P}} \times u(S)}_{\text{profit margin}}.
\end{aligned} \tag{7}$$

The above also decomposes the variable profit into revenue and profit margin. Both increase if a larger S is selected.

In the eyes of each refiner, adjusting for quality two suppliers differ only through their average costs. Hence, the refiner ranks suppliers based on p_j's. Then, he finds the optimal cut-point on the ladder of suppliers —where adding a new supplier does not any more cover fixed costs. The solution to the refiner's problem reduces to finding the number of suppliers rather than searching among all possible combinations of them.

Result 1. If the refiner selects L suppliers, its optimal decision is to select the L suppliers with the smallest average costs.

The refiner's maximized *total profit*, therefore, equals:

$$\Pi^\star = \max_{0 \leq L \leq J} \left[\pi(L) - LF\right]. \tag{8}$$

3.4 Specification

The average cost of a supplier contains four components: (i) price at origin p^{origin}, (ii) transport cost d, (ii) cost-advantage due to complexity ζ, (iii) unobserved component z. Specifically, for refiner x, for supplier j as a pair of source-type $i\tau$,

$$p_{i\tau}(x) = \underbrace{p_{i\tau}^{origin}(1 + d_i(x) + \zeta_\tau(x))}_{\text{observable}} \times \underbrace{z_{i\tau}(x)}_{\text{unobs.}} \tag{9}$$

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 vertical integration between refineries and suppliers, geopolitical forces, and unobserved location of infrastructure such as pipelines.

Transport costs are specified as $d_i(x) = (\gamma_i + \gamma_d\, distance_i(x))(\gamma_b)^{border_i(x)}$. Here, γ_i is a source-specific parameter, γ_d is distance coefficient, and γ_b is border coefficient. $distance_i(x)$ is the shortest distance between the capital city of country i and the *exact location* of refiner x within the US. The dummy variable $border_i(x) = 1$ if only if the *state* in which refiner x 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 f.o.b price, $p_0 = p_0^{origin}$.

Since the majority of heavy crude oil grades are also sour, I use a parsimonious specification in which low-quality type includes heavy and sour crude, and high-quality type includes the rest. The complexity effect ζ_τ equals $\beta_0 + \beta_{CI}CI(x)$ if τ is low-quality, and $-\beta_0$ if τ is high-quality. Here, $CI(x)$ is the complexity index of refiner x.[25]

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,

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

with $s_z = \Big[\Gamma(1 - 1/\theta)\Big]^{-\theta}$, 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 $z_0 = 1$.

Note the difference between z and ϵ. Unobserved z is fixed over time, but ϵ varies daily. Their dispersion parameters, in turn, reflect two different features of the data. θ (relating to z) represents the dispersion of variable costs faced by observably similar refineries with respect to a supplier. η (relating to ϵ) governs how much these costs fluctuate over time for every pair of refiner and supplier. As shown in Section 3.5, annual data on trade shares can be used to recover z's, while they inform only the dispersion of ϵ's.

Note the difference between three notions of trade costs. Hold refiner x fixed. Let $\hat{d}_j = 1 + d_j$, and for simplicity shut down the complexity effect $\zeta_j = 0$. At the first level, $p_j^{origin}\hat{d}_j(x)$ is the

[25] 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. $1 - \beta_0 = 1 + \beta_0 + \beta_{CI}CI^{\max} \Rightarrow \beta_0 = -\beta_{CI}CI^{\max}/2$.

unconditional average cost of supplier j at the location of refiner x. At the next level, the average cost is $p_j^{origin}\hat{d}_j(x)z_j(x)$ *conditional* on selecting supplier j. Since a refiner is more likely to select supplier j when z_j is small, conditional trade costs are likely to be smaller than unconditional ones. At the last level, the refiner pays $p_j^{origin}\hat{d}_j(x)z_j(x)\varepsilon_{jt}(x)$ to purchase from j at t. Since supplier j is the lowest cost supplier at t within the selected set, the actual payment is likely to be smaller than $p_j^{origin}\hat{d}_j(x)z_j(x)$.

Regarding the efficiency (Eq. 1), $\ln\lambda$ is a realization of a random variable drawn independently across refineries from a normal distribution G_λ with mean μ_λ and standard deviation σ_λ. I write fixed cost $F = \tilde{P}f$ 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.[26]

To summarize, each refiner is characterized by a vector of *observables* that consists of capacity R, complexity effect ζ, and transport costs $d = (d_j)_{j=1}^J$; and a vector of *unobservables* that consists of unobserved part of variable costs $z = (z_j)_{j=1}^J$, efficiency λ, and fixed costs f. While z, λ and f are known to the refiner, they are unobserved to an econometrician.

3.5 Mapping Between Observed Trade and Unobservables

Handling interdependent decisions for selecting suppliers in firm-based import models has proved challenging. This interdependency arises as selected suppliers jointly contribute to the marginal cost of a firm (here, refiner). For example, suppose the price of a supplier significantly rises. Here, the refiner not only drops that supplier but also its entire import decisions change. For example, the refiner may add a new supplier or purchase more from its existing suppliers. Traditional estimation approaches such as a Tobit formulation are not adequate to address these interdependencies.

The purpose of this subsection is to show how the model, by incorporating unobserved heterogeneity in trade costs, deals with these interdependencies. Specifically, I map the *observed* trade vector q to *unobserved* trade cost shocks z, efficiency λ, and fixed cost f. Then, in Section

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

4.1, I use this mapping to derive a tractable likelihood function that combines data on a refinery's purchased quantities and selection decisions.

Holding a refinery fixed, the set of suppliers is partitioned into the selected ones (part A), and the unselected ones (part B). For instance, q is partitioned into $q_A = [q_j]_{j\in S}$ and $q_B = [q_j]_{j\notin S} = 0$.

The mapping between q and (z, λ, f) has two parts. The first part maps import volumes of selected suppliers q_A to trade cost shocks of selected suppliers z_A and efficiency λ. (Note that z_A includes $|S| - 1$ unobserved entires, because for the domestic supplier, z_0 is normalized to one.) The second part of the mapping determines thresholds on trade cost shocks of unselected suppliers z_B and fixed cost f to ensure that the observed set S of suppliers is optimal. I first summarize the mapping in Proposition 1, then show how to construct the mapping.

Proposition 1. The mapping between the space of observed trade vector, q, and the space of unobservables (trade cost shocks z, efficiency λ, and fixed cost f) is as follows.

- *Conditional on $[q_A > 0,\ q_B = 0]$, purchased quantities of selected suppliers, q_A, map to trade cost shocks for selected suppliers and efficiency, $[z_A, \lambda]$, according to a one-to-one function h, to be derived below.*
- *Conditional on $[z_A, \lambda, f]$, the selections $[q_A > 0,\ q_B = 0]$ are optimal if and only if trade cost shocks of unselected suppliers, z_B, are larger than a lower bound $\underline{z}_B = \underline{z}_B(z_A, \lambda, f)$, and the draw of fixed cost, f, is smaller than an upper bound $\bar{f} = \bar{f}(\lambda, z_A)$.*

The following three steps provide a guideline to construct function h, $\underline{z}_B$, and $\bar{f}$ in closed form. Appendix B.2 presents the details.

Step 1. One-to-one function h. By specification of costs, $p_j = p_j^{origin}(1 + \zeta_j + d_j)z_j$ for $j = 0, 1, ..., J$; where $j = 0$ denotes the domestic supplier whose cost, p_0, is normalized to p_0^{origin}. According to equation (4), for $j \in S$

$$p_j = \tilde{k}_j p_0\,, \quad \textit{where} \quad \tilde{k}_j \equiv \left(\frac{k_j}{k_0}\right)^{-1/\eta} \tag{10}$$

Using equation (10),

$$z_j = \frac{\tilde{k}_j p_0}{p_j^{origin}(1 + \zeta_j + d_j)} \tag{11}$$

Replacing (10) in equation (5) delivers the following,

$$P = \Big[\sum_{j \in S} p_j^{-\eta}\Big]^{-1/\eta} = \tilde{K} p_0\,, \quad where \quad \tilde{K} = \Big[\sum_{j \in S} \tilde{k}_j^{-\eta}\Big]^{-1/\eta} \tag{12}$$

Replacing P from (12) in the first order condition, $\tilde{P} - P = \frac{\tilde{P}}{\lambda(1-u)^2}$, results in

$$\lambda = \frac{\tilde{P}}{(\tilde{P} - \tilde{K} p_0)(1-u)^2} \tag{13}$$

where $u = (\sum_{j \in S} q_j)/R$. Mapping h is given by equation (11) that delivers z_A and equation (13) that delivers λ. Note that h has a closed-form solution, and is one-to-one.

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. According to Result 1, it is not optimal to add the $k+1$st supplier when the kth supplier is not yet selected. In Appendix B.2.1, I show that the variable profit rises by diminishing margins from adding new suppliers.[27] Due to this feature, the gain from adding the kth supplier to a sourcing set that contains suppliers $1, 2, ..., k-1$ is *more* than the gain from adding the $k+1$st supplier to a sourcing set that contains suppliers $1, 2, ..., k$. 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 $\pi(S^+; p^+)$

[27] This feature appears because refineries are capacity constrained; when they add suppliers they face increasing costs of capacity utilization. In the model developed by Antràs et al. (2016), the variable profit 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 limit to the amount they can produce. This difference also highlights the medium-run horizon of my model as opposed to the long-run horizon of theirs.

be the associated variable profit. Then, the optimality of S implies that,

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

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 $\pi(S^+;\underline{p}_B) = \pi(S) + f$. See Appendix B.2.2 for the closed-form expression of $\underline{p}_B$. After solving for $\underline{p}_B$, I calculate the threshold on trade cost shocks $\underline{z}_B$. For $j \notin S$, $\underline{z}_B(j) = \frac{\underline{p}_B}{p_j^{origin}(1+d_j+\zeta_j)}$. Note that $\underline{p}_B \in \mathbb{R}$, but $\underline{z}_B \in \mathbb{R}^{J-|S|}$.

Step 3. Upper bound $\bar{f}$. 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

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

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

4 Estimation

I derive an estimation procedure that summarizes data on refineries' quantities of imports and their selection decisions in a single likelihood function. This estimation procedure has an advantage over its predecessors. In particular, the literature on firm-level import behavior makes an extreme timing assumption by which a firm learns about its unobserved component of variable trade costs, z, only after selecting its suppliers. Under this timing assumption, quantities of trade can be estimated independently from selection decisions (e.g. Halpern et al (2015), and Antràs et al (2014)). By departing from this timing assumption, my estimation allows the parameters that affect trade quantities to change the selections.

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).

The data consist of input volumes q_j, wholesale price of refinery output excluding taxes $\tilde{P}$, prices of crude oil at origin p_j^{origin}, refinery capacity R, complexity CI, and I^d as the information on distance and common border. Let $\mathbf{D}(x)$ summarize the following data:

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

4.1 Likelihood

Let $L_x(\Omega|\mathbf{D}(x), q(x))$ denote the likelihood contribution of refiner x, as a function of the vector of parameters Ω, given exogenous data $\mathbf{D}(x)$ and dependent variable $q(x)$.[28] As there is no strategic competition, the whole likelihood, is given by:

$$\prod_x L_x(\Omega|\mathbf{D}(x), q(x)).$$

The calculation of the likelihood function without using Proposition 1 involves high-dimensional integrals. Besides, simulated maximum likelihood is likely to generate zero values for tiny probabilities. I avoid these difficulties by deriving a likelihood function based on the mapping shown by Proposition 1. Focusing on one refiner, I drop x.

[28] 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 x, L_x, is given by

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

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

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

$$L = \underbrace{J(\lambda, z_A) g_\lambda(\lambda) \prod_{j \in S} g_Z(z_j)}_{L_A\text{, demanded quantities}} \times \underbrace{\int_0^{\bar{f}(\lambda, z_A)} \ell_B(\lambda, z_A, f)\, dG_F(f)}_{L_B\text{, selection decisions}} \tag{14}$$

where $\ell_B = Pr\Big\{z_B \geq \underline{z}_B(\lambda, z_A, f)\Big\}$*. Also,* $[\lambda, z_A]$*,* $\underline{z}_B$*, and* $\bar{f}$ *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*.*

Appendix B.3.1 contains the proof. This proposition summarizes data on import quantities and selection decisions into a single objective function. It also decomposes the likelihood L to the contribution of quantities L_A, and the contribution of selections L_B. The term L_A is the probability density of purchased quantities from selected suppliers. Translating it to the space of unobservables, it equals the probability density of efficiency λ times the probability density of trade cost shocks of selected suppliers z_A, corrected by a Jacobian term for the nonlinear relation between q_A and $[\lambda, z_A]$. The term L_B is the probability that the refiner selects the set S of suppliers among all other possibilities. It is an easy-to-compute one-dimensional integral with respect to the draw of f, and the integrand $\ell_B(\lambda, z_A, f)$ has a closed-form solution.[29]

The likelihood could be expressed as

$$\log L = \log L_A(\eta, \theta, \gamma, \beta_{CI}, \mu_\lambda, \sigma_\lambda) + \log L_B(\eta, \theta, \gamma, \beta_{CI}, \mu_\lambda, \sigma_\lambda, \mu_f, \sigma_f).$$

Here, $(\eta, \theta, \gamma, \beta_{CI}, \mu_\lambda, \sigma_\lambda)$ not only affect the purchased quantities, but may change the selections.[30] For this reason, a refiner not only buys less from its higher-cost suppliers, but also selects them with lower probability (from an econometrician's point of view). This channel proves important as shown in Section 4.3.

[29] 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, $\bar{f} = \infty$. In this case, we can infer no information from dropping a supplier simply because no foreign supplier is selected.

[30] As Proposition 1 shows, $\underline{z}_B$ depends on $[\lambda, z_A, f]$, and $\bar{f}$ depends on $[\lambda, z_A]$. In turn, $[\lambda, z_A]$ is a functions of η, γ, and β_{CI}. In addition, the density probability of λ depends on μ_λ and σ_λ, the density probability of z depends on θ, and the density probability of f depends on μ_f and σ_f.

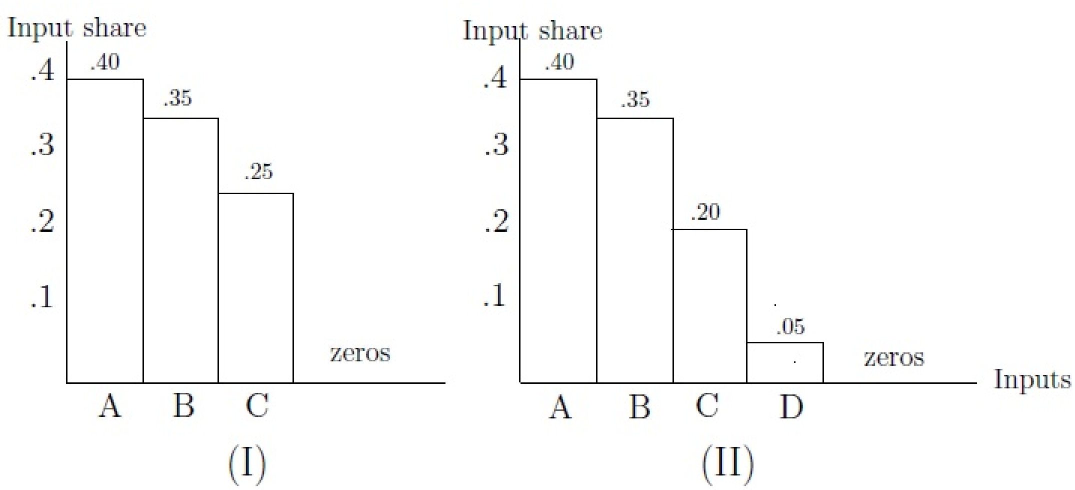
4.2 Identification

I discuss the intuition behind identification of the key model parameters. I first focus on fixed costs, then trade elasticity, then the rest of 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. Figure 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.

Figure 2: Identification of diversification gains and fixed costs



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 draw (which is normalized to one for the domestic supplier, $j = 0$). Equation (4) implies:

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

According to the above, the slope of $\ln(p_j^{obs}/p_0^{obs})$ identifies η if $E[\ln z_j \mid \ln p_j^{obs}/p_0^{obs}] = 0$. This orthogonality condition does not hold 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 trade cost shocks z's. Appendix A.2.1 contains a detailed discussion.

