Global Sourcing in Oil Markets[*]

Farid Farrokhi

Purdue University

August 2016

Abstract

Trade in oil accounts for 12-15% 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 $8.9 billion, annual profits of U.S. refineries decrease by $7.1 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.

1 Introduction

Trade in natural resources occupies a small part of international trade literature. Oil alone, as the most traded natural resource, accounts for 12-15% of world trade in recent years. 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, henceforth

---

[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.

EK). The downstream 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 adjusting the quantity of imports from other suppliers. 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 indepen-

2

dently from selection decisions (e.g. Halpern et al (2015), and Antràs et al (2014)).[2] I depart from this extreme 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 all-in-one 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. In particular, 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 the crude production and refining capacity of all countries to their factual values in 2013. The new equilibrium tightly predicts the change to the WTI price relative to Brent.[3] In the data this ratio declines by 9.6%. My model predicts it at 10.5%. 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

---

[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.

[3] This ratio is the relevant index for the price differential between the U.S. and the rest of the world, as WTI (West Texas Intermediate) is the benchmark price in the U.S., and Brent (North Sea, Europe) is the mostly used benchmark outside of the U.S.

counterfactual world where only U.S. crude oil production changes. Specifically, I consider a 36% rise in U.S. production corresponding to its rise from 2010 to 2013. The price of crude oil at refinery drops by 10.5% in the U.S., 9.4% in other countries of Americas, 8.8% in African countries, and on average 7.8% 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 my 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, the price of U.S. crude oil would have risen by 4.7%, the profits of the U.S. refineries would have decreased by 6.6%, and American end-users would have faced 0.1% higher prices of refined oil. Given the enormous values of oil trade, these small changes to prices translate to large dollar values: $8.9 billion increase in the annual revenue of U.S. crude oil producers, and $7.1 billion decrease in the annual profits of U.S. refineries.

Lastly, I study gains from oil trade. I first 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 EK as a benchmark for gains from trade in manufactures. U.S. gains from trade in oil compared with manufactures are at least thirteen times larger. I also consider another counterfactual where oil trade costs are at the autarky level between European countries and non-European countries. Even though this counterfactual is less extreme than a complete country-level auatarky, real wages of European countries drop by 15.8–26.9%. These losses are three to thirteen times larger than the case of country-level autarky in EK.

This study contributes to a vast literature on oil markets. A number of studies have identified 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 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 (see Arkolakis, Costinot, and Rodríguez-Clare (2012), and Costinot and Rodríguez-Clare (2014)). These

4

small gains are at odds with the critical role of trade in natural resources. My results indicate that gains from trade could be tremendously larger in a model designed for trade in oil compared to standard models typically designed for trade in manufactured products.

In addition, the trade literature has focused more on export than import behavior. 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, Fort, and Tintelnot (2014) 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. My alternative specification enables me to derive an all-in-one estimation of trade quantities and selections.

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.[4]

---

[4]A refinery produces a range of products that are largely joint. Refined oil products include gasoline, kerosene and jet fuels, diesel, oil fuels, and residuals. Typically, the heavier fuels are the byproduct of the lighter ones.

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 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 (1960, 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.[5] 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).[6] 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 agencies.[7]

Refineries heavily rely on a constant supply of crude oil as they operate 24/7 over the entire

---

[5] 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 annual surveys conducted by the Oil and Gas Journal titled Worldwide Refinery Survey & Complexity Analysis.

[6] 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)

[7] 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.

year. In particular, the costs of shutting down and restarting are large.[8] [9] As a result, careful scheduling for procurement of crude oil is important. The complications relate to the logistical arrangements in crude oil 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 paper here is Shah (1996) which formulates a refinery's optimal scheduling of multiple crude oil grades of different quality and origin.[10]

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."[11] Textbooks on engineering and economics of refineries assume that refineries take prices of refined products and prices of crude oil as given.[12] 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."[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] Since EIA does not assign id to refineries, I identified refineries using their location information and matched the three datasets. The merged

---

[8] 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.

[9] Refineries keep inventories of crude oil, 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.

[10] 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 on optimal scheduling from the port to refinery (see Pinto, Joro, and Moro, 2000). In the long term, the concerns include multiple orders as well as price and cost variability (see Chaovalitwongse, Furman, and Pardalos (2009), page 115).

[11] Peterson and Mahnovski (2003), page 7.

[12] As a widely used reference, see Gary, Handwerk, and Kaiser (2007), page 19.

[13] Economics of Petroleum Refining (2013), page 3. (Candian Fuels Association)

[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.

dataset contains 110 refineries in 2010 importing from 33 countries.[15] The merged dataset includes 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 data on the upgrading units and weights of each unit reported by the Oil and Gas Journal, I constructed Nelson complexity of refineries.

Using EIA data on before-tax price at the wholesale market of refinery products, I construct the price of the composite of refinery output. I construct a concordance between the 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 compile the prices of crude oil at each origin country for each type.[16] See the online appendix for more details on data.

I document the main facts in the above data, then explain how these facts motivate my model of refineries' sourcing. (I report supporting evidence in appendix).

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 foreign origins. 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 from more than one type of crude oil. Also, 36.1% of refineries import from 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] Not all refineries in one of the three datasets can be found in the other two. To match these datasets I have manually checked the entires of one dataset with the other two, often using online information on refineries to make sure of their correct geographic location. The merged sample accounts for 95% of total capacity and 90% of total imports of the U.S. refining industry in 2010.

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

[17] Specifically, 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%.

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.

Table A.5 reports how refineries' capacity, location, and complexity correlate with their imports. Each observation is the volume of imports of a refiner from a source of crude including zero trade flows. The distance coefficient is highly significant and equals -1.4, where distance is defined between the exact location of the refiner and the 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 relatively more low-quality inputs, but the correlation between complexity and imports of high-quality crude is not statistically significant. The evidence suggests 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.2 reports a set of detailed facts on differences in 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. 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 rate of capacity and complexity, across all refineries equal 1.1% and 0.8%.

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.[18]

Since crude oil is purchased by and large based on *annual* term contracts (Sec. 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 global equilibrium.

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 origin.

---

[18] It is worth mentioning that trade shares of refineries are often concentrated on few suppliers, and these suppliers are not the same across observably similar refineries. In line with my observation for American refineries, Blaum et al (2015, 2013) also report that the imports of French manufacturing firms are highly concentrated on few supplier countries suggesting that suppliers' costs vary across importing firms.

I index refineries by x. Each refiner has a technology that converts crude input to homogeneous 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 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 moments $t \in [0,1]$ that I call days. Let $p_{nj}(x)$ denote the *average cost* of 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 daily cost of supplier j equals

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

where ϵ is the daily variations in transport costs reflecting the availability of tankers, port storage tanks, and pipeline slots. 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.[19]

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 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 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

---

[19] Specifically, $Pr(1/\epsilon \leq 1/\epsilon_0) = \exp(-s_\epsilon \, \epsilon_0^{-\eta})$. Three points come in order: (i) I normalize $s_\epsilon = \left[\Gamma\left(1 + 1/\eta\right)\right]^\eta$ (Γ is the gamma function). This normalization ensures that the mean of ϵ equals one. (ii) Variance of ϵ equals $\Gamma(2/\eta + 1)\big/\left(\Gamma(1/\eta + 1)\right)^2 - 1$, which is decreasing in η. (iii) My independence assumption is observationally equivalent to a more general Fréchet distribution that allows ϵ's to correlate over $t \in [0,1]$. See Eaton and Kortum (2002), footnote 14.

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.[20]

On the sale side, the refiner enters into a contract with wholesale distributors.[21] 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,

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

The variable profit integrates profit flows over the year. 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,

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

[20] 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.

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

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 refinery's access to a wider range of lowest-cost suppliers over the year, so lowers the refinery's 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. The lower η, the more increase to the variable profit from adding a new supplier.[22] This relation delivers η as the *trade elasticity*, defined as the elasticity of demanded quantity from a supplier with respect to the cost of the supplier, 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 days 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,

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

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 price equals

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

---

[22] In an extreme case that $\eta \to \infty$, the cost of each supplier does not vary with ϵ, and the refiner chooses only one supplier.

Equation (5) measures the extent to which adding a new supplier lowers the input cost. (for example, in a special case where $p_j = p$ for all j, $P(S)$ equals $|S|^{-1/\eta}p$).

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,[23]

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

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

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

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 the revenue and the profit margin. Both increase if a larger S is selected.

Holding a refinery fixed, 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.

---

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

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

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

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 import costs across all pairs of refineries and 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 three fourth 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.[24]

---

[24] 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. $1 - \beta_0 = 1 + \beta_0 + \beta_{CI}CI^{\max} \Rightarrow \beta_0 = -\beta_{CI}CI^{\max}/2$.

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 $s_z = \left[\Gamma(1 - 1/\theta) \right]^{-\theta}$, 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 in the data. While θ (relating to z) represents the heterogeneity of variable trade costs in the industry; η represents trade elasticity induced by the dispersion of ϵ's. Moreover, data on annual trade shares can be used to recover z's, while they inform only the dispersion of ϵ's.

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 refinery'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.[25]

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* including 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 the econometrician.

3.5 Mapping Between Observed Trade and Unobservables

Handling interdependent decisions for selecting suppliers in firm-based import models has been a challenge. This interdependency arises because selected suppliers jointly affect the marginal cost of the firm (here, refiner). For example, suppose the price of a supplier significantly rises. In response, the refiner not only drops that supplier but also its entire import decisions change. For

---

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

example, the refiner may add a new supplier or adjust its quantity of imports with 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, here I map the *observed* trade vector q to *unobserved* trade cost shocks z, efficiency λ, and fixed cost f. Then, in Section 4.1, I use this mapping to derive a 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 observed trade vector, q, and the 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 $\bar{f}$ with closed-form solutions. 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 $j \in S$

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

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,

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

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

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

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.[26] 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

---

[26] 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 (2014), 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.

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 $\pi(S^+; p^+)$ 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 a likelihood function that summarizes data on refineries' quantities of imports and their selection decisions. 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. see 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 includes input volumes q_j, wholesale price of refinery output excluding taxes $\tilde{P}$, f.o.b. prices of crude oil inputs p_j^{origin}, refinery capacity R, complexity CI, and I^d as 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 $q(x)$.[27] As there is no strategic competition, the whole likelihood, is given by:

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

---

[27]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 c.d.f. and p.d.f. by F_Q and f_Q. The likelihood contribution of refiner x, L_x, is given by

$$\begin{aligned} L_x(\Omega | \mathbf{D}(x), q(x)) &\equiv L_x(\Omega | \mathbf{D}(x), [q_A(x), 0]) \\ &= f_{Q_A}(q_A(x) \mid Q_A(x) > 0, Q_B(x) = 0; \Omega, \mathbf{D}(x)) \times Pr(Q_A(x) > 0, Q_B = 0 \mid \Omega, \mathbf{D}(x)), \end{aligned}$$

where by construction, $q_A(x)$, is strictly positive.

The calculation of the likelihood without using Proposition 1 involves high-dimensional integrals (see Appendix B.3.1). 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.

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.3.2 contains the proof and more details. 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. In particular, $\ell_B = Pr\Big\{ z_B \geq \underline{z}_B(\lambda, z_A, f) \Big\}$ has a closed-form solution.[28]

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 selec-

---

[28] 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 contain 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.

tions.[29] 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.

4.2 Identification

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.

---

[29] 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.