Heterogeneity of variable trade 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 the trade cost shock, the smaller θ.

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 closely relates to the distribution of observed utilization rates.

In addition, I conduct a Monte Carlo analysis described in Appendix C.4. 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.

4.3 Estimation Results

Tables 1–2 in column "all-in-one" report the estimation results. Standard errors are shown in parenthesis. I also report the results based on estimating i) the parameters that govern refineries' variable profit using data on quantities of trade, labeled as "quantities only"; and ii) fixed costs using data on zero-one selections, given the estimates in stage i, labeled as "selections only".

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.[31] 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, unconditional 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 2). Putting these together, *unconditional* prices increase by more than 100% from origins to refineries. In addition, $\theta = 3.16$ implies $Var[z] = 0.38$,[32] 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.[33]

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 1). Besides, the source-specific parameters of variable trade costs are sizably smaller (see Table 2).[34]

---

[31] 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.

[32] Variance of z equals $\Gamma(1-2/\theta)\Big/\Big(\Gamma(1-1/\theta)\Big)^2 - 1$, which is decreasing in θ.

[33] 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.

[34] The only parameters that remain the same are μ_λ (and σ_λ) which govern the scale (and variation) of total input demand.

Table 1: 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 $\ln \lambda$	μ_λ	5.45	5.36	
		(0.14)	(0.14)	
standard deviation of $\ln \lambda$	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

Note: standard errors in parentheses.

Table 2: Estimation Results —Estimates of γ_i, source-specific parameters of variable trade costs

country	all-in-one	quantities only	country	all-in-one	quantities only
Canada	1.08	0.58	Colombia	1.11	0.39
	(0.11)	(0.14)		(0.15)	(0.16)
Mexico	1.27	0.24	Angola	0.95	0.63
	(0.14)	(0.15)		(0.15)	(0.31)
Saudi Arabia	0.86	0.58	Russia	0.91	0.51
	(0.12)	(.21)		(0.14)	(0.19)
Nigeria	0.99	0.32	Brazil	1.04	0.54
	(0.15)	(0.26)		(0.14)	(0.17)
Venezuela	1.24	0.27	Ecuador	0.90	0.43
	(0.18)	(0.17)		(0.13)	(0.18)
Iraq	0.95	0.59	Every other source	1.33	0.57
	(0.13)	(0.24)		(0.18)	(0.21)

Note: standard errors in parentheses.

4.3.1 Model fit & partial equilibrium implications

I simulate my model to evaluate its performance. Specifically, I draw (z, λ, f) for each observable (R, ζ, d) for two thousand times. Each $(z, \lambda, f, R, \zeta, d)$ represents a refiner for which I solve its problem. Then I calculate the average outcome in the industry.

Model Fit. The model predictions 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. In addition, the model predictions closely fit the actual annual input costs in the industry. Results are reported in Appendix A.2.2.

Model fit according to independent estimations. I also evaluate the model performance using the results of the independent estimations of quantities and selections. The model fit is largely poorer for both the distribution of the number of origins and annual input costs. The main reason behind this poor performance is the underestimation of the trade elasticity η. Sourcing from a greater number of suppliers benefits a refinery by reducing its annual input costs. This reduction, as equation 5 describes, is governed by η, with a smaller η implying larger diversification gains. With $\eta \approx 11$, the model generates input costs that are too small to be believable. See Appendix A.2.2 for further discussion.

Quantitative Implications. The following results point to the behavior of a typical individual refinery holding the prices of crude inputs and of composite output fixed.

I first simulate the effect of a 10% increase in variable trade costs, d, on the imports of an individual refinery. Total imports of a typical refinery drop by 26.7%. That is, the elasticity of imports of a typical individual refinery with respect to distance is -2.67.

In addition, by sourcing globally compared with buying only domestically, a typical refinery lowers its complexity-adjusted input costs by 8.2%, accompanied by 56.3% increase in its profits. Since a refinery is capacity constrained, the change to its profits is largely accounted for by the change to the difference between input cost and output price, rather than a change to its production. Here, 56.3% increase in profits is associated with 47.1% increase in the profit margin while only 4.1% increase in production.

Transition to Equilibrium. The above results inform a refiner's behavior rather than the aggregate industry behavior. In turn, the aggregate behavior is key to study how international oil prices endogenously change in response to policy. To this end, Sections 5–6 embed the analysis into a multi-country general equilibrium framework.

5 General Equilibrium

This section links upstream crude oil procurement to downstream trade and consumption of refined oil in a multi-country setting. International trade flows of refined oil, compared with crude, contain 2.5 times more number of nonzero entries; and are much more two-way (see Section 5.3). As these facts are in line with trade of manufactured products, I model refined oil trade using a standard setting similar to Eaton and Kortum (2002).

Embedding my earlier analysis into a multi-country equilibrium framework requires further assumptions about which parameters are universal. The limitation is that a subset of parameters can be identified only from refinery-level data while such data are available only for the US. This subset consists of trade elasticity η, fixed costs $f \sim logN(\mu_f, \sigma_f)$, trade cost shock $z \sim$ Fréchet distribution with dispersion parameter θ, and complexity coefficient β_{CI}. I continue to use my estimates of these parameters in the multi-country setting. I also use the same distribution for efficiency λ as $logN(\mu_\lambda, \sigma_\lambda)$. However, I will revise my estimates of mean of log-efficiency μ_λ, and observed part of variable trade costs d, because (i) μ_λ and d could be sensitive to the performance and geography of American refineries, and (ii) they could be estimated using country-level data (see Section 5.3).

5.1 Framework

Section 5.1.1 concerns the aggregation of refineries' sourcing decisions. Section 5.1.2–5.1.3 links crude oil markets to refined oil trade and consumption. Section 5.1.4 links refined oil markets to the rest of economy. Section 5.1.5 defines the equilibrium.

5.1.1 The Refining Industry & Crude Oil Trade

There are N countries. Each country has a continuum of refineries. A refinery is characterized by x in country n, where $x \equiv (z, f, \lambda, R, \zeta, d)$ —as (trade cost shocks, fixed cost, efficiency, capacity, complexity effect, observable trade costs). The distributions of z, f, and λ are already specified in Section 3.4. I maintain a seamless transition by using the same distributions. Considering the whole vector x, I denote the distribution of refineries in country n by $G_{x,n}$ with support X_n. Measure of incumbent refineries, denoted by M_n, is exogenously given.

Sections 3.1–3.3 describe the refiner's problem and the solution to this problem—to what extent the refiner utilizes its capacity, which suppliers it selects, and how much it buys from each selected supplier. The supply of refinery output to the domestic wholesale market of country n, denoted by $\tilde{Q}_n$, is given by:

$$\tilde{Q}_n = M_n \int_{x \in X_n} \tilde{q}_n(x) \, \mathrm{d}G_{x,n}(x), \tag{15}$$

where $\tilde{q}_n(x) = u_n(x)R(x)$ is refinery output. The aggregate trade flow of crude oil $j = (i, \tau)$ to country n is:

$$Q_{ni\tau} \quad = \quad M_n \int_{x \in X_n} q_{ni\tau}(x) \, \mathrm{d}G_{x,n}(x), \tag{16}$$

where $q_{ni\tau}(x)$ is the flow of crude oil (i, τ) to refiner x in country n (Eq. 4). Variable trade costs are paid to the labor in the importer country. $\tilde{F}_n$ and $\tilde{C}_n$ denote aggregate fixed costs and aggregate utilization costs, respectively. As before, both $\tilde{F}_n$ and $\tilde{C}_n$ are measured in units of refinery output.

The production flow of crude oil of type τ from country i is inelastically given by $Q_{i\tau}$. The nonzero pairs of (i, τ) list the menu of suppliers for refineries all around the world. As before, prices of crude oil at the location of suppliers, $p_{i\tau}$, and wholesale prices of refinery output, $\tilde{P}_n$, are given to a refiner.

5.1.2 Distributors of Refined Oil Products

In each country, refinery output is sold domestically at a competitive wholesale market to a continuum of *distributors*. Each distributor converts the refinery output to a refined oil product $\omega^e \in [0,1]$. The distributors carry out the retail sale of refined products, ω^e's, to the domestic or foreign markets.

The unit cost of ω^e in country i is $[\tilde{P}_i/\xi_i(\omega^e)]$ where $\tilde{P}_i$ is the wholesale price of refinery output in country i, and $\xi_i(\omega^e)$ is the efficiency shock drawn from a Fréchet distribution with dispersion parameter θ^e and location parameter m_i^e. Comparative advantage in refined oil depends not only on productivity in retail sale of refined oil m_i^e, but also on the equilibrium outcome of crude oil markets, summarized by $\tilde{P}_i$.

The *composite of refined oil products* combines the full set of $\omega^e \in [0,1]$ according to a CES aggregator with elasticity of substitution $\sigma^e > 0$. The composite of refined oil products is an input to downstream production.

5.1.3 Market Structure, Prices, and Trade Shares of Refined Oil

Markets of refined oil products are perfectly competitive, and their trade frictions take the standard iceberg form. Delivering a unit of ω^e from country i to country n requires producing d_{ni}^e units in i, where $d_{ni}^e \geq 1$, $d_{ii}^e = 1$, and $d_{ni}^e < d_{nj}^e d_{ji}^e$. Any good ω^e from country i is available for destination n at price $p_{ni}(\omega^e) = \tilde{P}_i d_{ni}^e/\xi_i(\omega^e)$. Country n buys ω^e from the lowest-cost distributor:

$$p_n(\omega^e) = \min\{p_{ni}(\omega^e);\ i = 1, 2, ..., N\}.$$

The share of country n's imports of refined oil products from country i is

$$\pi_{ni}^e = \frac{m_i^e(\tilde{P}_i d_{ni}^e)^{-\theta^e}}{\Phi_n^e}, \qquad \text{with} \qquad \Phi_n^e = \sum_{i=1}^{N} m_i^e(\tilde{P}_i d_{ni}^e)^{-\theta^e}. \tag{17}$$

Assuming that $\sigma^e < \theta^e + 1$, the price index is given by

$$e_n = \gamma^e \Big(\Phi_n^e\Big)^{-1/\theta^e}, \tag{18}$$

where γ^e is a constant[35], and e_n is before-tax price index of refined oil products in country n.

5.1.4 Downstream

Downstream production consists of two sectors: one oil-intensive sector that uses refined oil and labor; and one non-oil-intensive sector that only uses labor. The oil-intensive sector produces a measure one of goods under constant returns to scale. Its unit cost in country n is c_n, where

$$c_n \equiv c(w_n, e_n) = \left(b_n^\rho w_n^{1-\rho} + (1-b_n)^\rho [(1+t_n)e_n]^{1-\rho}\right)^{\frac{1}{1-\rho}}. \tag{19}$$

Here, w_n is wage in country n. e_n is given by equation (18). $t_n \in (-1, \infty)$ is the tax rate on refined oil consumption ($t_n < 0$ refers to subsidy).[36] b_n and $(1-b_n)$ are factor intensities; and $\rho \geq 0$ is the elasticity of substitution between labor and oil. The production is Leonteif if $\rho = 0$, it collapses to Cobb-Douglas at $\rho = 1$, and converges to a linear production if $\rho \to \infty$. Let β_n and $1-\beta_n$ be respectively spending share of producers on labor and oil, then cost minimization results

$$\beta_n = \frac{b_n^\rho w_n^{1-\rho}}{b_n^\rho w_n^{1-\rho} + (1-b_n)^\rho [(1+t_n)e_n]^{1-\rho}}. \tag{20}$$

Producers in the oil-intensive sector sell their products to the domestic market only. I suppose at least there is some output in the non-oil-intensive sector that can be traded at no cost. This output is the numériare. Wages are pinned down by the productivity of the non-oil-intensive sector, and so are exogenous to the oil-intensive sector.

Finally, each country n is endowed by a fixed measure of human capital augmented labor L_n. Consumers in country n spend α_n share of their income on the oil-intensive sector, and $1-\alpha_n$ on the other. The price index faced by final consumers, then, equals

$$P_n^{Final} = w_n^{\alpha_n} c_n^{1-\alpha_n}. \tag{21}$$

[35] $\gamma^e = \left[\Gamma\left(\frac{\theta^e+1-\sigma^e}{\theta^e}\right)\right]^{1/(1-\sigma^e)} \Big/ \Gamma\left(\frac{\theta^e+1}{\theta^e}\right)$

[36] Fuel taxes and subsidies vary largely across countries. For instance, in 2010, price of gasoline in terms of cents per gallon was 954 in Turkey while only 9 in Venezuela. The model, thus, allows for tax-driven shifts to demand schedules.

5.1.5 Equilibrium

Oil revenues of country i is given by $O_i = \sum_{\tau=1}^{2} p_{i\tau} Q_{i\tau} I_{i\tau}$, where $I_{i\tau}$ equals zero if country i does not produce crude oil (i, τ). Aggregate profits of the refining industry is denoted by Π_i. GDP is given by

$$Y_i = w_i L_i + O_i + \Pi_i + \text{Taxes}_i, \tag{22}$$

where taxes are distributed equally across the domestic population. Expenditures of country i on refined oil products is denoted by $Y_i^e = \alpha_i(1-\beta_i)Y_i$. From every $1+t_i$ dollars spent on refined oil products, 1 dollar is paid to sellers and t_i dollars to the tax authority. So, $\text{Taxes}_i = \frac{t_i}{1+t_i}\alpha_i(1-\beta_i)Y_i$, and GDP, Y_i, equals $\left(1 - \frac{t_i}{1+t_i}\alpha_i(1-\beta_i)\right)^{-1}\left(w_i L_i + O_i + \Pi_i\right)$. The market clearing condition for the wholesale market of refinery output in country i is given by

$$\sum_{n=1}^{N} \frac{\pi_{ni}^e Y_n^e}{1+t_n} = \tilde{P}_i \tilde{Q}_i - \tilde{F}_i - \tilde{C}_i \tag{23}$$

The LHS is the spending of oil distributors on country i's refinery output. The RHS is the value of the net supply of refineries to the wholesale market of country i. π_{ni}^e and $\tilde{Q}_i$ are respectively given by (17) and (15). $\tilde{F}_i$ are $\tilde{C}_i$ are aggregate fixed costs and aggregate utilization costs, which are measured in units of refinery output. Lastly, the supply and demand for crude oil $j = (i, \tau)$ equalize:

$$Q_{i\tau} = \sum_{n=1}^{N} Q_{ni\tau}. \tag{24}$$

where $Q_{ni\tau}$ is given by (16).

Definition 1. Given L_i, w_i, t_i, α_i, b_i, $Q_{i\tau}$, $G_{x,i}$, d_{ni}, d_{ni}^e, and M_i, for all n, i, τ, an ***equilibrium*** *is a vector of crude oil prices $p_{i\tau}$ and refinery output prices $\tilde{P}_n$ such that:*

1. *Imports of crude oil and production of refinery output are given by 4–8 for individual refineries, and by 15–16 for the industry.*

2. *Trade shares and price indices of refined oil products are given by 17–18.*
3. *Unit cost and share of spendings on labor for the oil-intensive sector are given by 19–20. The price index of final goods is given by 21.*
4. *Markets of refined oil products, wholesale refinery output, and crude oil clear according to 22-24.*

5.2 Country-Level Data

Table A.12 summarizes all country-level variables that are taken from data as well as sources of these data. Table A.13 lists all countries and their crude oil production, total refining capacity, average complexity index, average utilization rate, refined oil consumption, and ad valorem equivalent tax rate on refined oil consumption.

Domain. The sample uses data of year 2010. A country is chosen if its crude oil production is more than 0.750 million bbl/day or otherwise its refining capacity is more than 0.750 million bbl/day. This criterion selects 33 countries, accounting for 89% of world crude oil production and 81% of world refining capacity. The rest of the world is divided into six regions: rest of Americas, rest of Europe, rest of Eurasia, rest of Middle East, rest of Africa, and rest of Asia and Oceania —summing up to 39 countries and regions covering the whole world.

Trade Flows. Aggregate trade flows of crude oil and refined oil products are available by UN Comtrade Dataset. For crude oil, there are 359 nonzero trade flows plus 31 own-purchases, summing up to 390 nonzero entries in the trade matrix —nonzeros are 32% of all entries when defined between 31 producers and 39 destinations.[37] Trade in refined products compared with crude, is 2.1 times less in value while 2.5 times more in the number of nonzero entries (there are 926 nonzero trade flows for refined oil). Also, in terms of value, 89.5% of refined oil trade is two-way compared to 26.4% for crude. Finally, in 2010, global trade in crude and refined oil accounts for 12.3% of world trade.

Other Data. The source of GDP and population data is Penn World, and of human capital

[37] From the total of 39 countries/regions, 10 of them produce both types of crude oil, 21 countries produce only one type, and 8 countries produce none. From the total of 41 suppliers (pairs of source-type), 27 of them produce high-quality crude accounting for 61% of world's production, and the rest produce low-quality crude. Further, 16 countries do not import crude oil; 9 countries do not export; and the rest both import and export.

data is Barro and Lee (2012). Crude oil production and aggregate refining capacity are reported by EIA. Country-level complexity index and the maximum refinery capacity are from the Oil and Gas Journal. Data on utilization rate at the country level are taken from World Oil and Gas Review published by Eni. The source of refined oil prices is International Fuel Prices by German Agency for International Cooperation. The source of taxes on refined oil consumption is International Energy Agency. The online data appendix contains the details of data construction.

Accounting of Oil Flows. Aggregate data on trade flows of crude oil do not necessarily match the more reliable data on countries' total exports and total purchases of crude oil. This discrepancy is presumably due to the mismeasurement of international trade flows of crude oil. The problem of modifying the reported trade entries can be formalized as a contingency table with given marginals. I use Ireland and Kullback (1968) algorithm to modify the trade entries. The problem reduces to minimizing deviations from reported entries subject to marginal constraints. I define these constrains such that trade flows add up to aggregate exports and aggregate input uses. I explain the details of this algorithm in the online data appendix.

5.3 Quantifying the Framework

I first explain how I solve for equilibrium given all parameters. Then, I quantify the entire model by using my earlier estimates and by calibrating the parameters introduced by the transition to the general equilibrium setting.

5.3.1 Simulation Algorithm

I can not use the method of exact hat algebra, popularized by Dekle et al. (2007), to calculate counterfactual equilibrium outcomes. The reason is that in my setup trading relationships endogenously change in response to shocks. Instead, I parametrize the entire model, and solve the equilibrium by simulation.

Prior to running the simulation, I draw artificial refineries $x = (z, \lambda, f, R, \zeta, d)$ for each country n for T times from the distribution $G_{n,x}$.[38] I hold these realizations fixed as I search for equi-

[38] Here, I assume that the observed part of variable trade cost, d, is the same for all refineries within a country.

librium variables.

The simulation algorithm consists of an inner and an outer loop. In the inner loop, given a vector of crude oil prices, $p_j\ \forall j$, I solve for refinery output prices, $\tilde{P}_n\ \forall n$, such that all markets expect for crude oil suppliers clear. Specifically, I solve the refiner's problem for each artificial refiner x, aggregate refinery-level to country-level variables, and update $\tilde{P}_n$ until all equilibrium conditions, except crude oil market clearing, hold. In the outer loop, I update my guess of crude oil prices, $p_j\ \forall j$, until the market for each supplier of crude oil $j = (i, \tau)$ clears. Appendices C.1 and C.2 describe the numerical integration and simulation algorithm in details.

5.3.2 Calibration

I explain the entire task of calibrating my framework in four steps. Appendix C.3 contains details that are not presented here. The list of parameters is given by Table 3.

Step 1. I use the estimates in Section 4, reported in Table 1, for the trade elasticity η, distribution of fixed costs $G_F \sim \log N(\mu_f, \sigma_f)$, distribution of trade cost shocks $G_z \sim$ Fréchet distribution with dispersion parameter θ, and complexity coefficient β_{CI}. I keep my specification of the distribution of λ as a log-normal distribution. Here, I let efficiency of refineries in country n to have different mean log-efficiency. Specifically, 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 from Section 4, but will calibrate $\mu_{\lambda,n}$ in Step 4. Besides, my earlier estimates of the observed part of variable trade costs of crude oil, d, might reflect the geography of American refineries. Step 4 also calibrates d using country-level data on crude oil trade flows.

Step 2. A subset of parameters, reported in Table 4, are taken from auxiliary data or related empirical bodies of literature. 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 to 50'000 b/d, and $R_n^{\max}$ is taken from the Oil and Gas Journal. The best fit to the data on U.S. refinery capacity is achieved at $\phi = .11$.[39] I assume that all refineries

[39] Specifically, $G_{R,n} = \frac{1-\left(R/R_n^{\min}\right)^{-\phi}}{1-\left(R_n^{\max}/R_n^{\min}\right)^{-\phi}}$. I estimate ϕ using maximum likelihood and data on U.S. refinery capacity.

within a country has the same complexity index equal to its average in that country. I interpret the oil-intensive sector as manufacturing and transportation. Accordingly, the share of expenditures on manufacturing and transportation sectors is used to set α_n. In addition, using data on prices and consumption of refined oil products, together with equation (20), I calibrate the parameter of oil intensity, $1 - b_n$ (see Appendix C.3.1). I set the dispersion parameter of the efficiency of the retail sale of refined oil, θ^e, to 20, equal to the value of trade elasticity I estimated for crude oil.[40] I calibrate the location parameter of the efficiency of country i in retail sale of refined oil, m_i^e, in Step 4. The elasticity of substitution across refined oil products, σ^e, is set to 5. This value plays no role in my comparative statics analysis since it only appears in the constant term before refined oil price index.

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. In an influential study, 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.

Step 3. Trade costs of refined products, d_{ni}^e, are estimated according to a gravity equation delivered from (17),

$$\ln(\pi_{ni}^e / \pi_{nn}^e) = V_i^e - V_n^e - \theta^e \ln d_{ni}^e$$

[40] This value lies in the range of estimates in the literature. Broda and Weinstein report 11.53, Caliendo and Parro report 51.08.

where $V_i^e = \ln m_i^e (\tilde{P}_i)^{-\theta^e}$. Trade costs are specified as

$$\ln d_{ni}^e = exporter_i^e + \gamma_d^e \ln distance_{ni} + b_{ni}^e + l_{ni}^e + \epsilon_{ni}^e$$

Here, $exporter_i^e$ is the exporter-specific parameter of trade cost for country i.[41] $distance_{ni}$ is distance between exporter i and importer n, b_{ni}^e and l_{ni}^e are dummy variables for common border and language. Following Eaton and Kortum (2002), I estimate these parameters using the method of Generalized Least Squares. The results are reported in Tables 5–6.

The estimates of exporter-specific parameters, $exporter_i^e$, represent barriers that are not explained by geographic variables. As column 6 in Table A.13 shows, refined oil consumption is heavily subsidized in a subset of oil-abundant countries. These subsidies –that aim at increasing domestic consumption– are reflected in the estimates as export barriers. At the other extreme, among *non-producers* of crude oil, the estimates of exporter fixed effects are exceptionally large for the Netherlands and Singapore. Their large exporter-specific parameters reflect that these two countries are the oil trade hubs in Europe and Asia.

Step 4. All parameters listed in Table 3 are set in steps 1–3 except mean log efficiency $\mu_{\lambda,n}$, variable trade costs of crude oil d_{ni}, and efficiency in retail sale of refined oil products, m_n^e.

I calibrate μ_λ, d, and m^e 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 μ_λ, d, and m^e, I use these draws to construct artificial refineries $x = (z, \lambda, f, R, \zeta, d)$ in each country n according to distribution $G_{n,x}$. I solve the refiner's problem for each refinery x in every country n, aggregate refinery-level to country-level variables, then match the model to three sets of moments, as explained below.

The first set of moments, A^1, consists of *total use* of crude oil n, $A_n^1 = \sum_{i=1}^{N} \sum_{\tau=1}^{2} Q_{ni\tau}$ for all n. The second set of moments, A^2, contains all crude oil trade shares, denoted by A_{ni}^2, as the ratio of imports from i to n relative to total input use in n, $A_{ni}^2 = \frac{\sum_{\tau=1}^{2} Q_{ni\tau}}{A_n^1}$. The third set of moments A^3 consists of $A_n^3 \equiv \tilde{P}_n / P_n^{avg}$ where P_n^{avg} is average price of crude oil at the location of refineries in

[41] See Waugh (2010) for the advantage of allowing for export fixed effect over import fixed effect.

country n.[42] These three sets of moments sum up to $N^2 + N$ known entries.

The parameters to be calibrated, also, sum up to $N^2 + N$ unknowns: N for $\mu_{\lambda,n}$'s, $N^2 - N$ for d_{ni}'s (by normalization $d_{ii} = 1$), and N for m_n^e's. The set of parameters are just-identified with respect to the set of moments. Given all other parameters, $[\mu_{\lambda,n}]_{n=1}^N$, $[d_{ni}]_{n\neq i}$, and $[m_n^e]_{n=1}^N$ target $[A_n^1]_{n=1}^N$, $[A_{ni}^2]_{n\neq i}$, and $[A_n^3]_{n=1}^N$, respectively. Refinery efficiency governs *total* crude oil purchases. All else being equal, the higher $\mu_{\lambda,n}$, the larger A_n^1. Variable trade costs determine the allocation of total purchases across suppliers. All else being equal, the larger d_{ni}, the smaller A_{ni}^2. Efficiency in retail sale of refined products governs demand for refinery output in the wholesale market. All else being equal, the larger m_n^e, the higher A_n^3. Appendix C.3 describes the calibration algorithm in details.

[42] P_n^{avg}, also called *acquisition cost of crude oil* in country n, is given by

$$P_n^{avg} = \Big(\int_{x \in X_n} u(x) R P(x)\, dG_{x,n}(x) \Big) \Big/ \Big(\int_{x \in X_n} u(x) R \, dG_{x,n}(x) \Big),$$

where $P(x)$ is the input price index of refiner x described by equation (5).

Table 3: List of Parameters

1. Parameters related to refineries and trade in crude oil	
η	trade elasticity for crude oil
G_F	distribution of fixed costs, log-normal (μ_f, σ_f)
G_λ	distribution of efficiency, log-normal $(\mu_\lambda, \sigma_\lambda)$
G_z	distribution of trade cost shock, Fréchet with mean one and dispersion parameter θ
$G_{R,n}$	distribution of capacity R, Pareto with shape parameter ϕ over $[R_n^{\min}, R_n^{\max}]$
β_{CI}	coefficient of complexity index
d_{ni}	variable trade costs of crude oil
2. Parameters related to trade in refined oil products, and downstream production	
ρ	elasticity of substitution between labor and refined oil products
α_n	share of spending on oil-intensive sector
$1 - b_n$	oil intensity
d_{ni}^e	trade costs of refined oil products for flows from n to i
θ^e	dispersion parameter of the distribution of efficiency in retail sale of refined (Fréchet)
m_i^e	location parameter of the distribution of efficiency in retail sale of refined (Fréchet)
σ^e	elasticity of substitution across refined oil products

Table 4: Parameter Values set in Step 2

ϕ	ρ	θ^e	σ^e
0.11	0.25	20	5

Table 5: Refined oil trade costs —Estimates of distance, common border, and common language.

	$-\theta^e\gamma_d^e$	$-\theta^e border^e$	$-\theta^e language^e$
coef.	-1.72	0.90	0.22
s.e.	(0.10)	(0.42)	(0.28)

Note: standard errors in parentheses.

Table 6: Refined oil trade costs —Estimates of exporter-specific parameters, $-\theta^e exporter_i^e$. By normalization, $\sum_{i=1}^{N} exporter_i^e = 0$. For an estimated parameter b, its implied percentage effect on trade cost equals $100(\exp(-b/\theta^e) - 1)$.

Country	Estimate	% Effect
Algeria	-1.7	8.9
Angola	-6.9	41.5
Azerbaijan	-5.2	29.4
Brazil	1.7	-8.1
Canada	1.3	-6.2
China	1.7	-8.1
Colombia	-3.0	15.9
France	2.4	-11.2
Germany	1.9	-9.3
India	2.5	-11.5
Indonesia	-0.4	2.3
Iran	-5.0	28.2
Iraq	-8.2	51.0
Italy	2.3	-10.9
Japan	2.3	-11.0
Kazakhstan	-2.7	14.6
Korea	4.9	-21.6
Kuwait	-2.1	10.8
Libya	-2.9	15.5
Mexico	-0.8	4.1

Country	Estimate	% Effect
cont'd		
Netherlands	6.1	-26.1
Nigeria	-3.1	16.9
Norway	-3.5	19.2
Oman	-5.0	28.2
Qatar	-3.7	20.1
Russia	1.1	-5.5
Saudi Arabia	-2.2	11.7
Singapore	5.1	-22.4
Spain	1.7	-8.3
UAE	-2.3	12.1
United Kingdom	2.6	-12.1
United States	6.2	-26.7
Venezuela	-1.9	10.1
RO America	2.4	-11.4
RO Europe	3.9	-17.9
RO Eurasia	-1.5	8.0
RO Middle East	3.1	-14.3
RO Africa	4.0	-18.0
RO Asia & Oceania	5.0	-22.1

5.4 Model Fit

The calibration matches country-level crude oil trade, $Q_{ni} = \sum_{\tau=1}^{2} Q_{ni\tau}$, rather than country- and type-level trade, $Q_{ni\tau}$. (Because international trade data are available only at the country level). According to equilibrium definition, however, market clearing conditions hold for each supplier as a pair of source country and type, $Q_{i\tau}$. So, when I solve for equilibrium using the calibrated μ_λ, d, and m^e, the model does not have to match the moments exactly. However, the equilibrium outcome almost exactly fits the moments defined in Step 4 of Section 5.3. Specifically, Figures A.8 and A.9 show the model fit to crude oil trade shares and average utilization rates.

In addition, I look into the relation between the calibrated values of crude oil trade costs, d_{ni}, and geographic variables. Specifically, for the sample of nonzero trades, the following relation holds based on an OLS regression,

$$\log d_{ni} = imp_n + exp_i + \underset{(0.02)}{0.22} \log(distance_{ni}) - \underset{(0.07)}{0.03} border_{ni} + error_{ni},$$

where d_{ni} is the calibrated trade cost between importer n and exporter i, imp_n and exp_i are importer and exporter fixed effects. Standard errors are in parenthesis, number of observations are 359, and $R^2 = 0.69$. As expected, distance highly correlates with the calibrated trade costs.

6 Quantitative Predictions

The framework –developed in Sections 3, 5.1, 5.2, and quantified in Sections 4, 5.3– allows me to asses gains to suppliers, refineries, and end-users from changes to policy. Section 6.1 tests out-of-sample predictions of the model for factual changes of crude oil production and refinery capacity of all countries from 2010 to 2013. 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 (i) lifting the ban on US crude oil exports, (ii) reducing trade costs of crude oil from Canada to US, and (iii) ceasing international trade in oil for the US and Europe.

6.1 A Validation: 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 measure of total refining capacity M_i, are exogenously given. In Sections 4 and 5.3, 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. The equilibrium predicts prices and trade flows of crude and refined oil for 2013.

From 2010 to 2013, U.S. oil crude production grew by 36%. 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. Table A.14 reports the changes from 2010 to 2013 in crude oil production and refinery capacity of all countries.

Table 7 reports the data and my model predictions for 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 3.4%.[43,44] 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 2.6% relative to US average price of crude oil.

The model predicts the decline in US/ROW crude oil price ratio at 4.1% compared with 3.4% in the data. Further, the model predicts that US wholesale price of refined products relative to US price of crude oil increases by 4.6% compared with 2.6% in the data. These predictions are in the right direction, and their magnitudes are close to the factual changes. In addition, the model tightly predicts changes to volume of imports, number of trading relationships, and total use of crude oil for the US refining industry.

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

	import volumes	# 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	−15.43%	−12.93%	2.40%	4.68%	−4.11%

The above experiment considers *all* shocks to the location of supply and demand. The next section focuses on the effect of *one* shock on oil prices and trade flows.

[43] 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 to the production of suppliers.

[44] 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 oils 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.

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% equal to two million b/d 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. Here, I consider a counterfactual world where only US production would change by 36% from 2010 to 2013, then study how this boom would propagate across the globe.