Figure 2: Identification of diversification gains and fixed costs
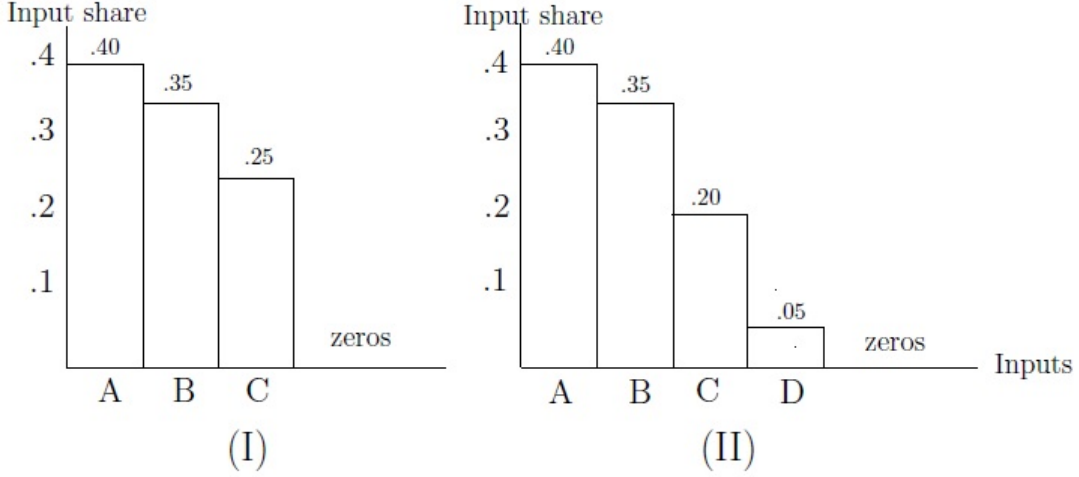
Trade elasticity. Holding a refiner fixed, the cost of supplier j can be written as $p_j = p_j^{obs} z_j$, where p_j^{obs} is the observable part of the cost, and z_j is the unobserved draw (which is normalized to one for the domestic supplier, $j = 0$). Equation (4) implies:

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

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 the 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.3 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 z, the smaller θ.[30]

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 un-

---

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

observed λ closely relates to the distribution of observed utilization rates.

4.3 Estimation Results

Tables 1–2 in column "all-in-one" report the estimation results. Standard errors are shown in parenthesis. The tables also report the results based on estimating (i) the parameters that govern refineries' variable profit using only data on quantities of trade (labeled as flows-only), and (ii) fixed costs only using data on selections, given the estimated variable profit (labeled as sparsity-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 distance coefficient is relatively small. If the f.o.b. price of crude oil is \$100/bbl, every 1000 km adds on average \$2/bbl to variable trade costs. If the state where the refiner is located shares a border with a supplier (either Canada or Mexico), trade costs reduce by 28%. Moreover, the complexity parameter β_{CI} is negative as expected.

The estimate of $\theta = 3.16$ implies that the variance of shocks to trade costs, $Var[z]$, equals 0.38. The variance should be compared to the source-specific estimates of trade costs which range from 0.86 to 1.33 (see Table 2). The source-specific estimates imply that trade costs are on average more than 100% when they are not conditional on formed trading relationships —which is somewhat large compared with benchmarks in the literature.[32] However, refineries select a supplier when they draw a favorable z with respect to that supplier. The estimates imply that the median of variable trade costs conditional on selection equals 17% —significantly smaller than the unconditional size. (Notice that this number is still larger than what refineries pay for trade costs. Because con-

---

[31] For example, Broda and Weinstein (2006) report that the median elasticity of substitution for 10-digit HTS codes is less than four, also 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-Weinstein and Soderbery are different from mine in a number of ways. They directly use unit costs at the location of consumption for homogeneous buyers using the sample of nonzero imports. In contrast, (i) I use firm-level data (heterogeneous rather than homogeneous buyers). (ii) I also estimate trade costs because I have only prices at origin because costs at the gate of refineries are not available. (iii) My sample includes not only imports but also domestic purchases. (iv) My estimation uses the sparsity of trade matrix.

[32] For example, Anderson and van Wincoop (2004) report that trade costs for aggregate trade flows are about 70%.

ditional on selecting set S of suppliers, the refiner buys from $j \in S$ only when j is the lowest-cost supplier among suppliers in S.)

If I separately estimate the flows and sparsity, 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). For example, γ_{Canada} equals 1.08 according to the all-in-one estimation compared with 0.58 in the flows-only estimation.[33]

Table 1: Estimation Results

description	parameter	all-in-one	flows only	sparsity 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.

---

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

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

country	all-in-one	flows only	country	all-in-one	flows 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 & quantitative 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.

I first simulate the effect of 10% increase in variable trade costs, d, on the imports of individual refineries. Total import volumes of a refinery, on average, drop by 26.7%. That is, the elasticity of total imports of the industry with respect to distance is -2.67.[34]

Table 3 reports the model prediction 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 percentile is 14 in the data, 12 according to the all-in-one, and 30 according to the the independent estimations.

---

[34] This elasticity equals -2.72 if I consider 1% (instead of 10%) increase in variable trade costs.

Table 3: 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, equal to the input cost per barrel of crude including transport costs and other fees paid by refineries. Notice that I have prices at origins and quantities at destinations, but I have *no data on values of trade, or equivalently no data on unit costs at destinations*.

My model predicts the annual input cost of a refinery only when it is adjusted for complexity effect (Eq. 5 and Eq. 9). 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. (However, in this exercise, refineries do not take their complexity into account when they decide about their imports. Thus, the exercise provides only an approximate rather than an exact decomposition.) According to my estimates and the above decomposition, the average input cost excluding the effect of complexity, equals 75.5 $/bbl. In the data, the average input cost equals 76.7 $/bbl. According to the results from separate estimations of the flows and the sparsity, the 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 an exogenous trade sparsity delivers trade elasticity $\eta \approx$ 11 instead of $\eta \approx 20$. The underestimation of η is the force behind the overestimation of the extent that refineries diversify (Table 3), and the overestimation of the gains from supplier diversification (Table 4).

Table 4: 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

Holding the prices of crude oil inputs and of the composite refinery output fixed, I now report the benefits of global sourcing to an individual refinery. A refinery on average lowers its (complexity-adjusted) input costs by 8.2% when sourcing globally compared with sourcing only domestically. This number jumps to 26.8% according to the independent estimations. Table 5 reports the results for profits, profit margin, and refinery production. (Eq. 7 expresses profit margin.) For the four variables reported below, the independent estimations predict values that are around three times larger.

Since refineries are capacity constrained, the change in their profits is largely accounted for by the change in the difference between output and input prices, rather than a change in their production. As shown in the 1st row of Table 5, global sourcing compared with only buying domestically increases refineries profits by 56.3%. This increase is associated with 47.1% increase in profit margin while only 4.1% increase in production.

Table 5: Global Souring vs Only Domestic Sourcing (Percentage change)

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

Transition to equilibrium. The above results (for the all-in-one estimation) show some quantitative implications of my estimates when prices of crude oil and prices of refinery output remain unchanged. However, endogenous changes to these prices is key to explore how a local shock, such as a boom in U.S. crude oil production, propagates around the world. To this end, Sections 5–6 embed refineries' sourcing developed in Sections 3–4 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 products. 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 Sec. 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 requires further assumptions about which parameters are universal. The limitation is that refinery-level data are not available for countries except the United States. In particular, a set of parameters could be identified only from refinery-level data, including trade elasticity η, distribution of fixed costs $f \sim logN(\mu_f, \sigma_f)$, distribution of trade cost shock z as a Fréchet with dispersion parameter θ, and complexity coefficient β_{CI}. I continue to use these parameter estimates in my multi-country equilibrium framework. I also continue to 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 cost d, because μ_λ and d could be sensitive to the performance and geography of American refineries. Specifically, I estimate μ_λ and d according to aggregate data on a sample of many countries (see Section 5.3).

5.1 Framework

Section 5.1.1 concerns the aggregation of refineries' sourcing decisions. Sections 5.1.2–5.1.3 link 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 specified in Sec-

tion 3.4. To maintain a seamless transition, I use 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 refinery 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:

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

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:

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

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 suppliers, $p_{i\tau}$, and the wholesale price of refinery output, $\tilde{P}_n$, are given to refineries in country n.

5.1.2 Distributors of Refined Oil Products.

In each country, refinery output is sold domestically in a competitive wholesale market to a continuum of *distributors*. Each distributor converts the homogeneous 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 that is produced in country i equals $[\tilde{P}_i / \xi_i(\omega^e)]$ in country i, 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

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

Assuming that $\sigma^e < \theta^e + 1$, the price indices are

$$e_n = \gamma^e \left(\Phi_n^e\right)^{-1/\theta^e}. \tag{18}$$

where γ^e is a constant[35], and e_n is the price of the composite refined oil products in country n.

---

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

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 as before-tax price of the composite refined oil products in country n. $t_n \in (-1, \infty)$ 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 demand elasticity of refined oil products (or, 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

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

---

[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_i or 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 prices of refinery output $\tilde{P}_n$ 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.

Domain. The sample uses data of year 2010. A country is chosen if its crude oil production is more than 0.75 million bbl/day or otherwise its refining capacity is more than 0.75 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/regions covering the whole world. Table A.10 (in appendix A.1) lists countries/regions and their crude oil production, total refining capacity, average complexity, average utilization rate, as well as taxes on and consumption of refined oil products.

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.

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 0.32 of the trade matrix when defined between 31 producers and 39 destinations. Further, 16 countries do not import crude oil; 9 countries do not export; and the rest both import and export. 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 for GDP and population is Penn World Dataset, and for human capital 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 fuel prices and taxes is International Fuel Prices by GIZ (Federal Ministry for Economic Corporation and Development, Germany).

Accounting of oil flows. Aggregate data on trade flows of crude oil do not necessarily match the aggregate data on countries' exports and total purchases of crude oil. For this reason, I assume that reported trade flows of crude oil are not accurate. 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. Section 2 of the online appendix explains the details of the algorithm.

5.3 Quantifying the Framework

I first explain how I solve my general equilibrium model given all model 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.

Simulation Algorithm. I can not use the method of exact hat algebra, as Dekle, Eaton, and Kortum (2007), to calculate counterfactual equilibrium outcomes. The reason is that in my setup the sparsity of trade endogenously changes in response to shocks. Instead, I parametrize the entire model, and solve the equilibrium by simulation.

For a given distribution of refineries $G_{n,x}$, I simulate the model equilibrium. Prior to running the simulation, I draw artificial refineries $x = (z, \lambda, f, R, \zeta, d)$ from distribution $G_{n,x}$ for each country n for T times.[37] I hold the realizations of refineries x fixed as I search for equilibrium variables. My algorithm to solve for equilibrium consists of an inner and an outer loop. In the inner loop,

---

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

given a vector of crude oil prices $p_{i\tau}$, I solve for the vector of refinery output prices, $\tilde{P}_n$, such that markets of wholesale refinery output, and markets of refined oil products clear. In this inner loop, I solve refinery problem for each realization of x, calculate aggregate 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 $p_{i\tau}$, until aggregate demand for each supplier of crude oil $j = (i, \tau)$ equals the inelastic aggregate supply of crude oil j.

Calibrating/estimating the entire Framework. I explain the entire task of quantifying the framework in four steps. The list of parameters is given by Table 6. The list of countries is given by Table A.10.

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 of 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 revises d using country-level data on crude oil trade flows.

Step 2. A subset of parameters, reported in Table 7, can be taken from either the literature or auxiliary data. 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 bbl/day; $R_n^{\max}$ is given by the Oil and Gas Journal. The best fit to the data on U.S. refinery capacity is achieved at $\phi = .11$.[38] I assume that all refineries 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

---

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

refined oil products, together with equation (20), I calibrate the parameter of oil intensity, $1 - b_n$. 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.[39] 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 predictions as shown by Eaton and Kortum (2002).