Table 8 reports model predictions for changes to three variables: P_n^{avg} as average price of crude oil at the location of refineries defined in footnote 42, u_n^{avg} as average utilization rate of refineries[45], and e_n as before-tax price of refined oil at the location of end-users described by equation 18. Three 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 14.2% in the U.S. and on average 11.2% in the rest of the world. Price of crude oil *at refinery*, drops by 13.2% in the U.S., 12.5–12.6% in Canada and Mexico, 12.1–12.4% in Venezuela and Colombia, 11.6–12.0% in Brazil, rest of Americas, Angola, Algeria, and Nigeria; while less than 11.6% in the rest of the world, and only 9.5–9.8% in Singapore, Japan, and part of Eurasia. Compared with Americas and Africa, countries in Europe, Russia, and part of Asia are less integrated with the U.S. market.

Second, 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 and Nigeria whose initial utilization rates are the lowest among all countries exemplify this mechanism.[46]

[45] u_n^{avg} is given by

$$u_n^{avg} = \Big(\int_{x \in X_n} u(x) R \, dG_{x,n}(x) \Big) \Big/ \Big(\int_{x \in X_n} R \, dG_{x,n}(x) \Big),$$

[46] Correlation between percentage change to P_n^{avg} and log distance of n to US is 0.31 with standard error 0.07. However, an OLS regression of percentage change to u_n^{avg} against log distance of n to US controlling for initial u_n^{avg} yields no statistically significant coefficient on distance.

Third, prices of refined oil products fall less than prices of crude oil. To make the point more clearly, I present a simple one-country model with homogeneous refineries in Appendix B.4. I analytically show there that by an increase in crude oil production, crude oil price drops more than refined oil price. The intuition is straightforward. When worldwide supply of crude oil increases while refineries' capacity has remained 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.

Table 8: Percentage change to crude oil price at refinery P^{avg}, utilization rate u^{avg}, and refined oil price e, in response to 36% rise in U.S. crude oil production.

Country	P^{avg}	u^{avg}	e
Algeria	-11.8	1.2	-7.5
Angola	-12.0	3.5	-8.0
Azerbaijan	-10.9	7.8	-7.9
Brazil	-11.6	1.3	-7.3
Canada	-12.5	1.9	-7.3
China	-11.0	1.2	-7.5
Colombia	-12.4	3.9	-7.8
France	-10.6	4.3	-7.7
Germany	-10.2	3.9	-7.7
India	-11.2	0.5	-7.3
Indonesia	-10.7	2.3	-7.7
Iran	-10.6	0.6	-7.5
Iraq	-11.2	7.4	-8.2
Italy	-10.3	5.6	-7.8
Japan	-9.8	7.1	-8.5
Kazakhstan	-10.4	4.5	-7.9
Korea	-10.3	4.3	-7.8
Kuwait	-10.8	0.6	-7.7
Libya	-10.8	1.2	-7.6
Mexico	-12.6	1.6	-7.2
cont'd			
Netherlands	-10.5	4.5	-7.7
Nigeria	-12.0	7.2	-8.9
Norway	-10.8	1.4	-7.5
Oman	-10.8	1.0	-8.4
Qatar	-10.2	1.0	-7.5
Russia	-10.5	1.0	-7.5
Saudi Arabia	-11.3	1.1	-8.1
Singapore	-9.5	9.5	-7.8
Spain	-10.5	6.5	-7.9
UAE	-10.5	2.7	-7.9
United Kingdom	-10.7	3.1	-7.6
United States	-13.2	1.8	-7.3
Venezuela	-12.1	2.4	-7.4
RO America	-11.8	3.7	-7.6
RO Europe	-10.6	3.3	-7.6
RO Eurasia	-9.7	6.8	-8.0
RO Middle East	-10.5	1.1	-7.7
RO Africa	-11.3	2.6	-7.7
RO Asia & Oceania	-10.7	2.4	-7.7

6.3 Changes to Trade Barriers and Gains from Trade

I study the implications of lifting the ban on US crude oil exports in Section 6.3.1, of reductions in trade costs from Canada to US in Section 6.3.2, and of shutting down international trade in oil in Section 6.3.3.

6.3.1 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 calibrated trade costs and geographic variables to predict these 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).

Table 9: The effects of removing crude oil export restrictions on the U.S. oil industry

import volumes	# of trading relationships	total use of crude oil	refineries' profits	US refined oil price	US crude oil price
15.27%	8.11%	−0.49%	−6.35%	0.10%	4.59%

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 4.59%, US refining industry would have lost 6.35% of its profits, while American end-users would have faced a negligibly higher price of refined oil.[47] Translating these percentage changes to dollar values, revenues of U.S. crude oil producers would have increased by \$8.41 billion and profits of US refineries would have decreased by \$6.51 billion.

Table 10 shows model predictions for US crude oil imports and exports across cases that the ban is or is not lifted. In 2010, US exports were negligible (and only to Canada). Had the ban been lifted when production rose about 2 million b/d, crude oil exports and imports would have increased respectively by 1.41 and 1.34 million b/d.

Table 10: The effects of lifting the export ban on US exports, imports and use of crude oil (million b/d)

	production	exports	imports	total use
Baseline	5.47	0.04	10.42	15.85
36% rise in US production & ban in place	7.45	0.05	8.73	16.13
36% rise in US production & ban is lifted	7.45	1.46	10.07	16.06

Implications for the sources of the decline in the relative price of US crude oil. I study the importance of the ban on US crude oil exports and the boom in US crude oil production for changes to US/ROW crude oil price ratio between 2010 and 2013. In particular, I compare three cases: (i) crude oil production and refining capacity of all countries are set to their values in 2013, (ii) crude oil production and refining capacity of all countries are set to their 2013 values and the ban on US crude oil exports is lifted (iii) only US crude oil production is set to its value in 2013. Table 11 reports the change to US/ROW crude oil price ratio from the baseline to each of these three cases. (In the data, this ratio has changed by -3.40% as reported in Table 7). According to the model predictions, if the ban was lifted, the decline in the US/ROW crude oil price ratio would virtually disappear (compare 1st and 2nd rows). In addition, in the presence of US export ban, out of 4.11% decline in US/ROW price ratio, 3.42% is explained by the boom in US crude oil production while the rest is attributable to changes in supply and demand elsewhere (compare 1st and 3rd rows).

[47] These findings are in line with the views by some of the experts on oil markets. For example, see (Kilian, 2015, 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.".

Table 11: Percent change to US/ROW crude oil price ratio between 2010 and 2013

From baseline to:	
Production and capacity of all countries are set to their 2013 values & US export ban is in place	-4.11%
Production and capacity of all countries are set to their 2013 values & US export ban is lifted	-0.12%
Only US production is set to its 2013 value & US export ban is in place	-3.42%

A comparison to data after lifting the ban. In December 2015, the US government removed the 40-year-old ban on US crude oil exports. I compare the above predictions with the most recent available data by EIA. This comparison is not exact because in my experiment the ban is lifted in a different year than 2016. In the data, in January to August of 2016 compared with January to August of 2015, US crude oil imports increased by 0.529 million b/d and US crude oil exports increased by 0.497 million b/d. In addition, the gap between US crude oil price relative to the rest of the world narrowed between 2015 and 2016. In particular, the crude oil price ratio of West Texas Intermediate in Oklahoma relative to Brent in the North Sea was 0.930 in 2015 while it rose to 0.987 in 2016 (as the average between January to September of 2016). These predictions are in the right direction with the magnitudes that are not far from the data.

6.3.2 Reduction in Trade Costs from Canada to US

The model also can be used to evaluate gains from large infrastructure projects that facilitate oil trade. A notable example is the Keystone pipeline system designed to carry crude oil from Alberta in Canada to the Midwest and Texas in the United States. Canadian crude oil has been considerably cheap compared with international prices.[48] Thus, less costly trade could generate gains particularly to Canadian suppliers and American refiners. Here, I consider counterfactual experiments in which trade costs of crude oil from Canada to US are lowered by 10% and 20%.

Under 20% reduction of trade costs, for the Canadian side, average price of crude oil would rise by 7.83% and profits of refineries would decrease by 21.47%. For the American side, average price of crude oil would fall by 0.55% and profits of refineries would increase by 2.66%. In US dollars, these percentage changes translate to $5.17 billion increase in annual revenues of Canadian crude oil producers, $1.66 billion decrease in annual profits of Canadian refineries, $0.86 billion

[48] In 2010, average price of crude oil at the location of suppliers was 12.8% lower in Canada relative to the rest of the world.

decrease in annual revenues of American crude oil producers, and \$2.32 billion increase in annual profits of American refineries. Total generated gains amount to \$4.97 billion. In addition, in both countries consumers would face only a negligibly higher price of refined oil. Results for 10% reduction in trade costs are also reported in Tables 12-13.

I compare these gains with a back of the envelope calculation of costs associated with construction and operation of the pipelines. According to available estimates, total capital investments to build the Keystone pipeline system sum up to \$12 billion. Assuming a 5% annual cost of capital, a 40% share of capital costs in total construction costs, plus an additional 10% margin due to maintenance, annual costs amount to $\$1.65 = \frac{0.05\times 12\times(1+0.10)}{0.40}$ billion.[49] My analysis admittedly does not incorporate the entire range of gains and costs. In particular, environmental costs, gains and costs due to change in crude oil production, and gains associated with job creations are not part of this analysis. Despite these limitations, it is interesting that the gains outweight the costs if trade costs reduce only by 10% (2.24 compared to 1.65), and do so by a large margin if trade costs decrease by 20% (4.97 compared to 1.65).

Table 12: Effects of 10% and 20% reductions in trade costs of crude oil from Canada to US (percentage change)

	Crude oil price		Refineries' profits		Refined oil price	
	USA	Canada	USA	Canada	USA	Canada
10% reduction	-0.23%	3.68%	1.21%	-11.36%	0.01%	0.03%
20% reduction	-0.55%	7.83%	2.66%	-21.47%	0.04%	0.09%

[49] For the capital cost, see the report by the Keystone Pipeline System, TransCanada, February 2011. The other numbers are taken from the IHS economic report (O'Neil et al., 2016) for pipelines of 20-inch diameter within the United States. Specifically, the share of capital costs in total construction costs equals 40.9 percent, and the ratio of annual costs of operation and maintenance relative to total construction costs for newly constructed pipelines equals 9.4 percent. I use these numbers as rough estimates particularly by noting that the Keystone pipelines do not have the same features as the pipelines studied in the report e.g. it is designed for 30- and 36-inch diameters.

Table 13: Effects of 10% and 20% reductions in trade costs of crude oil from Canada to US (billion US dollars)

	10% reduction			20% reduction		
	USA	Canada	Total	USA	Canada	Total
Profits of refiners	1.05	-0.87	0.18	2.32	-1.66	0.66
Revenues of crude oil producers	-0.37	2.43	2.06	-0.86	5.17	4.31
Total	0.68	1.56	2.24	1.46	3.51	4.97

6.3.3 Gains from Oil Trade

I examine gains from oil trade by simulating counterfactual experiments in which oil trade between countries or regions of the world is prohibitive. I then compare my results to the literature on gains from trade.[50]

Gains from oil trade for the United States. I start with a counterfactual world where oil trade between the United States and the rest of the world is prohibitive. Specifically, I raise trade costs of both crude and refined oil 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 oil trade in my framework to typical gains from trade in the literature.

In the U.S. economy, in the autarky compared with the baseline, average price of crude oil at source increases by 1178.0%, input costs of refineries increase by 1289.2%, profits of refineries drop by 94.0%, price index of refined oil increases by 998.0%, and price index of final goods rises by 23.2%. Because of the increase in U.S. crude oil revenues, gdp rises by 15.3%. Consequently, real gdp (gdp divided by price index of final goods) decreases by 6.4%.

I compare my results on gains from trade in oil, to the results in the literature on gains from trade in manufactures. From baseline to autarky, U.S. real wage (wage divided by price index of final goods) drops by 18.8%.[51] Eaton and Kortum (2002) provide a benchmark for gains from

[50] 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.

[51] Since wage is exogenously pinned down by a non-oil-intensive sector, the whole change to the real wage comes from the price index of final goods.

trade in manufactures when wages are pinned down in a non-manufacturing sector. When they shut down trade in manufactures, real wage in the U.S. drops by 0.8% (Table IX under column "mobile labor"). According to the benchmark provided by Costinot and Rodriguez-Clare (2014), gains from trade would equal 1.8% in a one-sector gravity-based trade model. (Table 4.1 under column "one sector"). In terms of changes to real wages from baseline to autarky, US gains from oil trade are at least ten times larger than its gains from trade in the benchmark models.

Gains from oil trade for Europe. Consider a counterfactual world where oil 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 both crude and refined oil between European countries and all non-European countries to infinity.

Across European countries the price of crude oil at the location of refinery increases by 1916-2284%, price index of refined oil products increases by 1670-1991%, and price index of final goods rises by 41-71%. 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 sources. In addition, real wages across these countries decrease between 29% for the United Kingdom and 41% for the Netherlands. See Table 14.

Even though this counterfactual is less extreme than a complete autarky at the level of individual countries, changes to real wages here are large compared with those in the literature. As a benchmark, welfare gains from trade for European countries according to the one-sector version of Costinot and Rodriguez-Clare is in the range of 3-8%.

Table 14: From baseline to shutting down crude and refined oil trade between Europe and the rest of the world (percentage change)

	price of crude oil at refinery	price of refined oil	price of final goods	real wage
France	1966.6%	1786.8%	44.2%	−30.7%
Germany	1918.9%	1774.0%	42.4%	−29.8%
Italy	2284.1%	1991.0%	44.5%	−30.8%
Netherlands	1916.7%	1754.8%	71.0%	−41.5%
Norway	2073.7%	1849.6%	43.4%	−30.3%
Spain	2233.5%	1932.8%	48.5%	−32.7%
UK	1840.4%	1670.6%	41.6%	−29.4%
RO Europe	2134.7%	1868.4%	52.7%	−34.5%

What are the sources of gains from oil trade in my framework? The features in my model that matter for gains from oil trade could be distinguished in connection with the literature that aims to put numbers on welfare gains from trade as studied in details in Costinot and Rodriguez-Clare. Three features are key: (i) share of a country's trade in oil with itself, (ii) elasticity of substitution across oil suppliers, and (iii) elasticity of substitution across oil and other factors of production. The smaller each of these, the larger the gains from oil trade. The observed domestic share of trade in crude oil is a major factor for a subset of countries that produce small amount of crude oil. The high elasticity of substitution across oil suppliers is certainly not a source of large gains because it means that oil from one supplier is highly substitutable for oil from other ones. However, sectoral elasticity of oil is very small, meaning that end-users can hardly substitute oil with other products. This small elasticity is another source of large gains in my model compared to multi-sector models that assume Cobb-Douglas production function.

6.3.4 Discussion on Interpretation of Results

I have studied how shocks to oil production and trade costs affect international oil prices and trade. A brief discussion is noteworthy about interpretation and limitations of the results.

The model is designed for a short or medium-run horizon in which refining capacity and crude oil supply tend to be inelastic. During 2008-2012, average annual growth rate of national refining capacity equals 1.08% across countries with more than 0.1 million b/d capacity, and 2.2% across the ten largest countries in terms of refining capacity. In addition, average annual growth rate of the number of refineries in the world equals -0.07%. As for crude oil production, average annual growth in the same period equals 0.04% across countries with more than 0.1 million b/d production, and 1.9% across the ten largest producers. At the level of individual oil wells, Anderson, Kellogg, and Salant (2016) document that production in existing wells in Texas is largely dominated by a long-term trend with only a negligible response to crude oil spot prices.[52]

In addition, in the long run, trade costs could endogenously change in response to crude oil supply shocks. For instance, developments in the US transport system altered oil transport costs within the United States with a two or three year lag after the boom in US crude oil production.[53] Rather than endogenizing crude oil supply or crude oil trade costs, which are tasks beyond the scope of this paper and much more relevant to long-run outcomes, I have focused on medium-run general equilibrium effects that operate through the trade dimension.

7 Conclusion

This paper develops a general equilibrium framework that incorporates crude oil purchases by refineries and refined oil demand by downstream end-users. I model refineries' sourcing from international suppliers, and derive an estimation procedure that combines refinery-level data on selected suppliers and purchased quantities. I use my estimates in the general equilibrium framework to perform counterfactual experiments. 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 or building pipelines between Canada and US can generate net gains. In particular, lifting the ban generates distributional impacts across U.S. crude oil producers and U.S. refineries, with negligible effect on U.S. final consumers. Lastly, gains from oil trade in my framework are substantially larger than

[52]It is worth-mentioning that crude oil supply could be more flexible if a supplier had considerable spare capacity such as Saudi Arabia in certain historical periods, or in case of hydraulic fracturing due to technological features.

[53]As for pipelines, total mileages covered by crude oil pipelines in the US, increased by an average annual rate of 2.4% between 2009 and 2013, with a surge to an average annual rate of 11.6% between 2013 and 2015. (U.S. Liquids Pipeline Usage & Mileage Reports by Association of Oil Pipelines (AOPL) and American Petroleum Institute (API))

gains from trade in standard models that are originally designed for manufactures trade.

The model of refineries' sourcing developed in this paper can be used in other applications in which input users select among available suppliers and purchase non-negative amounts from each selected supplier. The tools developed in Sections 3–4 allow for estimating such models by incorporating the heterogeneity in trade frictions between individual buyers and sellers.

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.

References

Antràs, P., Fort, T. C., and Tintelnot, F. (2016). The Margins of Global Sourcing: Theory and Evidence from U.S. Firms.

Association of Oil Pipelines (AOPL) and American Petroleum Institute (API) (2015). U.S. Liquids Pipeline Usage & Mileage Report. Technical report.

Barro, R. J. and Lee, J.-w. (2012). A New Data Set of Educational Attainment in the World.

Blaum, J., Lelarge, C., and Peters, M. (2013). Non-Homothetic Import Demand: Firm Productivity and Quality Bias.

Blaum, J., Peters, M., and Lelarge, C. (2015). The Gains from Input Trade in Firm-Based Models of Importing.

Broda, C. and Weinstein, D. E. (2006). Globalization and the Gains from Variety. *The Quarterly Journal of Economics*, 121(2):541–585.

Caliendo, L. and Parro, F. (2015). Estimates of the trade and welfare effects of NAFTA. *Review of Economic Studies*, 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. Technical report.

Chaovalitwongse, W., Furman, K., and Pardalos, P. (2009). *Optimization and Logistics Challenges in the Enterprise*. Series: Springer Optimization and Its Applications.

Costinot, A. and Rodriguez-Clare, A. (2014). Trade Theory with Numbers: Quantifying the Consequences of Globalization. In *Handbook of International Economics*, volume 4, pages 197–261. Elsevier B.V.

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

Dekle, B. R., Eaton, J., and Kortum, S. (2007). Unbalanced Trade. *The American Economic Review*, 97(2):351–355.

Eaton, J. and 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 Economics*, 20(3):273–295.

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

Gary, J., Handwerk, G., and Kaiser, M. (2007). *Petroleum Refining: Technology and Economics*. CRC Press, fifth edition.

Halpern, L., Koren, M., and Szeidl, A. (2015). Imported Inputs and Productivity. *American Economic Review*, 105(12):3660–3703.

Hughes, J. E., Knittel, C. R., and Sperling, D. (2008). Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand. *Energy Journal*, 29(1):113–134.

Hummels, D., Jorgensen, R., Munch, J., and Xiang, C. (2014). The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data. *The American Economic Review*, 104(6):1597–1629.

Ireland, C. T. and Kullback, S. (1968). Contingency Tables with Given Marginals. *Biometrika*, 55(1):179–188.

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

Kilian, L. (2015). The Impact of the Shale Oil Revolution on U.S. Oil and Gasoline Prices.

Kilian, L. and Murphy, D. (2014). No Title. *Journal of Applied Econometrics*, 29:454–478.

McRae, S. (2015). Vertical Integration and Price Differentials in the U.S. Crude Oil Market.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6):1695–1725.

Miranda, M. J. and 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 Journal*, September:216.

Nelson, W. L. (1960b). How to Describe Refining Complexity? *Oil & Gas Journal*, March 14:189.

Nelson, W. L. (1961). How to Compute Refinery Complexity? *Oil & Gas Journal*, June 19:109.

O'Neil, B., Hopkins, P., and Gressley, J. (2016). The Economic Impact of Crude Oil Pipeline Construction and Operation. Technical report, IHS Economics Report.

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

Pinto, J., Joly, M., and Moro, L. (2000). Planning and Scheduling Models for Refinery Operations. *Computers & Chemical Engineering*, 24(9-10):2259–2276.

Platts (2010). The Structure of Global Oil Markets. Technical Report June.

Santos Silva, J. and Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, 88(November):641–658.

Shah, N. (1996). Mathematical Programming Techniques For Crude Oil Scheduling. *Computers & Chemical Engineering*, 20(2):S1227–S1232.

Soderbery, A. (2015). Estimating Import Supply and Demand Elasticities: Analysis and Implications. *Journal of International Economics*, 96(1):1–17.

Sweeney, R. L. (2014). Environmental Regulation , Imperfect Competition and Market Spillovers : The Impact of the 1990 Clean Air Act Amendments on the US Oil Refining Industry.

Appendix A Tables, Figures, and Notes

A.1 Tables, figures, and notes for Section 2

A.1.1 Tables

Table A.1: 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.2: Distribution of number of import origins for American refineries, 2010

percentile	P25	P50	P75	P90	P99	Max
# of supplier countries	1	2	7	10	14	16

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) $\times$ (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%.

Number of import origins vs capacity, geography, and complexity: Larger refineries systematically import from a higher number of sources —the coefficient of logarithm of capacity is positive and highly significant. At the median number of import origins (which equals 2), adding one source is associated with 67% increase in capacity. Refineries that are close to coastal areas, significantly import from a higher number of sources. Moreover, more complex refineries tend to import from a higher number of sources.

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 Figure 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 r, possibly zero		
	(1)	(2)
log distance$_{ri}$	-1.389	-2.168
	(0.245)	(0.342)
border$_{ri}$	0.788	0.717
	(0.404)	(0.422)
log (f.o.b. price)$_{i\tau}$	-4.681	-4.413
	(2.449)	(1.866)
Type L	-4.514	-4.412
	(1.448)	(1.866)
Type L$\times$log CI$_r$	1.449	1.827
	(0.401)	(0.826)
Type H$\times$log CI$_r$	-0.408	–
	(0.501)	
log capacity$_r$	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 $TypeH \times log(CI)$ or $TypeL \times log(CI)$.[a]

[a] 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).

A.1.2 Figures

Figure 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.

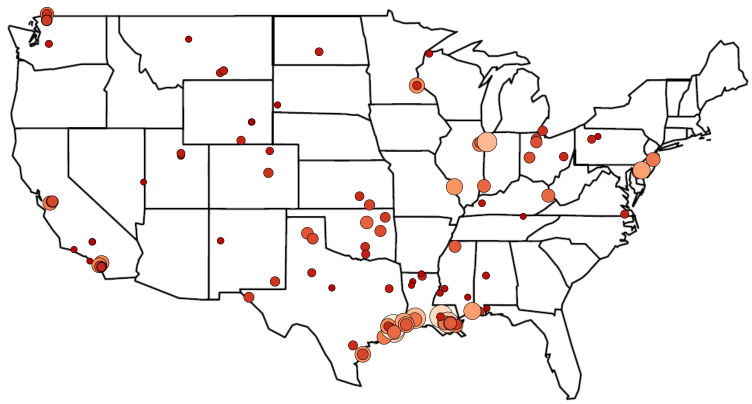


Figure A.2: Distribution of refinery capacity in the U.S. refining industry, 2010.

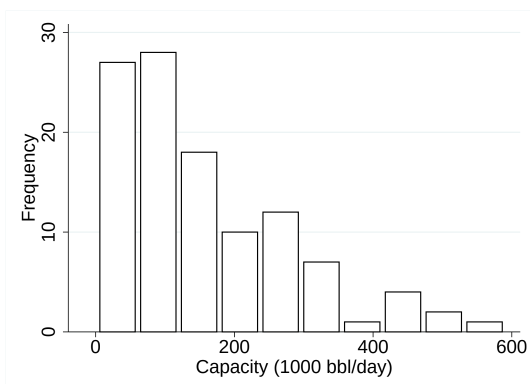


Figure A.3: Distribution of refinery distance to coastline for the U.S refineries, 2010. Distance to coastline is defined as the distance between location of a refinery to the closest port in the U.S.

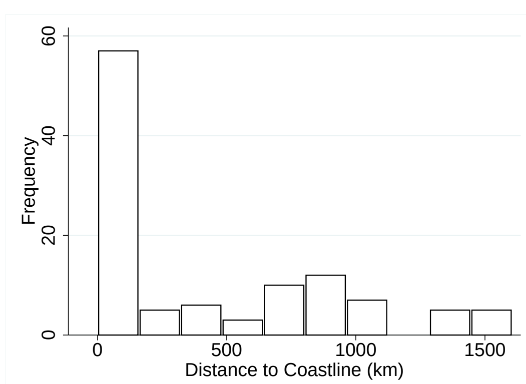


Figure A.4: Distribution of complexity index in the U.S. refining industry, 2010.

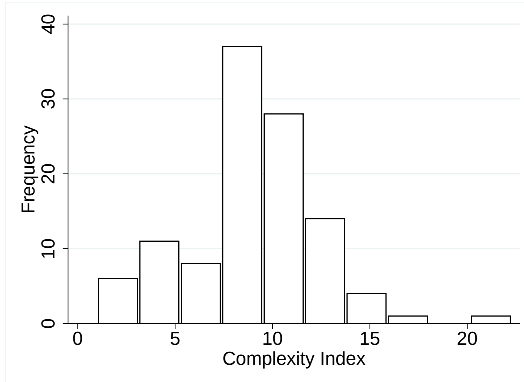


Figure A.5: Petroleum Administration for Defense Districts (PADD)

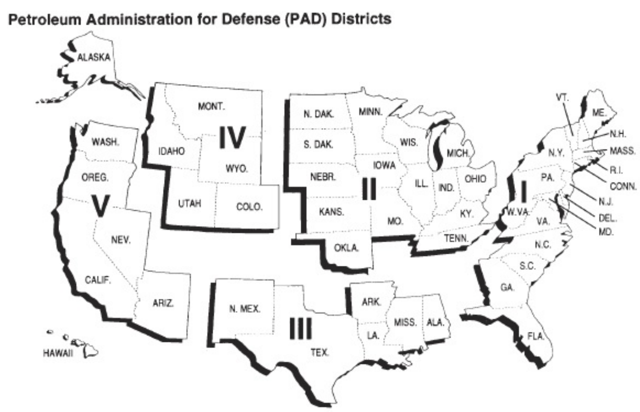
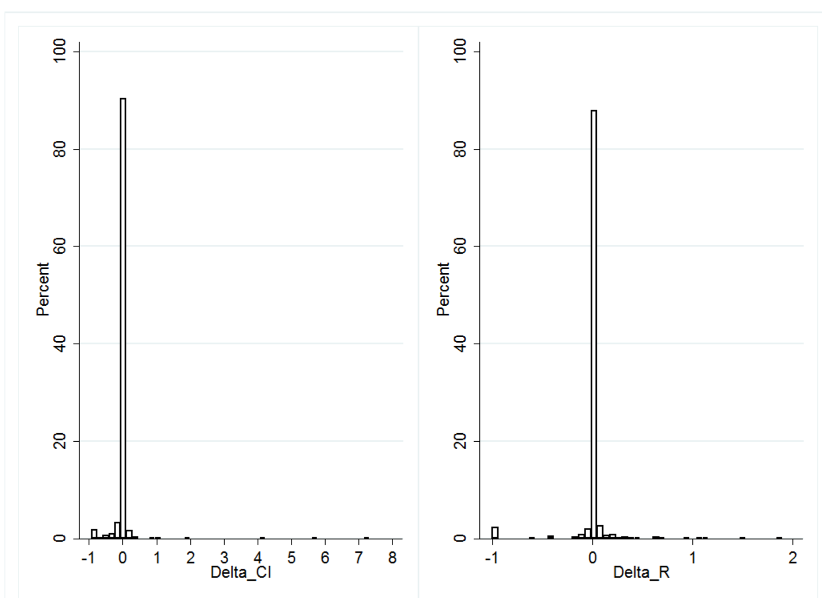


Figure A.6: Distribution of annual percentage change of complexity (left) and capacity (right) in the U.S. refining industry, 2008-2013.



A.1.3 Notes

Notes on Fact 3.b. I consider three samples of refineries: (i) all refineries, (ii) refineries located near the Gulf coast[54], (iii) refineries that are located within 40 kilometers (or 25 miles) to coastline.

I divide each of these samples into nine groups, as (small capacity, medium capacity, large capacity) $\times$ (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.6–A.8 report the results for each of the three samples. For example, take Table A.6 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 more than half of the trading relationships remain unexplained for 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.

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

Table A.6: 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.7: 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.8: 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*,

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

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,

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

$distance_T(x_1, x_2)$ 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.9: Distance in import shares

sample (i): All				*sample: (ii) Gulf*				*sample: (iii) Coastline*			
	R1	R2	R3		R1	R2	R3		R1	R2	R3
C1	.04	.05	.31	C1	0	.1	.34	C1	.16	.13	.67
C2	0	.13	.58	C2	0	.05	.69	C2	0	.27	.54
C3	0	.34	.65	C3	0	.35	.69	C3	0	.41	.66

A.2 Notes for Sections 3 and 4

A.2.1 Trade elasticity: identification and sample selection bias

Let $j = 0$ be the domestic supplier, then equation (4) implies:

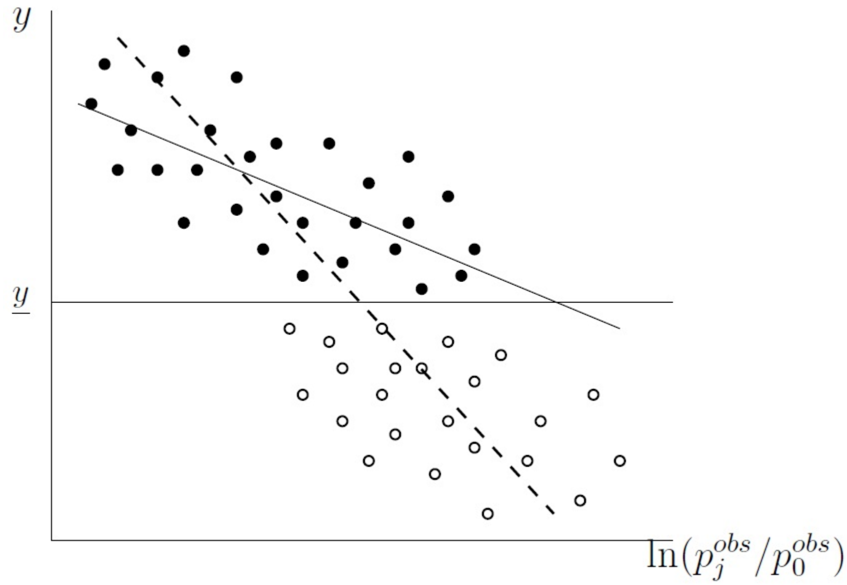
$$\underbrace{\ln \frac{q_j}{q_0}}_{y_j | j \in S} = -\eta \ln \frac{p_j^{obs}}{p_0^{obs}} - \eta \ln z_j, \qquad \text{if } j \in S. \tag{A.1}$$

The slope of $\ln(p_j^{obs}/p_0^{obs})$ identifies η if $E[\ln z_j \mid \ln p_j^{obs}/p_0^{obs}] = 0$. I continue to discuss that (i) this orthogonality condition does not hold, and (ii) using A.1 results in an under-estimation of η.

Start with the refiner's observed set S of suppliers. According to the model, $j \in S$ when the draw of z_j is favorable, i.e. only if z_j is smaller than a threshold that I call $\underline{z}$. (The construction of this threshold is explained by Proposition 1). For supplier $j \notin S$, consider a counterfactual case where j is added to S. In this counterfactual, the model predicts a trade quantity from j that I call q_j^{CF}, and a new quantity from the domestic supplier that I call q_0^{CF}. I define a variable, call it y_j, as follows: y_j equals $\ln(q_j/q_0)$ if $z_j \leq \underline{z}$, and $\ln(q_j^{CF}/q_0^{CF})$ if $z_j > \underline{z}$. Then, write a similar equation as (A.1) when $j \notin S$,

$$\underbrace{\ln \frac{q_j^{CF}}{q_0^{CF}}}_{y_j | j \notin S} = -\eta \ln \frac{p_j^{obs}}{p_0^{obs}} - \eta \ln z_j, \qquad \text{if } j \notin S. \tag{A.2}$$

Figure A.7: Identification and sample selection bias in estimating trade elasticity η. solid bullets: selected suppliers, circles: unselected suppliers. See the text for details.