A wide range of studies have estimated the elasticity of demand for refined oil products. For the long run, Hausman and Newey (1995) find the price elasticity of gasoline to be 0.80; Yatchew and No (2001) find slightly higher estimates around 0.90; Graham and Glaister (2002) report a range from 0.60 to 0.80; Chang and Serletis (2014) find a range from 0.57 to 0.74. For the short run, the literature suggests a value between 0.05 and 0.34, e.g. see Hughes, Knittel, and Sperling (2008) and Hamilton (2009). Here, the elasticity of demand for refined oil products, ρ, is set to 0.50. This value lies between the short-run and long-run estimates in the literature. The choice reflects the 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),

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

where $V_i^e = \ln m_i^e (\tilde{P}_i)^{-\theta^e}$. Trade cost d_{ni}^e is specified by

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

Here, $exporter_i^e$ is the exporter-specific parameter of trade cost for country i.[40] $distance_{ni}$ 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 8–9.

The estimates of exporter-specific parameters, $exporter_i^e$, represent barriers that are not explained by geographic variables. In some oil-abundant countries, refined oil products are heavily

---

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

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

subsidized. See column 6 in Table A.10. In my estimates, these subsidies are reflected as export barriers. At the other extreme, among *non-producers* of crude oil, the estimates of $exporter^e$ are exceptionally large for the Netherlands and Singapore. Their large $exporter^e$'s reflect that these two countries are the oil trade hubs in Europe and Asia.

Table 6: List of Parameters

---

1. Parameters related to refineries and trade in crude oil

η	trade elasticity of 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 7: Parameter Values set in Step 2

ϕ	ρ	θ^e	σ^e
0.11	0.50	20	5

Table 8: 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 9: 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	Country	Estimate	% Effect
Algeria	-1.7	8.9	*cont'd*		
Angola	-6.9	41.5	Netherlands	6.1	-26.1
Azerbaijan	-5.2	29.4	Nigeria	-3.1	16.9
Brazil	1.7	-8.1	Norway	-3.5	19.2
Canada	1.3	-6.2	Oman	-5.0	28.2
China	1.7	-8.1	Qatar	-3.7	20.1
Colombia	-3.0	15.9	Russia	1.1	-5.5
France	2.4	-11.2	Saudi Arabia	-2.2	11.7
Germany	1.9	-9.3	Singapore	5.1	-22.4
India	2.5	-11.5	Spain	1.7	-8.3
Indonesia	-0.4	2.3	UAE	-2.3	12.1
Iran	-5.0	28.2	United Kingdom	2.6	-12.1
Iraq	-8.2	51.0	United States	6.2	-26.7
Italy	2.3	-10.9	Venezuela	-1.9	10.1
Japan	2.3	-11.0	RO America	2.4	-11.4
Kazakhstan	-2.7	14.6	RO Europe	3.9	-17.9
Korea	4.9	-21.6	RO Eurasia	-1.5	8.0
Kuwait	-2.1	10.8	RO Middle East	3.1	-14.3
Libya	-2.9	15.5	RO Africa	4.0	-18.0
Mexico	-0.8	4.1	RO Asia & Oceania	5.0	-22.1

Step 4. All parameters listed in Table 6 are set in steps 1–3 except mean of log efficiency of refining cost $\mu_{\lambda,n}$, variable trade costs of crude oil d_{ni}, and location parameter of the distribution of 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 for the calibration. As I search for μ_λ, d, and m^e, I use the fixed 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, then aggregate refinery-level to

country-level variables, then match the model to three sets of moments, as I continue to explain.

On the one hand, the parameters of $\mu_{\lambda,n}$ and d_{ni} sum up to N^2 unknowns: N unknown $\mu_{\lambda,n}$, and $N^2 - N$ unknown d_{ni} (by normalization $d_{ii} = 1$). On the other hand, aggregate trade flows of crude oil sum up to N^2 known entires (including domestic purchases). Given all other parameters, the N^2 unknown $\mu_{\lambda,n}$ and d_{ni} are just-identified with respect to the N^2 moments of crude oil trade flows. Specifically, I summarize the trade flows in two sets of moments. The first set of moments A^1, includes *total use* of crude oil for every country n,

$$A_n^1 = \sum_{i=1}^{N} \sum_{\tau=1}^{2} Q_{ni\tau}.$$

The second set of moments A^2, contains all trade shares, denoted by A_{ni}^2, as the ratio of imports from i to n relative to total input use in n,

$$A_{ni}^2 = \frac{\sum\limits_{\tau=1}^{2} Q_{ni\tau}}{A_n^1}.$$

The model predictions match A_n^1's and A_{ni}^2's if and only if they match all trade flows at the country level, $Q_{ni} = \sum_{\tau} Q_{ni\tau}$. (In aggregate data, I only observe country-level flows $\sum_{\tau=1}^{2} Q_{ni\tau}$, but not country- and type-level flows $Q_{ni\tau}$). Given all other parameters, $\mu_{\lambda,n}$ governs the scale or total input demand $[A_n^1]_{n=1}^{N}$, and $[d_{ni}]_{n\neq i}$ governs the shares $[A_{ni}^2]_{n\neq i}$.

I confront this computationally intensive problem by exploiting some useful properties of the model. In the calibration problem, I allow f.o.b. prices of crude oil to be given by data. Therefore, for two importers n and n', and for two exporters i and i', trade from i to n does not depend on trade costs between n' and i', that is $\partial Q_{ni}/\partial d_{n'i'} = 0$.[41] Consequently, the large Jacobian matrix of the excess demand function is very sparse —from every N entries, $N - 1$ are zero.[42] This property significantly reduces the computational burden.

In addition, I calibrate the efficiency parameters in retail sale of refined oil products, m_n^e. I define A_n^3 as the ratio of refinery output price $\tilde{P}_n$ to the average price of a barrel of crude in country

---

[41] Note that this property does not hold if I solve for f.o.b. prices within the equilibrium.

[42] The excess demand function is a $N^2 \times 1$ vector, where each row refers to the excess demand for a destination-source pair.

n. The larger m_n^e, the more local demand for refinery output in *n*, the higher $\tilde{P}_n$, the higher A_n^3. I assume that for all countries, A_n^3 equals A_{USA}^3 which is known by the estimates in Section 4. Given all other parameters and f.o.b. prices of crude oil, I update my guess of $\mu_{\lambda,n}$, d_{ni}, and m_n^e through successive iterations until my model predictions fit A_n^1, A_{ni}^2, and A_n^3.

5.4 Model Fit

In the calibration, I have matched country-level crude oil trade flows, $Q_{ni} = \sum_{\tau=1}^{2} Q_{ni\tau}$, rather than country- and type-level crude oil trade flows, $Q_{ni\tau}$. (Because international trade data are available only at the country level). In my definition of equilibrium, however, market clearing conditions hold for each supplier as a pair of source country and type, $Q_{i\tau}$. So, when I use calibrated μ_λ, d, and m^e together with other parameters to solve for equilibrium, the results do not necessarily equal to the actual f.o.b. prices of crude oil that I initially fed into the calibration. However, the correlation between the predicted f.o.b. prices and their actual values remains high; equal to 0.83.[43] In addition, the equilibrium outcome almost exactly fits to 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, consider

$$\log d_{ni} = imp_n + exp_i + \alpha_d \log(distance_{ni}) + \alpha_b border_{ni} + error_{ni},$$

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. The OLS results are reported in Table 10. As expected, distance highly correlates with the calibrated trade costs.

---

[43]In addition, the ratio of average price of high- to low-quality crude oil is predicted at 1.098 compared with 1.094 in data.

Table 10: Geographical Variables & Calibrated Trade Costs

$\log(distance_{ni})$	$border_{ni}$	importer FE	exporter FE
0.22	−0.03	yes	yes
(0.025)	(0.077)	–	–

Notes: Standard errors in parenthesis. Number of observations = 359. $R^2 = 0.69$.

6 Quantitative Predictions

The framework –developed in Sections 3, 5.1, 5.2, and quantified in Sections 4, 5.3, 5.4– 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 U.S. production propagates around the world. Section 6.3 examines the implications of (i) lifting the export ban on U.S. crude oil, and (ii) gains from oil trade for the U.S. 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 production and demand from 2010 to 2013. Recall that in my framework, the flows of crude oil production $Q_{i\tau}$, and the measure of total refining capacity M_i, are exogenously given. In Sections 4 and 5.3–5.4, I quantified the framework using cross sectional data from 2010. Here, I re-calculate the equilibrium when crude 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. crude production grew by 36%. While the total production in the rest of the world remained stagnant, its composition changed to some extent. Production in Europe, Libya, and Iran declined; and in Canada, and part of the Middle East rose. On the demand side, refining capacity increased by 1.5 million b/d in India. About the same size of capacity was also added to a collection of other countries in Asia. Table A.11 reports the changes from 2010 to 2013

in crude oil production and refinery capacity of all countries.

What did these changes imply for the relative prices of crude oil across countries? How did they change imports and production of U.S. refineries? To what extent can my model predict these changes in the data? Table 11 reports the data, as well as my model predictions.

Regarding the prices, two observations are noteworthy. Between 2010 and 2013, the crude oil price of West Texas Intermediate (WTI) in the U.S. relative to the price of Brent in Britain decreased by 9.6%. This price ratio is the relevant index for the price differential between the U.S. and the rest of the world.[44] Moreover, prices of refined oil products in the U.S. did not track the price of WTI. Specifically, the wholesale price of the composite of refinery output (including gasoline, diesel, jet fuels, etc.) in the U.S. increased by 8.4% relative to the price of WTI.

My model predicts the drop in the WTI/Brent ratio at 10.5% compared with 9.6% in the data.[45] Further, the model predicts that the U.S. wholesale price of refined products relative to WTI has increased by 6.4% compared with 8.4% in the data. Not only the predictions are on the right direction, but also their magnitudes are close to the factual changes.

Moreover, the model closely predicts the changes in import volumes, number of trading relationships, and the total use of crude oil for the U.S. refining industry. Recall that I have quantified my model only by cross-sectional data from 2010. These predictions, therefore, show a validation on the performance of the model.

Table 11: 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 price relative to WTI	WTI/Brent crude price ratio
Data	−16.1%	−15.2%	2.2%	8.4%	−9.6%
Model	−15.5%	−13.0%	2.4%	6.4%	−10.5%

The above experiment considers *all* shocks to the location of supply and demand. The framework also allows me to focus on *one* shock, and quantify how it propagates around the world. I

---

[44] WTI is the benchmark price in the U.S., and Brent is the mostly used benchmark outside of the U.S.

[45] The counterpart of WTI/Brent in the model is the relative f.o.b. prices of high quality crude oil of the U.S. to high quality crude oil of the U.K.

now do so for the recent boom in U.S. production.

6.2   A Boom in U.S. Crude Oil Production

The Unconventional Oil Revolution –the boom in U.S. production of crude oil– is at the heart of the conversation on U.S. energy policy. The production of crude oil in the United States grew 36% from 2010 (for which my model is calibrated) to 2013 (as the most recent year with available data on production, capacity, and prices). I consider a counterfactual world where only U.S. production changes by 36% (equal to two million barrels per day). How does this shock change the relative prices of crude oil across the world?

The price of crude oil at the location of refineries, also called *acquisition cost* of crude oil, is defined as the average cost of a barrel of crude for the refining industry of a country,

$$\bar{P}_n = \left(\int_{x \in X_n} u(x) RP(x) \, dG_{x,n}(x) \right) \Big/ \left(\int_{x \in X_n} u(x) R \, dG_{x,n}(x) \right),$$

where $P(x)$ is the input price index of refiner x (given by equation 5). Moreover, average utilization rate for the refining industry of each country equals

$$\bar{u}_n = \left(\int_{x \in X_n} u(x) R \, dG_{x,n}(x) \right) \Big/ \left(\int_{x \in X_n} R \, dG_{x,n}(x) \right),$$

Table 12 reports the model predictions for changes in the acquisition cost $\bar{P}$, average utilization rate $\bar{u}$, and the price index of refined oil products e (given by equation 18) for all countries/regions. Three results stand out:

Most importantly, there is a systematic regional effect on the prices of crude oil, but is not large. The results are also illustrated in Figure 3. The average prices of crude *at source* falls by 11.5% in the U.S. and on average 8.4% in the rest of the world. The acquisition cost, or the price of crude at the gate of refineries, drops by 10.5% in the U.S., 9.8–9.9% in Canada and Mexico, 9.4–9.6% in Venezuela and Colombia, 8.9–9.3% in Brazil, West Africa, and Algeria; while less than 8.5% in the rest of the world, and only 6.7–7.1% in Singapore and Japan. The results imply that

markets of crude oil are not entirely integrated, but fragmented to a modest degree. In particular, compared with Americas and Africa, countries in Europe, Russia, and part of Asia are less integrated with the U.S. market.