Consider two suppliers j and j' with the same observable costs $p_j^{obs}/p_0^{obs} = p_{j'}^{obs}/p_0^{obs}$. Suppose the refiner has selected j while it has not selected j'. The fact that $j \in S$ and $j' \notin S$ means that $z_j < z_{j'}$. Thus, according to equations A.1-A.2, $y_j > y_{j'}$. That is, *selected supplies map to larger y's*.

Figure A.7 shows the selected and unselected suppliers in the space of y and $\ln(p_j^{obs}/p_0^{obs})$. For the sake of illustration, the figure is drawn by a simplification as if there is one threshold for all pairs of refiner-supplier's.[55] This simplified diagram illustrates the bias in estimating η when selections are taken as exogenous. Because *selected supplies map to larger y's*, the slope of the solid line is smaller than the slope of the dashed line. The smaller slope when there is a sample selection means an under-estimation of η.

A.2.2 Model fit

Table A.10 reports the model predictions versus data on the distribution of the number of import origins. It also shows the predictions according to the independent estimations. The median is 2 in the data, 2 according to the all-in-one, and 4 according to the independent estimations. The 99th

[55] In fact, for each refinery, there is a different $\underline{z}$, that can be constructed by Proposition 1. Then, holding the refiner fixed, for each supplier j, there is a threshold on y, denoted by $\underline{y_j}$, equal to $-\eta \ln(p_j^{obs}/p_0^{obs}) - \eta \ln \underline{z}$. For the sake of simplification, in the figure $\underline{y_j}$'s are the same.

percentile is 14 in the data, 12 according to the all-in-one, and 30 according to the the independent estimations.

Table A.10: 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

I compare my estimates with available data at the aggregate of the industry. Specifically, EIA reports the annual acquisition cost of crude oil for the U.S. refining industry, that is equal to the input cost per barrel of crude including transport costs and other fees paid by refineries. Notice that my data includes prices at origins and quantities at refineries, but not prices at refineries.

The model predicts the annual input cost of a refinery only when it is adjusted for complexity effect. As a result, I can not directly compare what my estimates predict with what EIA reports. According to my estimates, the average crude oil input costs (adjusted for complexity) equals 73.4 \$/bbl. To disentangle the effect of complexity, I set $\beta_{CI} = 0$, and re-do the simulation. Since the simulation excludes the effect of complexity, I consider its result as the unadjusted input cost.

According to the estimates and the above decomposition, average input cost excluding the effect of complexity, equals 75.5 \$/bbl. In the data, average input cost equals 76.7 \$/bbl. However, according to the results from separate estimations of selections and quantities , average input cost excluding the effect of complexity equals 59.7 \$/bbl which is far below 76.7 \$/bbl in the data.

Estimating trade flows by assuming exogenous selections delivers trade elasticity $\eta \approx 11$ instead of $\eta \approx 20$. The underestimation of η is the force behind the overestimation of the extent to which refineries diversify (Table A.10), and the overestimation of the gains from global sourcing (Table A.11).

Table A.11: Average input cost in the industry, accroding to the estimates and data, (dollars per bbl, 2010)

	average input cost (adjusted for complexity)	average input cost (not adjusted for complexity)
Data	–	76.7
All-in-one estimation	73.4	75.5
Separate estimation	58.5	59.7

A.3 Tables, figures, and notes for Sections 5 and 6

A.3.1 Tables

Table A.12: Country-level variables taken from data and sources of these data

Variable	Source	Units
Crude oil production	Energy Information Administration & The Oil and Gas Journal	barrels per day
Free on board prices of crude oil	Bloomberg	dollars per barrel
Total refining capacity	Energy Information Administration & The Oil and Gas Journal	barrels per stream day
Maximum refinery capacity	The Oil and Gas Journal	barrels per stream day
Refined oil consumption	Energy Information Administration	barrels per day (quantity) dollars (values)
Capacity of upgrading units	The Oil and Gas Journal	barrels per stream day
Utilization rate	World Oil and Gas Review published by Eni	
Refinery processing gain	Energy Information Administration	
Production of refined oil	Energy Information Administration	barrels per day (quantity) & dollars (values)
Share of expenditures on manufacturing and transportation	The World Input-Output Table	
Retail prices of gasoline and diesel	International Fuel Prices by German Agency for International Cooperation	cents per litre
Before tax prices of gasoline, diesel, fuel oil	Energy Prices and Taxes (International Energy Agency)	domestic currency per litre
Excise and value added taxes on gasoline, diesel, fuel oil	Energy Prices and Taxes (International Energy Agency)	domestic currency per litre
International trade flows of crude and refined oil	UN Comtrade Data	dollars (values)
Population	Penn World	
Years of schooling	Barro and Lee (2012)	
GDP	Penn World	US 2010 dollars

Table A.13: List of countries & selected variables related to oil production and consumption, 2010.

Country	Crude oil production (1000 b/d)	Total refining capacity (1000 b/d)	Avg complexity index	Avg utilization rate	Refined oil tax rate (%)	Refined oil consumption (1000 b/d)
Algeria	1540	450	1.34	0.89	-60	354
Angola	1899	39	1.79	0.72	-21	104
Azerbaijan	1035	399	3.89	0.30	-9	83
Brazil	2055	1908	4.28	0.87	81	2699
Canada	2741	2039	8.14	0.84	42	2283
China	4078	10521	2.73	0.88	30	8938
Colombia	786	413	4.67	0.69	54	270
France	0	1984	6.96	0.78	127	1833
Germany	0	2411	7.90	0.90	136	2467
India	751	3996	3.20	0.95	68	3305
Indonesia	953	1012	3.75	0.74	-25	1487
Iran	4080	1451	3.91	0.95	-91	1811
Iraq	2399	638	4.05	0.56	1	641
Italy	0	2337	6.87	0.69	117	1544
Japan	0	4624	7.84	0.75	75	4429
Kazakhstan	1525	345	5.25	0.65	-13	234
Korea	0	2702	4.98	0.81	94	2269
Kuwait	2300	936	5.02	0.92	-68	397
Libya	1650	378	1.57	0.86	-76	331
Mexico	2621	1540	7.62	0.86	20	2080
Netherlands	0	1206	7.52	0.81	143	1020
Nigeria	2455	505	4.43	0.36	9	283
Norway	1869	319	4.39	0.83	137	222
Oman	865	85	2.56	0.85	-48	150
Qatar	1129	339	4.25	0.85	-72	199
Russia	9694	5428	4.38	0.90	13	3135
Saudi Arabia	8900	2080	3.79	0.91	-82	2580
Singapore	0	1357	5.29	0.63	57	1149
Spain	0	1272	7.04	0.72	92	1441
UAE	2415	773	2.44	0.65	-8	615
United Kingdom	1233	1866	8.41	0.76	173	1620
United States	5471	17584	9.77	0.85	21	19180
Venezuela	2216	1282	5.41	0.80	-98	688
RO_America	1408	3022	4.79	0.74	6	2824
RO_Europe	662	5659	7.01	0.75	135	5219
RO_Eurasia	324	2032	4.51	0.48	0	877
RO_Middle East	1028	944	3.72	0.85	-55	1011
RO_Africa	2257	1906	3.11	0.75	-43	2464
RO_Asia & Oceania	2047	2882	3.84	0.77	60	5904

Table A.14: Percentage change to crude production and refining capacity of countries from 2010 to 2013

Country	production	capacity
Algeria	-3.6	0.0
Angola	-6.0	0.0
Azerbaijan	-15.6	0.0
Brazil	-1.5	0.5
Canada	22.2	-6.4
China	2.1	3.2
Colombia	27.7	1.7
France	0.0	-11.9
Germany	0.0	-6.8
India	2.8	53.2
Indonesia	-13.4	0.0
Iran	-21.6	0.0
Iraq	27.3	0.0
Italy	0.0	-6.1
Japan	0.0	2.9
Kazakhstan	2.8	0.0
Korea	0.0	9.5
Kuwait	15.2	0.0
Libya	-44.3	0.0
Mexico	-2.3	0.0

Country	production	capacity
Netherlands	0.0	-0.9
Nigeria	-3.5	-11.9
Norway	-18.2	0.0
Oman	8.7	0.0
Qatar	37.6	0.0
Russia	3.3	1.3
Saudi Arabia	8.8	1.5
Singapore	0.0	0.0
Spain	0.0	0.0
UAE	16.8	0.0
United Kingdom	-34.3	-10.0
United States	36.3	1.3
Venezuela	3.8	0.0
RO_America	-4.6	2.7
RO_Europe	-8.2	-1.7
RO_Eurasia	16.2	0.1
RO_Middle East	-77.7	0.0
RO_Africa	-12.5	0.0
RO_Asia & Oceania	-11.5	44.0
WORLD	2.2	3.3

A.3.2 Figures

Figure A.8: Calibrated utilization rates

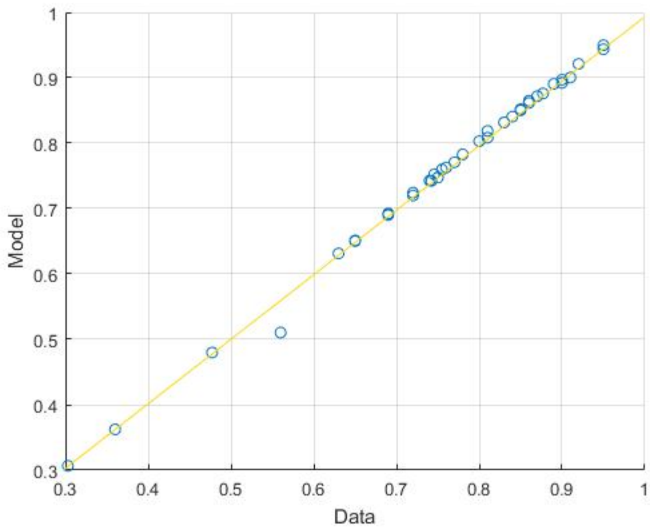
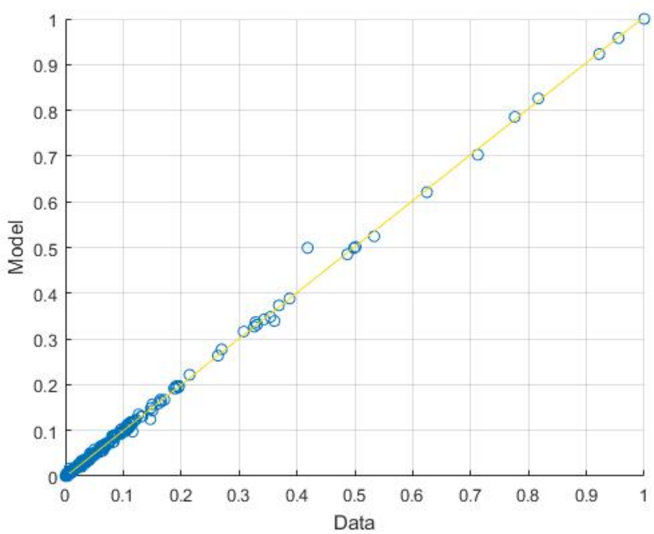


Figure A.9: Calibrated trade shares



A.3.3 Notes

Counterfactual trade costs from U.S. to elsewhere. According to the data, U.S. exported crude oil only to Canada in 2010, so $d_{n,USA} = \infty$ for all $n \neq$ Canada. To predict the after-lifting-ban trade costs, I use the estimates in Section 5.4 —which parametrizes the calibrated trade costs as a function of distance, common border, and fixed effects. 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. The estimates of refined oil trade costs highlight that among all countries/regions, barriers to export is the smallest for the United States (Panel B of Table 17). In addition, similar estimations in other researches show that U.S. has the smallest export barriers for manufactured products (see Table 3 in Waugh (2010) where he reports trade costs for a sample of 77 countries). In the absence of the export ban, it is then reasonable to assume that, relative to the other suppliers, U.S. faces small barriers for exporting crude oil. Accordingly, I let U.S. exporter fixed effect be equal to 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 remain at infinity.

Appendix B Proofs & mathematical derivations

B.1 Derivation of equation 5

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 $W(t) = \min_j \{p_j/\tilde{\epsilon}_j(t);\ j \in S\}$, where by a change of variable, $\tilde{\epsilon} \equiv 1/\epsilon$. Also, $Pr(\tilde{\epsilon}_j(t) \leq \tilde{\epsilon}) = \exp(-s\ \tilde{\epsilon}^{-\eta})$ where $s = \Big[\Gamma\Big(1 + 1/\eta\Big)\Big]^{\eta}$ ensuring that $E[1/\tilde{\epsilon}] \equiv E[\epsilon] = 1$. The probability distribution of random variable W is given by

$$
\begin{aligned}
G_W(w) &\equiv Pr(W \leq w) = 1 - Pr(W > w) \\
&= 1 - \prod_{j \in S} Pr(\tilde{\epsilon}_j < \frac{p_j}{w}) \\
&= 1 - \exp(-\Phi w^{\eta}),
\end{aligned}
$$

where $\Phi = s \sum_{j \in S} p_j^{-\eta} = \left[\Gamma\left(1 + 1/\eta\right)\right]^{\eta} \sum_{j \in S} p_j^{-\eta}$. The annual input cost, P, aggregates the flow of input costs over the entire period. Thus,

$$\begin{aligned} P &= \int_0^\infty w \, dG_W(w) \\ &= \int_0^\infty w \, \Phi \eta w^{\eta-1} \exp(-\Phi w^\eta) \, dw \\ &= \Gamma\left(1 + \frac{1}{\eta}\right) \Phi^{-1/\eta} \\ &= \left(\sum_{j \in S} p_j^{-\eta}\right)^{-1/\eta}. \end{aligned}$$

B.2 Notes on Proposition 1

B.2.1 Diminishing gains from adding suppliers

The variable profit function features decreasing differences if

$$\pi(L+1) - \pi(L) \geq \pi(L+2) - \pi(L+1), \quad \text{for } L = 1, ..., J-2.$$

I provide a sufficient condition under which the above inequality holds. The proof uses the calculus of continuous functions for dealing with the originally discrete functions. I define an *auxiliary problem* in which there is a continuum of suppliers $[0, J]$ on the real line; compared with the *original problem* in which there is a discrete number of suppliers $J \in \mathbb{N}_+ = \{1, 2, ...\}$. Variable x in the original problem has its counterpart x^{aux} in the auxiliary problem. $p^{aux}(\ell)$ denotes the cost of supplier ℓ where $\ell \in [0, J]$ is a real number. I choose p^{aux} such that (i) evaluated at integer numbers, p^{aux} equals p, i.e. $p^{aux}(1) = p(1)$, $p^{aux}(2) = p(2)$, ..., $p^{aux}(J) = p(J)$; (ii) $p^{aux}(\ell)$ is weakly increasing in ℓ by possible re-indexing; (iii) $p^{aux}(\ell)$ is continuous and differentiable. Note that (ii) and (iii) imply that $dp^{aux}(\ell)/d\ell$ is well-defined and positive.

In the auxiliary problem, a refiner's decisions reduce to choosing L suppliers[56] with noting that L can be a real number. For a refiner that buys from the measure L of the lowest cost suppliers, define $u^{aux}(L)$ as the utilization rate, $C^{aux}(L) \equiv C(u^{aux}(L))$ as the utilization cost, and $y(L) \equiv$

[56]This is implied by a straightforward generalization of Result 1 in Section 3.3.2 to the auxiliary problem.

$C'(u)|_{u=u^{aux}(L)}$ as the marginal cost of utilization. F.O.C implies that

$$y(L) = \tilde{P} - P^{aux}(L) = \tilde{P} - \Big[\int_0^L p^{aux}(\ell)^{-\eta}\, d\ell\Big]^{\frac{-1}{\eta}}. \tag{B.1}$$

W.l.o.g. I normalize refiner's capacity, $R = 1$. Variable profit, denoted by $\pi^{aux}(\ell)$, is given by

$$\begin{aligned}
\pi^{aux}(L) &= u^{aux}(L)(\tilde{P} - P^{aux}(L)) - C^{aux}(L) \\
&= u^{aux}(L)y(L) - C^{aux}(L).
\end{aligned}$$

Since by definition, $y(L) = \tilde{P}/[\lambda(1 - u^{aux}(L))^2]$, hence $u^{aux}(L) = 1 - \tilde{P}^{1/2}\lambda^{-1/2}y(L)^{-1/2}$. Then, variable profit as a function y is given by

$$\pi^{aux}(L) = y(L) - 2\Big(\frac{y(L)\tilde{P}}{\lambda}\Big)^{1/2} + \frac{\tilde{P}}{\lambda} \tag{B.2}$$

Now, consider the following lemma.