Figure 3: Worldwide propagation of a shock to U.S. crude oil production
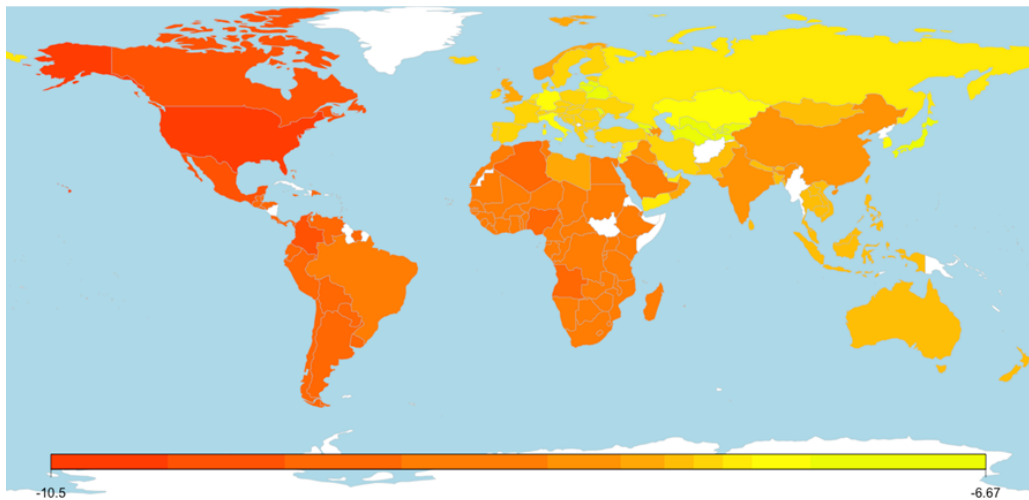
Note: The figure shows the percentage change in the predicted price of crude oil at the location of refineries when U.S. production increases by two million bbl/day, equal to its change from 2010 to 2013. The maximum drop in price belongs to the U.S., and the minimum belongs to Japan and Singapore. Section 6.2 provides the details.

Second, regional effects on refined oil prices disappear. 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 often supply refined oil more domestically than internationally, and since they 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 minimum among all –equal to 0.30 and 0.36, respectively– exemplify this mechanism.

Third, prices of refined oil products fall less than prices of crude oil. Appendix B.4, in a simple one-country model with homogeneous refineries, shows that by an increase in crude production, the price of crude drops more than the price of refined. The intuition is as follows: When worldwide supply of crude increases, refineries have to refine more crude oil in equilibrium. To have than happen, the price gap between crude and refined oil should rise so that refineries can afford the higher utilization costs imposed by capacity constraints. (see Appendix B.4 for details).

Table 12: Percentage change in acquisition price of crude $\bar{P}$, utilization rate $\bar{u}$, and refined oil price e, in response to 36% rise in U.S. production

Country	$\bar{P}$	$\bar{u}$	e	Country	$\bar{P}$	$\bar{u}$	e
Algeria	-9.1	1.2	-4.5	*cont'd*			
Angola	-9.3	3.7	-4.8	Netherlands	-7.7	4.5	-4.7
Azerbaijan	-8.2	8.4	-4.8	Nigeria	-9.3	5.5	-6.8
Brazil	-8.9	1.3	-4.2	Norway	-8.0	1.4	-4.5
Canada	-9.8	2.0	-4.3	Oman	-8.1	1.4	-4.8
China	-8.3	1.2	-4.4	Qatar	-7.4	1.0	-4.5
Colombia	-9.6	4.0	-4.8	Russia	-7.7	1.1	-4.5
France	-7.8	4.2	-4.7	Saudi Arabia	-8.5	1.2	-4.6
Germany	-7.4	3.9	-4.7	Singapore	-6.7	9.6	-4.9
India	-8.4	0.5	-4.2	Spain	-7.8	6.3	-4.9
Indonesia	-7.9	2.3	-4.8	UAE	-7.7	2.8	-4.9
Iran	-7.8	0.7	-3.8	United Kingdom	-8.0	3.1	-4.6
Iraq	-8.4	7.9	-5.1	United States	-10.5	1.8	-4.3
Italy	-7.5	5.8	-4.8	Venezuela	-9.4	2.5	-4.3
Japan	-7.1	6.7	-5.4	RO_America	-9.1	3.7	-4.6
Kazakhstan	-7.6	4.5	-4.9	RO_Europe	-7.8	3.3	-4.6
Korea	-7.5	4.3	-4.8	RO_Eurasia	-6.9	6.9	-5.1
Kuwait	-8.1	0.6	-4.6	RO_Middle East	-7.7	1.2	-4.5
Libya	-8.0	1.3	-4.5	RO_Africa	-8.5	2.7	-4.6
Mexico	-9.9	1.6	-4.3	RO_Asia & Oceania	-7.9	2.4	-4.7

6.3 Trade Barriers and Gains from Trade

6.3.1 Lifting the Ban on U.S. Crude Oil Exports

There has been much interest in implications of lifting the export ban on the crude oil produced in the United States. Had this ban overturned, how much would have U.S. imports risen? 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 relation between the calibrated trade costs and geographic variables to predict these costs. See Appendix A.5. Specifically, let $d^{new}_{n,USA}$ denote the

counterfactual trade costs when the ban is lifted. I perform a counterfactual experiment in which U.S. production rises by 36% and trade costs of exports from the U.S. are $[d_{n,USA}^{new}]_{n=1}^{N}$. Table 13 reports the percentage changes between two cases: (i) when the ban is lifted and U.S. production rises by 36%, and (ii) when the ban is maintained and U.S. production rises by 36%.

Table 13: Percent changes of trade, production, and prices in the U.S. refining industry due to removing export restrictions.

import volumes	# of trading relationships	utilization rate	refineries' profits	US refined oil price	US crude f.o.b prices
15.52%	8.38%	-0.50%	-6.62%	0.07%	4.73%

In both scenarios, U.S. crude production increases by 1.98 million barrels per day. If the ban is in place, 0.05 million barrels per day are exported (only to Canada). When the ban is lifted, U.S. exports 1.48 million barrels per day.

Had the ban been lifted when U.S. production rose from 2010 to 2013, the prices of crude oil at origin would have been higher by 4.73% in the U.S. and 0.60% in the rest of the world. The policy has notable effect on the producers in the U.S., and much smaller effect on the producers in the rest of countries. Regarding the pass-through in the U.S., the refining industry would have lost -6.62% of its profits, while American end-users would have faced the same refined oil prices (more accurately, 0.07% higher). Had the ban been lifted, the revenue of U.S. crude producers would have increased by $8.93 billion, and the profits of U.S. refineries would have decreased by $7.06 billion.[46]

Combining this experiment with the ones in Sections 6.1–6.2, I report the change to WTI/Brent price when (I) crude production and refining capacity of all countries are set to their values in 2013, (II) only U.S. crude production is set to its value in 2013, (III) U.S. crude production is set to its value in 2013 and the U.S. export ban is lifted. See Table 14. Comparing (II) and (III), lifting the ban accounts for half the gap between Brent and WTI prices.

---

[46] The finding is in line with the views by some of the experts on oil markets. For example, see Kilian (2015): "[...] 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." (page 20).

Table 14: Percent changes in WTI/Brent price ratio

	Data	I. production and capacity of all countries	II. production of U.S. with ban	III. production of U.S. with no ban
WTI/Brent	-9.6%	-10.5%	-5.1%	-2.4%

6.3.2 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.

A few comments are noteworthy. First, if crude oil trade between an exporter like Russia and the rest of the world is prohibitive, my model predicts that the price of crude oil in Russia must be zero. Because Russian crude oil production exceeds its total refining capacity, Russian market at autarky does not clear at any positive price of crude oil. Hence, the price of crude oil in Russia is trivially zero at autarky. Moreover, 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 avoid these two extremes —the case of an exporter like Russia, and a non-producer like Germany. Instead, I focus on less extreme counterfactuals for which my model delivers more informative results.

Gains from oil trade depends on the change to the oil price from the baseline to autarky, which in turn depends on model parameters including the demand elasticity of refined oil products, ρ. I have estimated and calibrated all these parameters except ρ which I have taken from the literature. As discussed in Section 5.2, the literature offers a wide range of values for ρ. Datasets across these studies differ in terms of countries and years they cover, and what exactly they measure (e.g. wholesale versus retail, household consumption versus consumption per automobile). As the benchmark, I have set $\rho = 0.5$ which lies in the middle of this range. However, some studies suggest that at least in the United States this elasticity has recently decreased (e.g. Hughes, Knittel, and Sperling (2008)). For this reason, I also show the results for a smaller value of ρ. To do so, I

have set $\rho = 0.2$ and calibrated the whole model in the same exact way as I did in Section 5.3.

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, the average price of crude oil at source increases by 719.9%, input costs of refineries increase 791.3%, profits of refineries drop by 96.2%, the price index of refined oil increases by 606.3%, and the price index of final goods rises by 13.4%. Because of the increase in the price of U.S. crude oil, gdp rises by 8.6%. Consequently, real gdp (gdp divided by the price index of final goods) decreases by 4.2%.

I compare my results on gains from trade in oil, to the results in the literature on gains from trade in manufactures. As shown in the first row of Table 15, U.S. real wage (wage divided by the price index of final goods) drops by 11.8%.[47] Eaton and Kortum (2002) provides a benchmark for gains from trade in manufactures. When they shut down trade in manufactures, real wage in the U.S. drops by 0.8-0.9.[48] Accordingly, gains from trade in oil compared with manufactures is at least thirteen times larger in terms of real wage (also, at least four times larger in terms of real gdp). [49]

Table 15: Percentage changes from the baseline to the case that crude and refined oil trade between the United States and the rest of the world is prohibitive.

refined oil demand elasticity	price of crude oil at refinery	profits of refineries	price index of refined oil	price index of final goods	real wage
$\rho = 0.5$ (benchmark)	791.3	-96.2	606.3	13.4	-11.8
$\rho = 0.2$	1516.9	-93.1	1179.1	26.7	-21.1

In addition, I report the results when the demand elasticity of refined oil products, ρ, is set to

---

[47]Since wage is exogenous, the whole change to the real wage comes from the price index.

[48] See Eaton and Kortum (2002), Table IX, page 1769.

[49]According to the benchmark provided by Arkolakis, Costinot, and Rodríguez-Clare (2012), U.S. gains from oil trade ranges between 0.7–1.4% which is eight to sixteen times smaller than my finding on gains from oil trade in terms of real wage.

$\rho = 0.2$ (see the second row of Table 15). In this case, from the baseline to autarky, U.S. real wage decreases by 21.1% and U.S. real gdp decreases by 7.2%. Therefore, U.S. gains from oil trade is about as twice when $\rho = 0.2$ compared to the benchmark $\rho = 0.5$.

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.

In the benchmark where $\rho = 0.5$, across European countries price of crude oil at refinery increases by 798–998%, price index of refined oil products increases by 723–870%, and price index of final goods rises by 18.8–36.8%. 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. Profits of refineries do not necessarily drop. Particularly, Germany, Netherlands, and United Kingdom strengthen their comparative advantage in refined oil. Although in the new equilibrium refineries produce less, they may earn a higher profit per barrel of production. When the latter effect is stronger, refineries' profits increase. In addition, real wages across the countries decrease between 15.8% (United Kingdom) and 26.9% (Netherlands). See Table 16-Panel A.

In the case with $\rho = 0.2$, the increase in the prices of crude oil, refined oil, and final goods are as more than twice as the benchmark with $\rho = 0.5$. In addition, real wages across the countries decrease between 31.5% (United Kingdom) and 43.1% (Netherlands). See Table 16-Panel B.

Even though this counterfactual is less extreme than a complete autarky at the level of individual countries, the results remain striking. Table 17 reports the percentage change in the real wage of European countries according to

- my framework, if Europe as a whole would be at the oil autarky for $\rho \in \{0.2, 0.5\}$.

- Eaton and Kortum (2002, EK) and the one-sector version of Costinot and Rodríguez-Clare (2014, CR), if each of the European countries would be at the country-level autarky.

In any comparison, gains from oil trade are enormously larger. At the least, for the Netherlands at $\rho = 0.5$ compared to EK, gains from oil trade are three times larger. At the most, for Spain at $\rho = 0.2$ compared to EK, gains from oil trade are twenty five times larger.

Table 16: Percentage changes from the baseline to the case that crude and refined oil trade between Europe and the rest of the world is prohibitive.

	price of crude at refinery	profits of refineries	price index of refined oil	price index of final goods	real wage
Panel A. demand elasticty of refined oil products, $\rho = 0.5$ (benchmark)					
France	856.0	0.1	777.4	21.4	-17.6
Germany	837.3	61.4	773.1	20.0	-16.7
Italy	998.4	-58.8	870.0	21.3	-17.6
Netherlands	832.7	25.7	764.0	36.8	-26.9
Norway	902.6	-38.4	736.9	21.2	-17.5
Spain	984.5	-68.4	843.1	22.6	-18.4
United Kingdom	798.9	232.1	723.0	18.8	-15.8
RO Europe	933.1	-27.4	812.4	25.4	-20.3
Panel B. demand elasticty of refined oil products, $\rho = 0.2$					
France	2012.9	107.5	1830.0	52.3	-34.3
Germany	1965.4	233.7	1817.3	48.6	-32.7
Italy	2338.5	-15.1	2037.2	54.1	-35.1
Netherlands	1960.9	159.4	1798.0	75.7	-43.1
Norway	2121.9	671.1	1875.3	53.6	-34.9
Spain	2286.2	-33.2	1977.7	57.5	-36.5
United Kingdom	1884.1	607.9	1710.7	45.9	-31.5
RO Europe	2185.1	51.2	1910.9	62.5	-38.5

Table 17: Gains from trade, percentage change in the real wage from the baseline to the relevant version of autarky

	oil trade $\rho = 0.5$	oil trade $\rho = 0.2$	EK	CR
France	17.6	34.3	2.5	3.0
Germany	16.7	34.3	1.7	4.5
Italy	17.6	32.7	1.7	2.9
Netherlands	26.9	35.1	8.7	6.2
Norway	17.5	43.1	4.3	–
Spain	18.4	34.9	1.4	3.1
United Kingdom	15.8	36.5	2.6	3.2

Notes: Columns 1–2: results from my framework, if the whole Europe would be at the oil autarky for $\rho \in \{0.2, 0.5\}$. Column 3–4: results from Eaton and Kortum (2002, EK) and the one-sector version of Costinot and Rodríguez-Clare (2014, CR), if each of the European countries would be at the country-level autarky.

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 can be effective to a certain degree. 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 tremendously larger than gains from trade in standard models that are originally designed for manufactures trade.

My model of refineries' sourcing can be useful for other single agent discrete-and-continuous problems, in particular applications where input users choose among available suppliers and purchase continuous amounts from each. For example, a company not only hires different number of workers from certain academic disciplines but also selects from which academic disciplines to hire. Or, a firm not only produces certain quantities of a set of differentiated products but also selects which set of differentiated products to produce. The tools developed in Sections 3–4 allow for estimating such models by taking into account that selection decisions are endogenous.

An important direction for future research is modeling dynamic decisions of crude oil producers to explore and to extract, and of refiners to invest on refinery capacity and complexity. While my framework is designed for the medium run, these dynamic considerations are the key to study long-run outcomes. Developing a model of long-run equilibrium of oil markets, in turn, is necessary to address oil-related environmental questions.

References

Anderson J. and van Wincoop E. (2004), *"Trade Costs"*. Journal of Economic Literature, Vol. 42(2), pp. 691-751.

Antràs P., Fort T., and Tintelnot F. (2014), *"The Margins of Global Sourcing: Theory and Evidence from U.S. Firms"*. Working paper.

Arkolakis C., Costinot A., and Rodríguez-Clare A. (2012), *"New Trade Models, Same Old Gains?"*. The American Economic Review, 102(1): 94–130.

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

Blaum J., Lelarge C., and Peters M. (2015), *"The Gains From Input Trade in Firm-Based Models of Importing"*. Working Paper.

Broda C. and Weistein D. (2006), *"Globalization and the Gains From Variety"*. The Quarterly Journal of Economics, 121(2), pp. 541-585.

Brown S., Mason C., Krupnick A., and Mares J. (2014), *"Crude Behavior: How Lifting the Export Ban Reduces Gasoline Prices in the United States"*. Resources for the Future, Issue Brief 14-03-REV.

Caliendo L. and Parro F. (2015), *"Estimates of the Trade and Welfare Effects of NAFTA"*. The Review of Economic Studies, 82(1), pp. 1-44 .

Candian Fuels Association (2013), "Economics of Petroleum Refining: Understanding the business of processing crude oil into fuels and other value added products"

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

Chang D. and Serletis A. (2014), *"The Demand for Gasoline: Evidence from Household Survey Data"*. Journal of Applied Econometrics, 29(2), pp. 291-313.

Costinot and Rodríguez-Clare (2014) *"Trade Theory with Numbers: Quantifying the Consequences of Globalization"*. Handbook of International Economics, vol. 4, chapter 4.

Dekle, Eaton J., and Kortum S. (2007), *"Unbalanced Trade"*. The American Economic Review. 97(2), pp. 351-355.

Eaton J. and Kortum S. (2002), *"Technology, Geography, and Trade"*. Econometrica, 70(5), pp. 1741-1779.

Eaton J., Kortum S., and Kramarz F. (2011), *"An anatomy of international trade: Evidence from French Firms"*. Econometrica, 79(5), pp. 1453-1498.

Energy Charter (2007), *"Putting a Price on Energy: International Pricing Mechanism for Oil and Gas"*.

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.

Graham D. and Glaister S. (2002), *"The Demand for Automobile Fuel: A Survey of Elasticities"*. Journal of Transport Economics and Policy, 36(1), pp. 1-25.

Halpern L., Koren M., and Szeidl A. (2015), *"Imported Inputs and Productivity"*. The American Economic Review (Forthcoming).

Hausman J. and Newey W. (1995), *"Nonparametric Estimation of Exact Consumer Surplus and Deadweight Loss"*. Econometrica, 63, pp. 1445-1476.

Hughes, J., Knittel C., and Sperling D. (2008), *"Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand"*. Energy Journal, 29, no. 1, pp. 113-34.

Ireland T. and Kullback S. (1968), *"Contingency Tables with Given Marginals"*. Biometrica, 55(1), pp. 179-188.

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

Kilian L. (2015), *"The Impact of the Shale Oil Revolution on U.S. Oil and Gasoline Prices"*. CEPR Discussion Paper 10304.

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

Nelson, W.L. (1960), *"How to describe refining complexity"*. Oil & Gas Journal. Mar. 14, p. 189.

Nelson, W.L. (1960), *"How complexity of a refinery affects costs of processing"*. Oil & Gas Journal. Sept. 26, p. 216.

Nelson, W.L. (1961), *"How to compute refinery complexity"*. Oil & Gas Journal. June 19. 26, p. 109.

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

Pinto J. M., Joly M., Moro L.F.L (2000), *"Planning and scheduling models for refinery operations"*. Computers and Chemical Engineering 24. pp 2259–2276.

Platts. (2010), *"The Structure of Global Oil Markets"*. Backgrounder Reports.

Santos Silva J.M.C., and Tenreyro S. (2006), *"The Log of Gravity"*. The Review of Economics and Statistics, 88(4), pp. 641-658.

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

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

Sweeney R. (2014), *"Environmental Regulation, Imperfect Competition and Market Spillovers: The Impact of the 1990 Clean Air Act Amendments on the U.S. Oil Refining Industry"*. Working Paper.

Waugh M. (2010), *"International trade and income differences"*. The American Economic Review, 100(5), pp. 2093-2124.

Yatchew A. and No J. (2001), *"Household gasoline demand in Canada"*. Econometrica, 69, pp. 1697-1709.

Appendix A Complementary Data, Figures, and Tables

A.1 Tables & Figures

Table A.1: Capacity and number of refineries importing from none, one, and more than one foriegn origin

	Total	0	1	2+
		# of foreign origins		
# 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

one type	two types	three types	four types
Share of Importing Refiners from			
21.6%	12.4%	29.9%	36.1%

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

Figure A.1: U.S. Refineries and Capacity, year 2010. Diameter of circles is proportional to capacity size. For visibility of smaller refineries, the smaller capacity size, the darker it is.
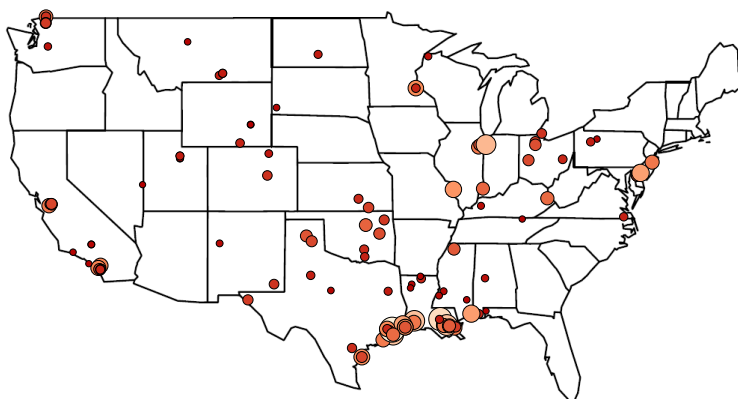
Figure A.2: Distribution of Refinery Capacity in the U.S. Refining Industry, year 2010.
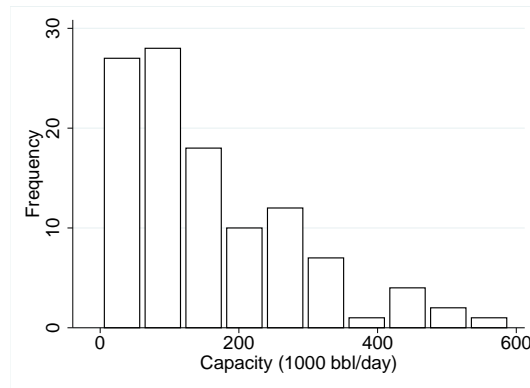

Figure A.3: Distribution of Refinery Distance to Coastline for the U.S Refineries, year 2010. Distance to Coastline is defined as the smallest distance between location of refinery to all ports in the U.S.
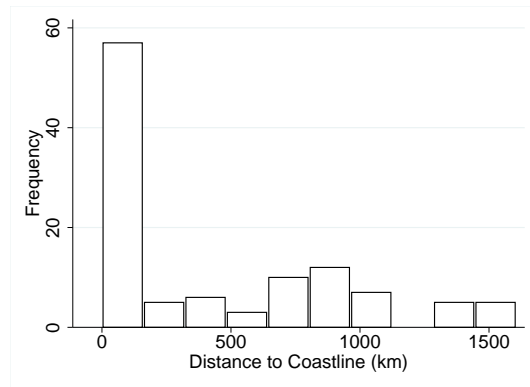

Figure A.4: Distribution of Complexity Index in the U.S. Refining Industry, year 2010. See Section 2.1 for definition of complexity index.
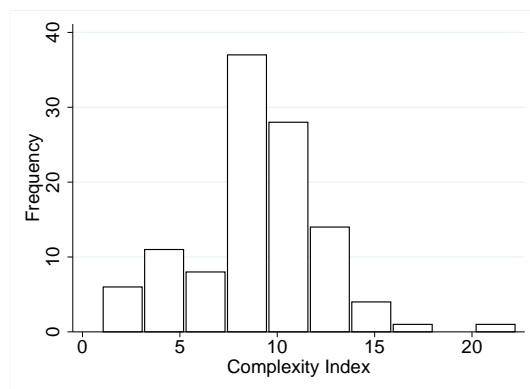

Table A.4: Poisson regression of "number of import origins" on variables related to capacity, geography, and complexity. Larger refiners systematically import from a higher number of

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. 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 below.