Lemma B.1. If the auxiliary variable profit function π^{aux} is concave, then the original variable profit function π features decreasing differences.

Proof. If π^{aux} is concave, then

$$\frac{\pi^{aux}(a) + \pi^{aux}(b)}{2} \le \pi^{aux}(\frac{a+b}{2}), \quad a, b \in [0, J] \ \text{(on the real line)}.$$

One special case of the above relation is where $a = L$ and $b = L + 2$ with L being an integer between 1 and $J - 2$. Evaluated at integers, the variables of the auxiliary problem equal to their counterparts in the original problem. Therefore, $\pi^{aux}(L) = \pi(L)$, $\pi^{aux}(L+1) = \pi(L+1)$, and $\pi^{aux}(L+2) = \pi(L+2)$. The above inequality, then, implies

$$\frac{\pi(L) + \pi(L+2)}{2} \le \pi(L+1) \Leftrightarrow \pi(L+1) - \pi(L) \ge \pi(L+2) - \pi(L+1); \quad \ell = 1, 2, ..., J-2$$

which is the definition of decreasing differences. □

According to lemma B.1, to show π features decreasing differences, it suffices to show $(\pi^{aux})'' \equiv$

$d^2\pi^{aux}(L)/dL^2 \leq 0$. By taking derivatives with respect to L in equation (B.2),

$$(\pi^{aux})''(L) = y''(L)\Big(1 - \tilde{P}(L)^{1/2}\lambda^{-1/2}y(L)^{-1/2}\Big) + \frac{1}{2}(y'(L))^2\tilde{P}(L)^{1/2}\lambda^{-1/2}y(L)^{-3/2}. \quad \text{(B.3)}$$

Using equation (B.1), I calculate $y'(L)$ and $y''(L)$,

$$y'(L) = \frac{1}{\eta}\Big[\int_0^L p^{aux}(\ell)^{-\eta}\, d\ell\Big]^{\frac{-1}{\eta}-1} p^{aux}(\ell)^{-\eta} \quad \text{(B.4)}$$

$$\begin{aligned} y''(\ell) &= \frac{-(1+\eta)}{\eta^2}\Big[\int_0^\ell p^{aux}(\ell)^{-\eta}\, d\ell\Big]^{\frac{-1}{\eta}-2} p^{aux}(L)^{-2\eta} \\ &\quad -\Big[\int_0^L p^{aux}(\ell)^{-\eta}\, d\ell\Big]^{\frac{-1}{\eta}-1} p^{aux}(L)^{-\eta-1}(p^{aux})'(L) \end{aligned} \quad \text{(B.5)}$$

It is straightforward to check that $y' > 0$ and $y'' < 0$. Equation (B.3) implies that $(\pi^{aux})'' \leq 0$ if and only if

$$\frac{(y')^2}{-y''} \leq \frac{2(1 - \tilde{P}^{1/2}\lambda^{-1/2}y^{-1/2})}{\tilde{P}^{1/2}\lambda^{-1/2}y^{-3/2}} = 2y(\tilde{P}^{-1/2}\lambda^{1/2}y^{1/2} - 1) \quad \text{(B.6)}$$

Since by construction $(p^{aux})' \geq 0$, it follows from equation B.5 that,

$$-y''(L) \geq \frac{(1+\eta)}{\eta^2}\Big[\int_0^L p^{aux}(\ell)^{-\eta}\, d\ell\Big]^{\frac{-1}{\eta}-2} p^{aux}(L)^{-2\eta}$$

Using the above inequality as well as equations B.4–B.5,

$$\frac{(y')^2}{-y''} \leq \frac{\Big\{\frac{1}{\eta}\Big[\int_0^L p^{aux}(\ell)^{-\eta}\Big]^{\frac{-1}{\eta}-1} p^{aux}(L)^{-\eta}\Big\}^2}{\frac{(1+\eta)}{\eta^2}\Big[\int_0^L p^{aux}(\ell)^{-\eta}\Big]^{\frac{-1}{\eta}-2} p^{aux}(L)^{-2\eta}} = \frac{\Big[\int_0^L p^{aux}(\ell)^{-\eta}\Big]^{\frac{-1}{\eta}}}{1+\eta} = \frac{P^{aux}}{1+\eta} \quad \text{(B.7)}$$

Using (B.6) and (B.7), a sufficient condition for $(\pi^{aux})'' \leq 0$ is

$$\frac{P^{aux}}{(1+\eta)} \leq 2y(\tilde{P}^{-1/2}\lambda^{1/2}y^{1/2} - 1). \quad \text{(B.8)}$$

I replace for $y = \tilde{P} - P^{aux}$, define $\kappa \equiv \tilde{P}/P^{aux}$, and rearrange the terms in inequality B.8,

$$\frac{1 + 2(1+\eta)(\kappa - 1)}{2(1+\eta)(\kappa-1)(\frac{\kappa-1}{\kappa})^{1/2}} \leq \lambda^{1/2} \tag{B.9}$$

Inequality B.9 is a sufficient condition for $(\pi^{aux})'' < 0$. I relate this condition to observed data. By F.O.C.,

$$\tilde{P} - P^{aux} = \frac{\tilde{P}}{\lambda(1-u^{aux})^2}$$

implying that

$$\lambda = \frac{\kappa}{(\kappa-1)} \frac{1}{(1-u^{aux})^2} \geq \frac{\kappa}{(\kappa-1)} \frac{1}{(1-u_{\min})^2},$$

where $u_{\min}$ is the observed minimum utilization rate in the data. Combining the above relation with inequality B.9,

$$\frac{1 + 2(1+\eta)(\kappa - 1)}{2(1+\eta)(\kappa-1)} \leq \frac{1}{1-u_{\min}}$$

or, equivalently

$$\eta \geq \frac{1-u_{\min}}{2(\kappa-1)u_{\min}} - 1 \tag{B.10}$$

Note that $P \leq p_0$, where p_0 is the cost of the domestic supplier. This is true because refineries always buy domestically and the annual input cost decreases by adding new suppliers. Thus, $\kappa \equiv \tilde{P}/P \geq \tilde{P}/p_0$. In the data $\tilde{P}/p_0 = 1.174$ and $u_{\min} = 0.52$. A simple calculation shows that as long as $\eta \geq 1.65$, inequality B.10 holds —or equivalently, inequality B.9 holds, or equivalently the variable profit function in the original problem features decreasing differences.

B.2.2 Notes on Proposition 1: Constructing the lower bound $\underline{p}_B$

Let $y \equiv C'$. Then, variable profit is given by $\pi = R(y - 2(\tilde{P}y/\lambda)^{1/2})$. Let $\tilde{y} \equiv y^{1/2}$. Then,

$$\tilde{y}^2 - 2(\tilde{P}/\lambda)^{1/2}\tilde{y} - \pi/R = 0$$

Since $\tilde{y} > 0$, the above equation has only one qualified root,

$$\tilde{y} = \sqrt{\frac{\tilde{P}}{\lambda}} + \sqrt{\frac{\tilde{P}}{\lambda} + \frac{\pi}{R}}$$

which then implies a mapping between the marginal cost of utilization y and variable profit π:

$$y = \frac{2\tilde{P}}{\lambda}\Big(1 + \sqrt{1 + \frac{\pi\lambda}{\tilde{P}R}}\Big) + \frac{\pi}{R} \tag{B.11}$$

Consider a counterfactual sourcing in which the refiner adds a new supplier to its sourcing set. I use superscript *new* for variables associated with this hypothetical sourcing. Particularly, equation B.11 implies: $y^{new} = \frac{2\tilde{P}}{\lambda}\Big(1 + \sqrt{1 + \frac{\pi^{new}\lambda}{\tilde{P}R}}\Big) + \frac{\pi^{new}}{R}$. The maximum variable profit such that adding a supplier is still not profitable is achieved at $\pi^{new} = \pi + f$. It is at this maximum that we can find the lower bound $\underline{p}_B$ (that is, if the cost of an unselected supplier is below $\underline{p}_B$, it would be profitable to add that supplier).

On the one hand, let $\underline{P}$ be the lower bound on P^{new} associated with adding a supplier with cost $\underline{p}_B$. Since, by F.O.C., $P^{new} = \tilde{P} - y^{new}$, we get

$$\underline{P} = \tilde{P} - \frac{2\tilde{P}}{\lambda}\Big(1 + \sqrt{1 + \frac{(\pi + f)\lambda}{\tilde{P}R}}\Big) + \frac{\pi + f}{R}$$

On the other hand, using equation (5), $\underline{P} = \Big[\sum_{j\in S} p_j^{-\eta} + \underline{p}_B^{-\eta}\Big]^{\frac{-1}{\eta}} = \Big[P^{-\eta} + \underline{p}_B^{-\eta}\Big]^{\frac{-1}{\eta}}$, which implies:

$$\underline{p}_B = \Big[\underline{P}^{-\eta} - P^{-\eta}\Big]^{\frac{-1}{\eta}}.$$

Finally, note that by Result 1, the added supplier must not be cheaper than any selected supplier. In case $\underline{p}_B \leq \max\{p_j;\ j \in S\}$, replace $\underline{p}_B$ with $\max\{p_j;\ j \in S\}$.

B.2.3 Re-stating Proposition 1

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:

$$
\begin{array}{ll}
(R.1) & \Big\{Q_A = q_A \mid Q_A > 0,\ Q_B = 0\Big\} \longleftrightarrow \Big\{X_A = h(q_A) \mid Q_A > 0,\ Q_B = 0\Big\} \\
(R.2) & \Big\{Q_A > 0,\ Q_B = 0 \mid X_A = x_A,\ F = f\Big\} \longleftrightarrow \Big\{Z_B \geq \underline{z}_B(x_A, f)\Big\}\ \mathit{and}\ \Big\{F \leq \bar{f}(x_A)\Big\}
\end{array}
$$

B.3 Notes on Proposition 2

B.3.1 Proof of Proposition 2

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

$$
\begin{aligned}
L &= g_{Q_A}\Big(q_A \mid S \text{ is selected}\Big) \times Pr\Big\{S \text{ is selected}\Big\} \\
&= g_{Q_A}\Big(q_A \mid Q_A > 0,\ Q_B = 0\Big) \times Pr\Big\{Q_A > 0,\ Q_B = 0\Big\} \\
&= \Big|\partial h(q_A)/\partial q_A\Big| \times g_{X_A}\Big(h(q_A) \mid Q_A > 0,\ Q_B = 0\Big) \times Pr\Big\{Q_A > 0,\ Q_B = 0\Big\} \\
&= \Big|\partial h(q_A)/\partial q_A\Big| \times g_{X_A}\Big(h(q_A)\Big) \times Pr\Big\{Q_A > 0,\ Q_B = 0 \mid X_A = h(q_A)\Big\} \\
&= \Big|\partial x_A/\partial q_A\Big| \times g_{X_A}(x_A) \times Pr\Big\{Q_A > 0,\ Q_B = 0 \mid X_A = x_A\Big\}.
\end{aligned} \tag{B.12}
$$

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 f.o.b. price, and $|S|$ is the number of suppliers in S. So, 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 in (B.12), I use the first relation in Proposition 1, (R.1). Suppose that w.l.o.g. h is strictly increasing.[57] Then,

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

Taking derivatives with respect to q_A delivers the result:

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

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 equation (B.12) as follows:

$$\begin{aligned} Pr\Big(Q_A > 0,\ Q_B = 0 \mid X_A = x_A\Big) &= \int_0^\infty Pr\Big(Q_A > 0,\ Q_B = 0 \mid X_A = x_A,\ F = f\Big)\ dG_F(f|\mu_f,\sigma_f) \\ &= \int_0^\infty Pr\Big(Z_B \geq \underline{z}_B(x_A,f)\Big) \times I\Big(f \leq \bar{f}(x_A)\Big)\ dG_F(f|\mu_f,\sigma_f) \\ &= \int_0^{\bar{f}(x_A)} \ell_B(x_A,f)\ dG_F(f|\mu_f,\sigma_f), \end{aligned} \tag{B.13}$$

where $I\Big(f \leq \bar{f}(x_A)\Big)$ is an indicator function to be equal one only if $f \leq \bar{f}(x_A)$; and by definition, $\ell_B(x_A,f) = Pr\Big\{Z_B \geq \underline{z}_B(x_A,f)\Big\}$. Plugging (B.13) into equation (B.12),

$$L = \Big|\partial x_A/\partial q_A\Big| \times g_{X_A}(x_A) \times \int_0^{\bar{f}(x_A)} \ell_B(x_A,f)\ dG_F(f|\mu_f,\sigma_f). \tag{B.14}$$

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

$$g_{X_A}(x_A) = \Big|\partial[\lambda, z_A]/\partial q_A\Big| \times g_\lambda(\lambda)\prod_{j\in S} g_Z(z_j), \tag{B.15}$$

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 $[\lambda, z_A]$

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

is a vector with $|S|$ random variables). It follows that

$$L = |\partial[\lambda, z_A]/\partial q_A| \times g_\lambda(\lambda) \prod_{j\in S} g_Z(z_j) \times \int_0^{\bar{f}(\lambda,z_A)} \ell_B(\lambda, z_A, f)\, dG(f|\mu_f, \sigma_f). \tag{B.16}$$

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

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

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

B.4 A simple closed economy with one supplier and homogeneous refineries

This section presents a simplified version of the main model in the text. I analytically show the effect of a change in this economy (such as a boom in crude oil production) on the prices of crude and refined oil.

There is one country with a measure one of homogeneous refineries each with capacity R; and one supplier with inelastic production Q. In this economy $Q < R$, and there are no trade costs for either crude or refined oil. Let p denote the price of crude oil at the location of supplier. Let $\tilde{P}$ be the price of the composite refinery output at the location of refineries. With m denoting the productivity of the retail sale of refined oil products, the price index of refined oil products is given by $e = \tilde{P}/\tilde{m}$. That is, e is the price of refined oil products at the location of end-users. [58]

Let Y, w, and L denote GDP, wage, and population. Then, $Y = wL + pQ$.[59] Consumers spend α share of their income on manufacturing sector. Manufacturing producers spend $1-\beta$ share of their expenditures on oil products. Rewriting Eq. (20) from the main text for this simple economy,

$$1 - \beta = \frac{b^{-\rho} e^{1-\rho}}{(1-b)^{-\rho} w^{1-\rho} + b^{-\rho} e^{1-\rho}} \tag{B.17}$$

$$= \frac{\tilde{b} e^{1-\rho}}{w^{1-\rho} + \tilde{b} e^{1-\rho}} \tag{B.18}$$

[58] In relation with the notation in the main text, $\tilde{m} = (m^e)^{1/\theta^e}/\gamma^e$

[59] I assume no taxes on oil consumption. Also, the profit of the refining sector is dropped here as it accounts for a negligible share of GDP.

Here, $\tilde{b} = [b/(1-b)]^{-\rho}$. The market clearing condition for oil products is given by

$$\alpha(1-\beta)(wL + pQ) = eQ. \tag{B.19}$$

By equations B.17-B.19,

$$\frac{Q}{wL + pQ} = \frac{\alpha\tilde{b}e^{-\rho}}{w^{1-\rho} + \tilde{b}e^{1-\rho}} \tag{B.20}$$

On the side of demand for crude oil, refinery utilization cost equals $\tilde{P}c(u)R$. Refinery's problem is to choose utilization rate u to maximize $(\tilde{P} - p)uR - \tilde{P}c(u)R$. By F.O.C.,

$$\tilde{P}c'(u) = \tilde{P} - p \tag{B.21}$$

Also, by market clearing condition for crude oil $Q = uR$. It is assumed that $c'(Q/R) < 1$. Therefore,

$$p = \mu e, \quad \text{where } \mu \equiv m(1 - c'(Q/R)) \tag{B.22}$$

In this model, Q, R, and L are exogenous variables; α, b, $\tilde{b}$, ρ, and m, are known parameters; p, $\tilde{P}$, e, and β are endogenous variables. Treating labor as the unit of numeraire, I normalize $w = 1$.