Table A.4: Estimation Results, using Poisson Maximum Likelihood

Dependent variable: number of import origins		
	(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.

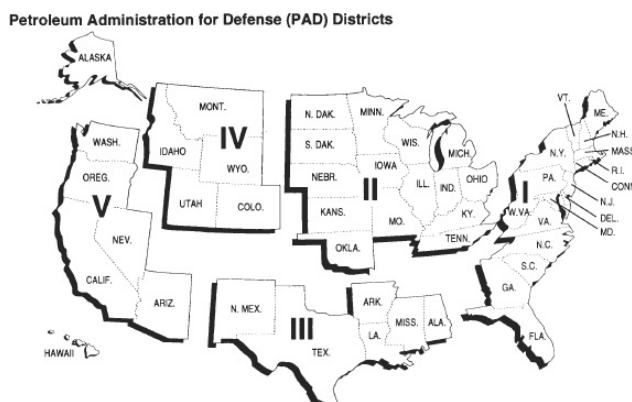Figure A.5: Petroleum Administration for Defense Districts (PADD)

Table A.5: Dependent variable: barrels of imports of individual refineries from suppliers including zero flows.

	(1)	(2)
Dependent variable: refinery-level imports (possibly zero)		
log(distance)	−1.389	−2.168
	(0.245)	(0.342)
border	0.788	0.717
	(0.404)	(0.422)
log(f.o.b. price)	−4.681	−4.413
	(2.449)	(1.866)
Type L	−4.514	−4.412
	(1.448)	(1.866)
Type L×log(CI)	1.449	1.827
	(0.401)	(0.826)
Type H×log(CI)	−0.408	–
	(0.501)	
log(capacity)	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. The results are based on a Poisson pseudo maximum likelihood (ppml) estimation. Each observation is a trade flow (possibly zero) from a source country to an American refinery in year 2010. In column 2, observations for non-importers are dropped; also, by inclduign refinery fixed effects capacity and either $TypeH \times log(CI)$ or $TypeL \times log(CI)$ should be dropped. A Tobit regression delivers the same signs and levels of significance for all coefficients. For details on the advantage of using ppml for estimating trade volumes, see Santos Silva and Silvana Tenreyro (2006).

Figure A.6: Distribution of annual percentage change of complexity (left) and capacity (right), for the U.S. refineries from 2008 to 2013.
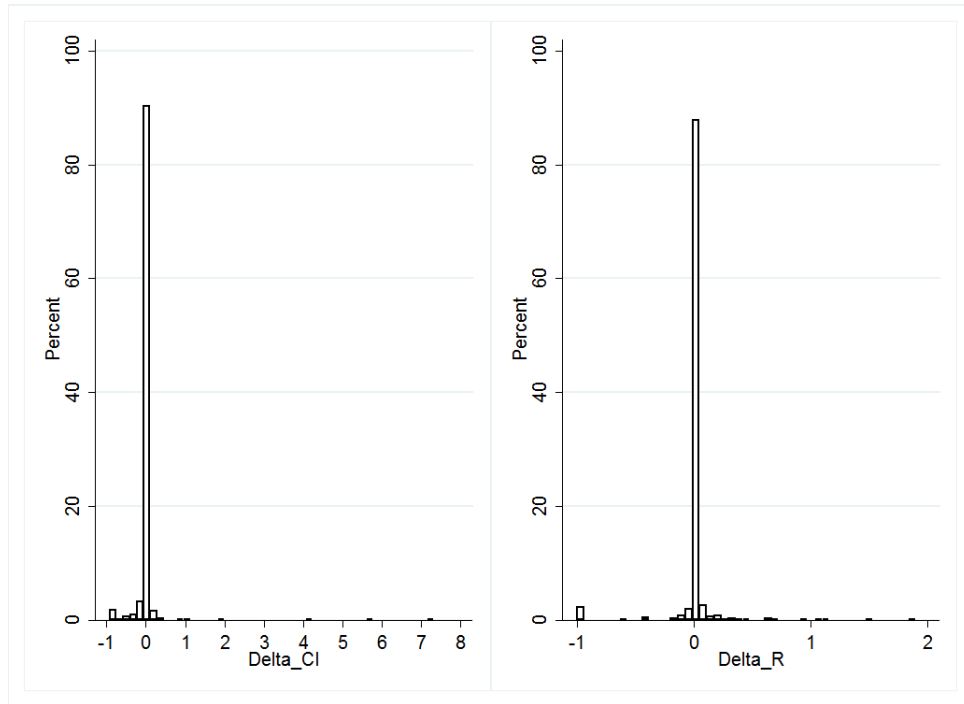
A.2 Complementary notes on Fact 3.b

I consider three samples of refineries: (i) all refineries, (ii) refineries located in the Gulf coast area[50], (iii) refineries that are located within 40 km 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 the groups 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 $commons_S(g)$ for group g

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

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. To help my reader to read the results, consider tables A.6. There are 18 refineries with large capacity and high complexity, see the number in (C3, R3). On average, these 18 refineries import from 8.3 origins. The average number of common origins across all 153 pairs of these refineries equals 3.6.

Table A.6: Common Selection, 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 Selection, 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 Selection, 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 i is an import origin, and $I = 33$. $T_i(x)$ is the import share of refiner x from i. For each pair of

refiners x_1 and x_2, I define *distance in imports*,

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

which is the summation of distances between import shares for those origins whom from both x_1 and x_2 import. I define $distance_T(g)$ for group g,

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

Specifically, $distance_T(x_1, x_2)$ equals zero, if refineries allocate the same share of their total use to their common suppliers. The maximum value of $distance_T(x_1, x_2)$ is two. Consider group $(C3, R3)$ in sample (i) that includes all refineries. The average distance in this group equals 0.65 which is far above zero. It is remarkable that the number does not change as we control for some features of geography in samples (ii) and (iii).

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.3 Trade elasticity: identification and sample selection bias

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

$$\underbrace{\ln \frac{q_j}{q_0}}_{y_j|j \in S} = -\eta \ln \frac{p_j^{obs}}{p_0^{obs}} - \eta \ln z_j, \quad \text{if } j \in S. \tag{25}$$

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 argue that this orthogonality condition does not hold because z's in equation (25) belong only to selected suppliers.

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. when z_j is smaller than a threshold that I call $\underline{z}$. (The construction of this threshold is explained by Proposition 1). On the other hand, the refiner does not select j when the draw of z_j is unfavorable, i.e. when $z_j > \underline{z}$. For $j \notin S$, the model predicts a demanded quantity from j in a counterfactual case where j is added to S. (in this counterfactual case, the refiner decides not to select j because the added gain does not cover the fixed cost.) In this counterfactual case, the refiner buys a quantity from j that I call q_j^{CF}, and a 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 $z_j > \underline{z}$. We can write a similar equation as (25) when $j \notin S$,

$$\underbrace{\ln \frac{q_j^{CF}}{q_0^{CF}}}_{y_j | j \notin S} = -\eta \ln \frac{p_j^{obs}}{p_0^{obs}} - \eta \ln z_j, \quad \text{if } j \notin S. \tag{26}$$

Figure A.7: Identification and sample selection bias in estimating trade elasticity η. solid bullets: selected suppliers, circles: unselected suppliers. See text for the definition of y and $\ln(p_j^{obs}/p_0^{obs})$.
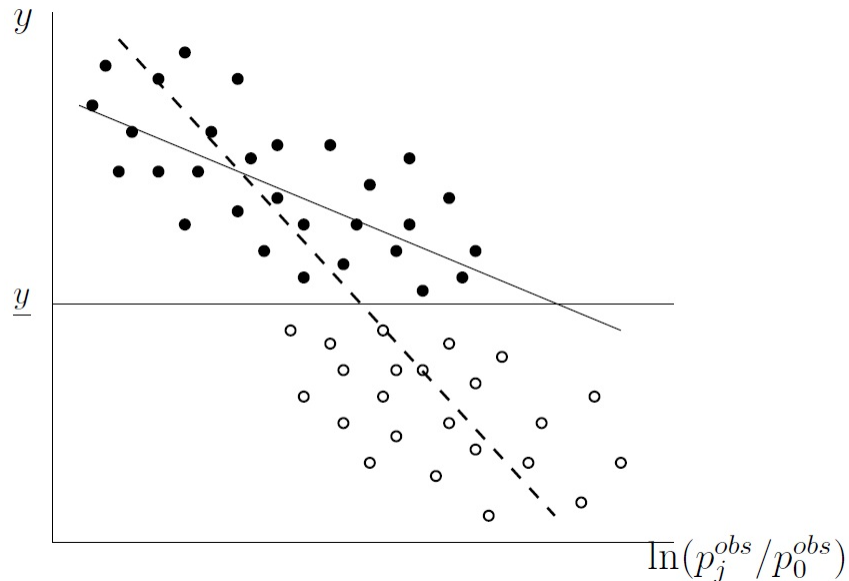
Consider two suppliers j and j' with the same observable costs $p_j^{obs} = p_{j'}^{obs}$. Suppose the refiner has selected j while has not selected j'. The model justifies $j \in S$ and $j' \notin S$ by a small z_j and a large z'_j. Since $p_j^{obs} = p_{j'}^{obs}$, it must be the case that $z_j < z_{j'}$. Thus, according to equations

25-26, $y_j > y_{j'}$. 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. (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 $-\eta \ln(p_j^{obs}/p_0^{obs}) - \eta \ln \underline{z}$. By simplification, in the figure, $\underline{y_j}$'s are the same). 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 (respectively, relating to the sample of the selected suppliers and the whole sample). The smaller slope for the selected sample means an under-estimation of η.

A.4 Country-level data

Table A.10: List of Countries, Selected variables on oil production and consumption, year 2010.

Country	Crude oil production (1000 b/d)	Total refining capacity (1000 b/d)	Avg complexity	Avg utilization rate	Fuel tax rate (%)	Fuel consumption (1000 b/d)
Algeria	1540	450	1.34	0.89	-63	354
Angola	1899	39	1.79	0.72	-22	104
Azerbaijan	1035	399	3.89	0.30	-6	83
Brazil	2055	1908	4.28	0.87	95	2699
Canada	2741	2039	8.14	0.84	65	2283
China	4078	8116	2.73	0.88	55	8938
Colombia	786	286	4.67	0.69	70	270
France	0	1984	6.96	0.78	166	1833
Germany	0	2411	7.90	0.90	157	2467
India	751	2836	3.20	0.95	156	3305
Indonesia	953	1012	3.75	0.74	-27	1487
Iran	4080	1451	3.91	0.95	-92	1811
Iraq	2399	638	4.05	0.56	-4	641
Italy	0	2337	6.87	0.69	156	1544
Japan	0	4624	7.84	0.75	113	4429
Kazakhstan	1525	345	5.25	0.65	-12	234
Korea	0	2702	4.98	0.81	106	2269
Kuwait	2300	936	5.02	0.92	-68	397
Libya	1650	378	1.57	0.86	-78	331
Mexico	2621	1540	7.62	0.86	10	2080
Netherlands	0	1206	7.52	0.81	176	1020
Nigeria	2455	505	4.43	0.36	-13	283
Norway	1869	319	4.39	0.83	197	222
Oman	865	85	2.56	0.85	-50	150
Qatar	1129	339	4.25	0.85	-73	199
Russia	9694	5428	4.38	0.90	12	3135
Saudi Arabia	8900	2080	3.79	0.91	-84	2580
Singapore	0	1357	5.29	0.63	77	1149
Spain	0	1272	7.04	0.72	118	1441
UAE	2415	773	2.44	0.65	-15	615
United Kingdom	1233	1866	8.41	0.76	180	1620
United States	5471	17584	9.77	0.85	15	19180
Venezuela	2216	1282	5.41	0.80	-98	688
RO_America	1408	3022	4.79	0.74	19	2824
RO_Europe	662	5659	7.01	0.75	164	5219
RO_Eurasia	324	2032	4.51	0.48	-2	877
RO_Middle East	1028	944	3.72	0.85	-57	1011
RO_Africa	2257	1906	3.11	0.75	-44	2464
RO_Asia & Oceania	2047	2882	3.84	0.77	80	5904

Figure A.8: Calibrated utilization rates

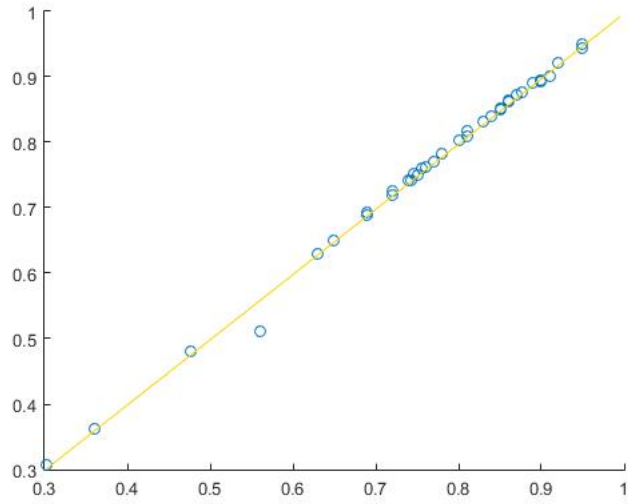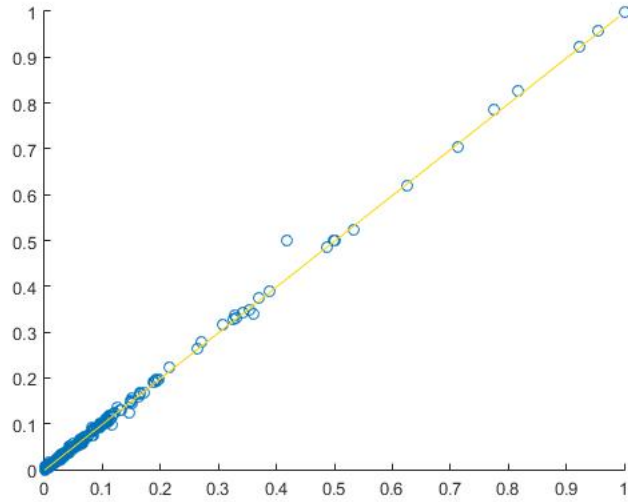Figure A.9: Calibrated trade shares

Table A.11: Percentage change in crude production and refining capacity of countries from 2010 to 2013

Country	production	capacity	Country	production	capacity
Algeria	-3.6	0.0	Netherlands	0.0	-0.9
Angola	-6.0	0.0	Nigeria	-3.5	-11.9
Azerbaijan	-15.6	0.0	Norway	-18.2	0.0
Brazil	-1.5	0.5	Oman	8.7	0.0
Canada	22.2	-6.4	Qatar	37.6	0.0
China	2.1	3.2	Russia	3.3	1.3
Colombia	27.7	1.7	Saudi Arabia	8.8	1.5
France	0.0	-11.9	Singapore	0.0	0.0
Germany	0.0	-6.8	Spain	0.0	0.0
India	2.8	53.2	UAE	16.8	0.0
Indonesia	-13.4	0.0	United Kingdom	-34.3	-10.0
Iran	-21.6	0.0	United States	36.3	1.3
Iraq	27.3	0.0	Venezuela	3.8	0.0
Italy	0.0	-6.1	RO_America	-4.6	2.7
Japan	0.0	2.9	RO_Europe	-8.2	-1.7
Kazakhstan	2.8	0.0	RO_Eurasia	16.2	0.1
Korea	0.0	9.5	RO_Middle East	-77.7	0.0
Kuwait	15.2	0.0	RO_Africa	-12.5	0.0
Libya	-44.3	0.0	RO_Asia & Oceania	-11.5	44.0
Mexico	-2.3	0.0	WORLD	2.2	3.3

A.5 Trade costs from the U.S. to elsewhere

According to the data, U.S. exported crude 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 Table 10 —which parametrizes the calibrated trade costs as a function of distance, common border, and fixed effects. Note that distance and border coefficients as well as importer fixed effects are exogenous to a change in U.S. export barriers. However, lifting the ban changes the U.S. exporter fixed effect. 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 of crude oil 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 Input Price Index

Given sourcing set S and utilization rate u, at each $t \in [0,1]$ the cost for the refiner is a random variable $W(t) = \min_{j} \{p_j / \tilde{\epsilon}_j(t); \ j \in S\}$. Here, by a change of variable $\tilde{\epsilon} = 1/\epsilon$ (in the main text ϵ is used). Also, $Pr(\tilde{\epsilon}_j(t) \leq \tilde{\epsilon}) = \exp(-s\,\tilde{\epsilon}^{-\eta})$ where $s = \left[\Gamma\left(1 + 1/\eta\right)\right]^{\eta}$ ensuring that the shock to p_j is unbiased, i.e. $E[\epsilon_j(t)] = E[1/\tilde{\epsilon}_j(t)] = 1$. The input price index is the average of $W(t)$ over the entire period,

$$P = \int_0^1 W(t)\, dt.$$

The probability distribution of random variable W is given by

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

where $\Phi = s \sum_{j \in S} p_j^{-\eta} = \left[\Gamma\left(1 + 1/\eta\right)\right]^{\eta} \sum_{j \in S} p_j^{-\eta}$. By rewriting the input price index,

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

B.2 Complementary notes on Proposition 1

B.2.1 Diminishing gains from adding suppliers

The variable profit function features decreasing differences if

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

I find a sufficient condition under which the above holds. Then, given the data that are used in my estimation, I show that for the entire space of parameters, variable profit features decreasing differences if $\eta > 1.65$.

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 $J \in \mathbb{N}_+ = \{1, 2, \dots\}$. 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, refiner's decision reduces to choose $\ell \in \mathbb{R}$ suppliers (implied by a

straightforward generalization of lemma 1 to the auxiliary problem). Define $u^{aux}(\ell)$ as the utilization rate, $C^{aux}(\ell) \equiv C(u^{aux}(\ell))$ as the utilization cost, and $y(\ell) \equiv C'(u)|_{u=u^{aux}(\ell)}$. F.O.C implies that

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

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

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

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

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

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

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

One special case of the above relation is where $a = \ell$ and $b = \ell + 2$ 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, as long as ℓ is an integer, we have $\pi^{aux}(\ell) = \pi(\ell)$, $\pi^{aux}(\ell+1) = \pi(\ell+1)$, and $\pi^{aux}(\ell+2) = \pi(\ell+2)$. The above inequality, then, implies

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

which is the definition of decreasing differences. $\square$

According to lemma B.1, to show π features decreasing differences, it suffices to show $(\pi^{aux})'' \equiv \partial^2 \pi^{aux}(L)/\partial L^2 < 0$ —where I have the luxury of taking derivatives.

By taking derivatives of equation (B.2),

$$(\pi^{aux})''(\ell) = y''(\ell)\left(1 - \tilde{P}(\ell)^{1/2}\lambda^{-1/2}y(\ell)^{-1/2}\right) + \frac{1}{2}(y'(\ell))^2\tilde{P}(\ell)^{1/2}\lambda^{-1/2}y(\ell)^{-3/2}. \qquad \text{(B.3)}$$

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

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

$$y''(\ell) = \frac{-(1+\eta)}{\eta^2}\left[\int_0^\ell p^{aux}(\ell')^{-\eta}\,d\ell'\right]^{\frac{-1}{\eta}-2}p^{aux}(\ell)^{-2\eta}$$

$$-\left[\int_0^\ell p^{aux}(\ell')^{-\eta}\,d\ell'\right]^{\frac{-1}{\eta}-1}p^{aux}(\ell)^{-\eta-1}(p^{aux})'(\ell) \qquad \text{(B.5)}$$

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

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

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

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

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

$$\frac{(y')^2}{-y''} \leq \frac{\left\{\frac{1}{\eta}\left[\int_1^L p^{aux}(\ell)^{-\eta}\right]^{\frac{-1}{\eta}-1}p^{aux}(L)^{-\eta}\right\}^2}{\frac{(1+\eta)}{\eta^2}\left[\int_1^L p^{aux}(\ell)^{-\eta}\right]^{\frac{-1}{\eta}-2}p^{aux}(L)^{-2\eta}} = \frac{\left[\int_1^L p^{aux}(\ell)^{-\eta}\right]^{\frac{-1}{\eta}}}{1+\eta} = \frac{P^{aux}}{1+\eta} \qquad \text{(B.7)}$$

Using (B.6) and (B.7), the following delivers a sufficient condition for $(\pi^{aux})'' < 0$,

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

Replacing for $y = \tilde{P} - P^{aux}$, defining $\kappa \equiv \tilde{P}/P^{aux}$, and rearranging inequality B.8 results

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

If inequality B.9 holds, then $(\pi^{aux})'' < 0$, then according to lemma B.1 the variable profit in the original problem features decreasing differences. $\square$

    Relation to the 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 minimum observed utilization rate in the sample. 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

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

Recall that input price index decreases by adding a new supplier (see Section 3.3.1). Therefore, $P^{aux} \leq p_0$, where p_0 is the price of the domestic input; which implies $\kappa > \tilde{P}/p_0$. In the data $\tilde{P}/p_0 = 1.174$ and $u_{\min} = 0.52$. A simple calculation shows that as long as $\eta > 1.65$, 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 On the construction of the lower bound $\underline{p}_B$

Part 1. Define $y \equiv C'$. Then, variable profit is given by

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

Let $\tilde{y} = y^{1/2}$. 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,

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

From the above,

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

The above shows a mapping between the marginal cost of utilization y, and variable profit π.

Part 2. I use superscript *new* for the counterfactual case where a new supplier is added. The maximum variable profit such that adding a supplier is not profitable is achieved at $\pi^{new} = \pi + f$. Using equation B.11, calculate y^{new}:

$$y^{new} = \frac{2\tilde{P}}{\lambda}\left(1 + \sqrt{1 + \frac{\pi^{new}\lambda}{\tilde{P}R}}\right) + \frac{\pi^{new}}{R}, \quad where \quad \pi^{new} = \pi + f.$$

Once we know y^{new}, from F.O.C., calculate $P^{new} = \tilde{P} - y^{new}$. On the other hand, by equation (5), P^{new} equals

$$P^{new} = \left[\sum_{j \in S} p_j^{-\eta} + x^{-\eta}\right]^{\frac{-1}{\eta}} = \left[P^{-\eta} + x^{-\eta}\right]^{\frac{-1}{\eta}},$$

where x is the price of the added supplier, and P is the current price index. From here, I calculate x:

$$x = \left[(P^{new})^{-\eta} - P^{-\eta}\right]^{\frac{-1}{\eta}}.$$

Also, note that by Result 1 the price of added supplier can not be smaller than $p^{maxS} = \max\{p_j; j \in S\}$. It follows that

$$\underline{p}_B = \max\{p^{maxS}, x\}.$$

Notice that by construction, $\underline{z}_B$ depends on λ, $p_A = [p_j; j \in S]$, and f.