The effect of a change in crude oil production on the prices of crude and refined. I calculate how a change in Q changes p and e. According to equations B.19-B.20,

$$Qw^{1-\rho} + Q\tilde{b}e^{1-\rho} - wL\alpha\tilde{b}e^{-\rho} - \mu Q\alpha\tilde{b}e^{1-\rho} = 0$$

By some algebra,

$$\frac{dQ}{Q} = -\tilde{\rho}\frac{de}{e} \tag{B.23}$$

where

$$\tilde{\rho} = \frac{(1-\beta)(1-\alpha\mu)}{\beta}(1-\rho) + \frac{1-\alpha\mu(1-\beta)}{\beta}\rho$$

Note that $\tilde{\rho} \to \rho$ as $\beta \to 1$. The elasticity of refined oil price e with respect to production Q converges to $-1/\rho$ when share of spending on oil goes to zero.

Using equation B.22 and B.23,

$$dp = m\Big[1 - c'(u) + \tilde{\rho}uc''(u)\Big]de \tag{B.24}$$

Since $p = m(1 - c'(u))e$ (Eq. B.22),the above results:

$$\frac{dp}{p} = \frac{de}{e}\Big[1 + \underbrace{\frac{\tilde{\rho}uc''(u)}{1 - c'(u)}}_{Buffer}\Big] \tag{B.25}$$

Here, *Buffer* is the portion of the shock that is absorbed by refineries. As $1 - c'(u) > 0$, $c''(u) > 0$,

$$\Big|\frac{dp}{p}\Big| > \Big|\frac{de}{e}\Big|$$

In addition, equation B.25 implies an assymetric response to changes in utilization rate. When there is an increase in production, i.e. $dQ/Q > 0$, then $0 > de/e > dp/p$. But, when there is a decrease in production, i.e. $dQ/Q < 0$, then $0 < de/e < dp/p$. This feature arises because of the convexity of utilization costs.

Appendix C Numerical and computational algorithms

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.[60]

For numerical integration, I use the method of Quasi Monte Carlo.[61] I generate Neiderreiter equidistributed sequence of nodes $U^z = (U_j^z)_{j=1}^J \in [0,1]^J$ for the vector of trade cost shocks

[60] 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.

[61] See Miranda and Fackler, Chapter. 5

$z = (z_j)_{j=1}^{J}$, $U^{\lambda} \in [0,1]$ 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 $T = 10,000$ vectors $U \equiv (U^z, U^{\lambda}, U^f, U^R)$, save all U's, and keep them fixed through the simulation.

Trade cost shock with respect to supplier j, z_j has a Frechet 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 Frechet distribution, $z_j = \Big(-\log(U_j^z)/s_z\Big)^{-1/\theta}$. 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 Simulation

The following algorithm describes the steps that I take to solve for equilibrium.

1. Guess crude oil prices at the location of suppliers, p_j^{origin} $\forall j$.

2. Inner Loop. Given crude oil prices p_j^{origin} $\forall j$, solve for composite output prices $\tilde{P}_n$ $\forall n$:

   2.a) Guess $\tilde{P}_n$ $\forall n$.

   2.b) Given prices of crude oil p_j $\forall j$, and of composite output $\tilde{P}_n$ $\forall n$, solve the refiner's problem for every individual refinery x in country n:

   Holding this individual refinery fixed,

   - Sort suppliers based on $p_j = p_j^{origin}(1 + d_j + \zeta_j)z_j$, i.e. average suppliers' costs at the location of the refinery.
   - Calculate total refinery profit when the refiner buys from the first L lowest cost suppliers for $L = 1, ..., J$. Then find optimal sourcing.
   - Save the refiner's optimal set of suppliers $S_n(x)$, input price index $P_n(x)$, utilization rate $u_n(x)$, trade shares $k_{nj}(x)$, trade quantities $q_{nj}(x) = u_n(x)k_{nj}(x)$, quantity of composite output $\tilde{q}_n(x) = u_n(x)R$, and utilization costs $C_n(u_n(x))R$.

2.c) Using results from 2.b, compute aggregate crude oil purchases by refineries in country n from supplier j, Q_{nj}, aggregate composite refinery output $\tilde{Q}_n$, aggregate costs of utilization $\tilde{C}_n$, and aggregate fixed costs $\tilde{F}_n$. The RHS of equation (23) delivers the net supply of refinery output $\tilde{H}_n^S = \tilde{P}_n \tilde{Q}_n - \tilde{C}_n - \tilde{F}_n$.

2.d) For every pair of countries n and i, compute trade shares of refined oil, π^e_{ni}, price index of refined oil products, e_n, share of spendings on oil products $1 - \beta_n$, gdp Y_n, and expenditures on oil products Y^e_n. The LHS of equation (23) delivers global demand for composite refinery output of country n, $\tilde{H}_n^D = \sum_{k=1}^N \pi^e_{kn} Y^e_k / (1 + t^e_k)$.

2.e) Calculate the excess demand function for refinery composite output $\tilde{H}_n = \tilde{H}_n^D - \tilde{H}_n^S$. If $|\tilde{H}_n / \tilde{H}_n^D| < \varepsilon$, then skip the rest of the Inner Loop and go to step 3. Otherwise, go on to step 2.f.

2.f) Construct a Jacobian matrix of the excess demand function $\tilde{\mathbf{J}} = [\tilde{H}'_{ni}]$ where $\tilde{H}'_{ni} = \partial \tilde{H}_n / \partial \tilde{P}_i$. I derive the Jacobian analytically and do not use approximations.

2.g) Update the guess for $\tilde{\mathbf{P}} = [\tilde{P}_n]_{n=1}^N$,

$$\tilde{\mathbf{P}} \longleftarrow \tilde{\mathbf{P}} - \tilde{\mathbf{J}} \tilde{\mathbf{H}}^{-1},$$

and, go to Step 2.b.

3. For crude oil supplier j, compute world demand $Q_j^D = \sum_n Q_{nj}$, and world excess demand $H_j = Q_j^D - Q_j$. If $|H_j / Q_j^D| < \varepsilon$, then the algorithm ends. Otherwise, go on to step 4.

4. Construct a Jacobian matrix for excess crude oil demand function, $\mathbf{J} = [H'_{jk}]$, where $H'_{jk} = \partial H_j / \partial p_k$.

5. Update the guess for $\mathbf{p^{origin}} = [p_j^{origin}]_{j=1}^J$

$$\mathbf{p^{origin}} \longleftarrow \mathbf{p^{origin}} - \mathbf{J}\mathbf{H}^{-1},$$

then, go to Step 2, that is the beginning of the Inner Loop.[62] □

[62] In practice, I use dampening for updating my guess, that is,

$$p^{origin} \longleftarrow \lambda(p^{origin} - H' H^{-1}) + (1 - \lambda) p^{origin},$$

Because the trade elasticity $\eta \approx 20$ which is very high, the slope of the surface of the excess demand function is very small. For this reason, I needed to choose a small value for dampening. In practice I set $\lambda = 0.01$ which makes the algorithm to some extent slow but ensures the convergence.

C.3 Calibration

As Section 5.3.2 in the main text describes, I conduct four steps to calibrate the multi-country framework. Here I describe more details for Steps 2 and 4.

C.3.1 Details of Steps 2 in Calibration

Equations to be used in the calibration of factor intensities. Define $\tilde{Y}_n = w_n L_n + O_n$ as the sum of wages and oil revenues. Recall that gdp equals $Y_n = \tilde{Y}_n + \text{Taxes}_n$. Here, because of the lack of reliable data, I abstract away from profits of the refining sector which is, in any event, only a tiny component of gdp. From every $1 + t_n$ dollars spent on refined oil products, 1 dollar is paid to sellers and t_n dollars to the tax authority. Moreover, spendings on refined oil products equal $Y_n^e = \alpha_n(1-\beta_n)Y_n$. So, $\text{Taxes}_n = \frac{t_n}{1+t_n}\alpha_n(1-\beta_n)Y_n$. By plugging taxes into gdp, and re-arranging the terms:

$$Y_n = \frac{\tilde{Y}_n}{1 - \frac{t_n}{1+t_n}\alpha_n(1-\beta_n)}$$

On the one hand, notice that $1 - \beta_n = \frac{Y_n^e}{\alpha_n Y_n}$, whereas Y_n itself depends on β_n according to the above equation. Solving for β_n:

$$\begin{aligned}
1 - \beta_n &= \frac{Y_n^e}{\alpha_n Y_n} \\
&= \frac{1 - \frac{t_n}{1+t_n}\alpha_n(1-\beta_n)}{\alpha_n(\tilde{Y}_n/Y_n^e)} \\
&= \frac{1 + t_n - t_n\alpha_n(1-\beta_n)}{(1+t_n)\alpha_n(\tilde{Y}_n/Y_n^e)} \\
\Longrightarrow (1-\beta_n)(1+t_n)\alpha_n(\tilde{Y}_n/Y_n^e) &= 1 + t_n - t_n\alpha_n(1-\beta_n) \\
\Longrightarrow (1-\beta_n)\Big[(1+t_n)\alpha_n(\tilde{Y}_n/Y_n^e) + t_n\alpha_n\Big] &= 1 + t_n \\
\Longrightarrow (1-\beta_n)\alpha_n\Big[(\tilde{Y}_n/Y_n^e) + \frac{t_n}{1+t_n}\Big] &= 1 \\
\Longrightarrow 1 - \beta_n &= \frac{1}{\alpha_n\Big[(\tilde{Y}_n/Y_n^e) + \frac{t_n}{1+t_n}\Big]} \qquad \text{(C.1)}
\end{aligned}$$

On the other hand, by cost minimization (equation 20),

$$\begin{aligned}\beta_n &= \frac{b_n^{\rho} w_n^{1-\rho}}{b_n^{\rho} w_n^{1-\rho} + (1-b_n)^{\rho}[(1+t_n)e_n]^{1-\rho}} \\ &= \frac{1}{1+\left(\frac{1-b_n}{b_n}\right)^{\rho}\left(\frac{(1+t_n)e_n}{w_n}\right)^{1-\rho}},\end{aligned}$$

which implies that

$$\hat{b}_n \equiv \left(\frac{1-b_n}{b_n}\right)^{\rho} = \frac{1-\beta_n}{\beta_n}\left[\frac{(1+t_n)e_n}{w_n}\right]^{-(1-\rho)} \tag{C.2}$$

Given a set of country-level data, and parameter ρ (elasticity of substitution between refined oil and labor), I follow three steps to calibrate $\hat{b}_n$ or equivalently b_n. Specifically, the country-level data I use here consist of wage w_n, human capital adjusted population L_n, oil revenues $O_n = \sum_{\tau} p_{n\tau} Q_{n\tau}$, ad valorem equivalent tax rate on refined oil consumption $t_n \in (-1, \infty)$, share of spendings on the oil-intensive sector (manufacturing and transportation) α_n, price of refined oil products e_n, and aggregate consumption of refined oil products Y_n^e. The three steps are:

1. Calculate $\tilde{Y}_n = w_n L_n + O_n$.
2. Plug $\tilde{Y}_n$, Y_n^e, α_n, and t_n into equation (C.1) to compute β_n.
3. Plug β_n, w_n, e_n, and t_n into equation (C.2) to compute $\hat{b}_n$ and b_n.

C.3.2 Details of Step 4 of Calibration

In the calibration procedure, crude oil prices at the location of suppliers, p_j^{origin} $\forall j$ are given by data. The following steps describe the calibration algorithm:

1. Guess trade costs d_{ni}, efficiency of utilization costs $\mu_{\lambda,n}$, and efficiency in retail sale of refined oil products m_n^e.
2. Solve for $\tilde{P}_n$. This step is an inner loop that is the same as Step 2 in the simulation algorithm.

- (At the current value of $\tilde{P}_n$ for refiner x in country n) This step solves for set of suppliers $S_n(x)$, input price index $P_n(x)$, utilization rate $u_n(x)$, trade shares $k_{nj}(x)$, trade quantities $q_{nj}(x) = u_n(x)k_{nj}(x)$, quantity of composite output $\tilde{q}_n(x) = u_n(x)R$, and utilization costs $C_n(u_n(x))R$.

3. Using the output of step 2, compute the aggregate crude oil demand by refineries in n for j, Q_{nj}, and total crude oil demand by refineries in n denoted by $Q_{n\star} \equiv \sum_j Q_{nj}$. Calculate $\pi_{nj} = Q_{nj}/Q_{n\star}$ as the share of supplier j in country n's crude oil purchases. Compute the ratio of the price of refinery composite output relative to the average price of crude oil at the location of refineries in country n denoted by $r_n \equiv \tilde{P}_n/P_n^{avg}$.

4. If $|\frac{\pi_{nj}}{\pi_{nj}^{data}} - 1| < \varepsilon$, $|\frac{Q_{n\star}}{Q_{n\star}^{data}} - 1| < \varepsilon$, and $|\frac{r_n}{r_n^{data}} - 1| < \varepsilon$, the algorithm ends. Otherwise, go on to the next step.

5. Update d_{ni}, $\mu_{\lambda,n}$, and m_n^e:

$$d_{ni} \longleftarrow d_{ni} \times \left(\frac{\pi_{nj}}{\pi_{nj}^{data}}\right)^{\kappa_1}$$

$$\mu_{\lambda,n} \longleftarrow \mu_{\lambda,n} \times \left(\frac{Q_{n\star}}{Q_{n\star}^{data}}\right)^{-\kappa_2}$$

$$m_n^e \longleftarrow m_n^e \times \left(\frac{r_n}{r_n^{data}}\right)^{-\kappa_3}$$

where $\kappa_1 > 0$, $\kappa_2 > 0$, and $\kappa_3 > 0$ govern the speed of convergence. Then, go to step 2. □

Intuitively, when predicted trade share π_{nj} is larger than its actual value π_{nj}^{data}, then I increase my guess of the value of trade costs between supplier j and country n, d_{nj}, in order to push down the predicted trade share toward its actual value. If predicted total demand for crude oil in country n, $Q_{n\star}$, is larger than its actual value $Q_{n\star}^{data}$, then I push down the demand by country n's refineries by decreasing the mean of their log efficiency in costs of capacity utilization. If predicted relative price of refinery output to crude oil, r_n, is larger than its actual value, r_n^{data}, then I decrease the efficiency in the retails sale of refined oil products m_n^e. Because crude oil prices are given by data here, a smaller m_n^e means pushing down the supply of refined oil products, and so, pushing down demand for refinery output. This effect makes the price of the composite refiner output to be lower in the next iteration. In practice, I set $\kappa_1 = 0.01$, $\kappa_2 = 0.10$, and $\kappa_3 = 5$.

C.4 Monte Carlo Analysis

I perform a Monte Carlo simulation to evaluate the ability of my estimation procedure to recover model parameters. A basic finding is that the estimation procedure is capable of recovering parameters with standard errors similar to those of the main estimation results.

I simulate artificial data using the "true" estimated parameters in Section 4, with the model of Section 3. For the simulated data, I run my procedure to estimate parameters, then compare them with the true parameters. I perform this exercise for 50 times. Each time, the true estimates and the estimation procedure remain fixed, whereas the artificial dataset varies because realizations of unobservable draws change.

Table C.1 reports the results. Columns "mean" and "std dev" show the average and standard deviation of estimates across 50 exercises. Comparing with the main results reported in Table 1, for each parameter, the mean is in a close distance to the true parameter, and the standard deviation is similar to that of the main estimate.

Table C.1: Monte Carlo Simulation Results

description	parameter	true	mean	std dev
trade elasticity	η	19.77	19.67	3.11
dispersion in trade costs	θ	3.16	3.20	0.47
distance coefficient	γ_d	0.02	0.02	0.01
border coefficient	γ_b	0.72	0.73	0.06
complexity coefficient	β_{CI}	-0.03	-0.03	0.01
mean of $\ln \lambda$	μ_λ	5.45	5.48	0.12
standard deviation of $\ln \lambda$	σ_λ	1.37	1.29	0.08
mean of $\ln f$	μ_f	4.13	4.17	0.29
standard deviation of $\ln f$	σ_f	1.99	1.94	0.18