B.3 Complementary notes on Proposition 2

B.3.1 High-dimensional integrals in the direct calculation of the likelihood

Let $\mathbf{I}(x)$ stack the exogenous data $\mathbf{D}(x)$ and parameters except the ones for fixed costs,

$$\mathbf{I}(x) = \left[(\eta, \gamma, \theta, \beta_{CI}, \mu_\lambda, \sigma_\lambda), \mathbf{D}(x) \right]. \tag{B.12}$$

The contribution of the refiner to the likelihood equals:

$$L_x(\Omega|\mathbf{D}(x), q(x)) = \int_0^\infty \ell_x(\eta, \gamma, \theta, \beta_{CI}, \mu_\lambda, \sigma_\lambda \mid \mathbf{D}(x), q(x), f) \, \mathrm{d}G_F(f|\Omega_F), \tag{B.13}$$

where, using $\mathbf{I}(x)$ as defined in (B.12), and

$$\ell_x(\eta, \gamma, \theta, \beta_{CI}, \mu_\lambda, \sigma_\lambda \mid \mathbf{D}(x), q(x), f)$$
$$= f_{Q_A}\Big(q_A(x) \,\Big|\, Q_A(x) > 0, \; Q_B(x) = 0; \mathbf{I}(x), f\Big) \times Pr\Big(Q_A(x) > 0, \; Q_B(x) = 0 \,\Big|\, \mathbf{I}(x), f\Big) \tag{B.14}$$

Here, L_x is a function of the vector of parameters Ω, given exogenous variables $\mathbf{D}(x)$, and the observed import volumes $q(x)$ (with $q_A(x) > 0$ and $q_B(x) = 0$). Equation (B.13) expresses L_x as a one-dimensional integration over the *conditional likelihood* ℓ_x, with respect to fixed cost f. Equation (B.14) shows that ℓ_x is the product of the joint p.d.f. of Q_A evaluated at $q_A(x)$ times the probability of $Q_B(x) = 0$, conditional on (i) all parameters except the parameters of fixed cost shocks, $(\eta, \gamma, \theta, \beta_{CI}, \mu_\lambda, \sigma_\lambda)$; (ii) exogenous variables, $\mathbf{D}(x)$; and, (iii) fixed cost f. Finally G_F is the c.d.f. of $f(x)$ as a log-normal distribution specified by (μ_f, σ_f).

Focusing on the contribution of one refiner in the likelihood function, drop superscript x.

Since refiners' decisions are inter-dependent across suppliers, direct calculation of ℓ (in equation B.14) involves high-dimensional integrals. In particular, the last term in equation (B.14) is given by

$$Pr(Q_A > 0, \ Q_B = 0 \mid \mathbf{I}, f) = \int Pr(Q_A > 0, \ Q_B = 0 \mid \lambda, [z_j]_{j \in S}, \ \mathbf{I}, f) \, dG_\lambda(\lambda) \prod_{j \in S} dG_Z(z_j).$$

The dimension of the above integral is $|S|$ which is a two-digit number for some of refiners. For this reason, the likelihood becomes too costly to compute.

B.3.2   Proof of Proposition 2

Re-stating Proposition 1.   It is more notationally correct to write Proposition 1 by using random variables. I 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 c.d.f. and p.d.f. by F_Q and f_Q.

Let $x_A \equiv [\lambda, z_A]$ stack efficiency λ and prices of chosen suppliers z_A, with corresponding random variable $X_A \equiv [\Lambda, Z_A]$. Then, Proposition 1 can be written as follows.

$$(R.1) \quad \left\{ Q_A = q_A \mid Q_A > 0, \ Q_B = 0 \right\} \longleftrightarrow \left\{ X_A = h(q_A) \mid Q_A > 0, \ Q_B = 0 \right\}$$

$$(R.2) \quad \left\{ Q_A > 0, \ Q_B = \mathbf{0} \mid X_A = x_A, \ f \right\} \longleftrightarrow \left\{ z_B \geq \underline{z}_B(x_A, f) \right\} \ and \ \left\{ f \leq \bar{f}(x_A) \right\}$$

The proof uses $(R.1)$ and $(R.2)$ and requires two steps as I explain below.

Step 1. From $[\mathbf{Q_A, Q_B}]$ to $[\mathbf{X_A, Q_B}]$.   The likelihood contribution of the refiner is given by

$$
\begin{aligned}
L \ &= \ f_{Q_A}\!\left(q_A \mid Q_A > 0, \ Q_B = 0\right) \times Pr\!\left\{ Q_A > 0, \ Q_B = 0 \right\} \\
&= \ \left| \partial h(q_A)/\partial q_A \right| \times f_{X_A}\!\left(h(q_A) \mid Q_A > 0, \ Q_B = 0 \right) \times Pr\!\left\{ Q_A > 0, \ Q_B = 0 \right\} \\
&= \ \left| \partial h(q_A)/\partial q_A \right| \times f_{X_A}\!\left(h(q_A) \right) \times Pr\!\left\{ Q_A > 0, \ Q_B = 0 \mid X_A = h(q_A) \right\} \\
&= \ \left| \partial x_A/\partial q_A \right| \times f_{X_A}(x_A) \times Pr\!\left\{ Q_A > 0, \ Q_B = 0 \mid X_A = x_A \right\}. \quad\quad\quad \text{(B.15)}
\end{aligned}
$$

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 second line from the first line in (B.15), I use the first relation in proposition 1, (R.1). Further, suppose that w.l.o.g. h is strictly increasing[51], then

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

Taking derivatives with respect to q_A delivers the result:

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

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

Step 2. From $[\mathbf{X_A}, \mathbf{Q_B}]$ to $[\mathbf{X_A}, \mathbf{Z_B}]$. Using the second relation in proposition 1, (R.2), the last term in equation (B.15) is given by

$$
\begin{aligned}
Pr\Big(Q_A > 0, \ Q_B = 0 \mid X_A = x_A\Big) &= \int_0^\infty Pr\Big(Q_A > 0, \ Q_B = 0 \mid X_A = x_A, \ f\Big) \, dG_F(f \mid \mu_f, \sigma_f) \\
&= \int_0^\infty Pr\Big(z_B \ge \underline{z}_B(x_A, f)\Big) \times I\Big(f \le \bar{f}(x_A)\Big) \, dG_F(f \mid \mu_f, \sigma_f) \\
&= \int_0^{\bar{f}(x_A)} \ell_B(x_A, f) \, dG_F(f \mid \mu_f, \sigma_f), \quad\quad\quad\quad \text{(B.16)}
\end{aligned}
$$

where, by definition, $\ell_B(x_A, f) = Pr\Big\{z_B \ge \underline{z}_B(x_A, f)\Big\}$. Plugging (B.16) into equation (B.15),

$$L = \Big|\partial x_A/\partial q_A\Big| \times f_{X_A}(x_A) \times \int_0^{\bar{f}(x_A)} \ell_B(x_A, f) \, dG_F(f \mid \mu_f, \sigma_f) \quad\quad \text{(B.17)}$$

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

$$f_{X_A}(x_A) = \Big|\partial[\lambda, z_A]/\partial q_A\Big| \times g_\lambda(\lambda) \prod_{j \in S} g_Z(z_j), \quad\quad \text{(B.18)}$$

---

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

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

$$L \;=\; |\partial[\lambda, z_A]/\partial q_A| \times g_\lambda(\lambda) \prod_{j \in S} g_Z(z_j) \times \int_0^{\bar{f}(\lambda, z_A)} \ell_B(\lambda, z_A, f) \, dG(f | \mu_f, \sigma_f). \tag{B.19}$$

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

$$
\begin{aligned}
\ell_B &= Pr\Big\{ z_B \geq \underline{z}_B \Big\} \\
&= 1 - \prod_{j \notin S} Pr\Big\{ z_j < \underline{z}_B(j) \Big\} \\
&= 1 - \prod_{j \notin S} G_Z\Big(\underline{z}_B(j) \Big) \tag{B.20}
\end{aligned}
$$

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

B.4   Simple economy with one supplier and homogeneous refineries

This section presents a simplified version of the main model introduced 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 production Q. In this economy $Q < R$. Let p denote the price of the supplier. For simplicity, assume that refineries directly sell to the end-users, and so, there is no distinction between the wholesale and retail sale markets. By this simplification, let e denote the price of refinery output at the gate of refiners, and also the price index of refined products at the location of end-users.

Let Y, w, and L denote GDP, wage, and population. Then, $Y = wL + pQ$. 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,

$$1 - \beta = \frac{b^{-\rho} e^{1-\rho}}{(1-b)^{-\rho} w^{1-\rho} + b^{-\rho} e^{1-\rho}} \tag{B.21}$$

$$= \frac{\tilde{b} e^{1-\rho}}{w^{1-\rho} + \tilde{b} e^{1-\rho}} \tag{B.22}$$

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

$$\alpha(1 - \beta)(wL + pQ) = eQ. \tag{B.23}$$

By equations B.21-B.23,

$$\frac{Q}{wL + pQ} = \frac{\alpha \tilde{b} e^{-\rho}}{w^{1-\rho} + \tilde{b} e^{1-\rho}} \tag{B.24}$$

On the side of demand for crude oil, refinery utilization cost equals $ec(u)R$. (note: As explained in the text, the unit cost of utilization is the price of refinery output. Here, $c(u) \equiv C(u)/e$.). Refinery's problem is to choose utilization rate u to maximize $(e - p)uR - ec(u)R$. By F.O.C.,

$$ec'(u) = e - p \tag{B.25}$$

Also, by market clearing condition for crude oil $Q = uR$. It is assumed that $c'(Q/R) < 1$. In this model, Q, R, and L are exogenous variables; α, b, ρ are known parameters; p, e, and β are endogenous variables.

The effect of a change in crude production on the prices of crude and refined. I calculate how a change in Q changes p and e. According to equations B.23-B.24,

$$\left[\frac{wL}{wL + pQ}\right] \frac{dQ}{Q} = \left[-\rho - (1 - \rho)\frac{eQ}{\alpha(wL + pQ)}\right] \frac{de}{e} \equiv -\tilde{\rho} \times \frac{de}{e}, \quad where \ \tilde{\rho} \equiv \rho + (1 - \rho)(1 - \beta) \tag{B.26}$$

Here, the elasticity of refined oil price e with respect to production Q approximately equals $1/\tilde{\rho}$ —when pQ/wL is small enough.

Using equation B.25,

$$c'(u)de + e\frac{1}{R}c''(u)dQ = de - dp$$

$$c'(u)de + euc''(u)\frac{dQ}{Q} = de - dp$$

$$c'(u)de - \tilde{\rho}uc''(u)de = de - dp$$

$$dp = de\left[1 - c'(u) + \tilde{\rho}uc''(u)\right] \tag{B.27}$$

Since by F.O.C., $e - p = ec'(u)$, $p = (1 - c'(u))e$. Dividing (B.27) by p,

$$\frac{dp}{p} = \frac{de}{e}\underbrace{\left[1 + \frac{\tilde{\rho}uc''(u)}{1 - c'(u)}\right]}_{Buffer} \tag{B.28}$$

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

$$\left|\frac{dp}{p}\right| > \left|\frac{de}{e}\right|$$

Asymmetric response to changes in utilization rate. In the above, 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 due to the convexity of utilization cost. This asymmetric behavior can be expressed more generally by taking derivatives of $B.25$,

$$\frac{dp}{p} = \frac{de}{e} - \underbrace{\frac{c''(u)}{1 - c'(u)}du}_{Buffer} \tag{B.29}$$