NBER WORKING PAPER SERIES

CRUDE BY RAIL, OPTION VALUE, AND PIPELINE INVESTMENT

Thomas R. Covert
Ryan Kellogg

Working Paper 23855
http://www.nber.org/papers/w23855

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2017

We are grateful to the Sloan Foundation for generous financial support, and we thank participants at the 2017 NBER Hydrocarbon Infrastructure and Transportation Conference and a seminar at Montana State University for valuable comments. We are grateful to Chris Bruegge for generously sharing code to process the STB waybill dataset and to Nathan Lash for valuable research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Crude by Rail, Option Value, and Pipeline Investment
Thomas R. Covert and Ryan Kellogg
NBER Working Paper No. 23855
September 2017
JEL No. L13,L71,L95,Q35

ABSTRACT

The recent large-scale use of railroads to transport crude oil out of newly discovered shale formations has no recent precedent in the U.S. oil industry. This paper addresses the question of whether crude-by-rail is simply a transient phenomenon, owing to delays in pipeline construction, or whether it will be a durable presence in the industry by reducing investment in pipeline infrastructure. We develop a model of crude oil transportation that highlights how railroads generate option value by: (1) giving shippers the ability to flexibly increase or decrease volumes shipped in response to price shocks; and (2) allowing shippers to opportunistically send oil to multiple destinations. In contrast, pipelines have low amortized costs but lock shippers into debt-like ship-or-pay contracts to a single destination. We calibrate this model to the recently constructed Dakota Access Pipeline and find that the elasticity of pipeline capacity to railroad transportation costs lies between 0.24 and 0.61, depending on parameters such as the upstream oil supply elasticity. These values are likely conservative because they neglect economies of scale in pipeline construction and the presence of cost-saving contracting in rail. Our results imply that crude-by-rail is an economically significant long-run substitute for pipeline transportation and that regulatory policies targeting environmental and accident externalities from rail transportation would likely substantially affect pipeline investments.

Thomas R. Covert
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
thomas.covert@chicagobooth.edu

Ryan Kellogg
University of Chicago
Harris School of Public Policy
1155 East 60th Street
Chicago, IL 60637
and NBER
kelloggr@uchicago.edu

1 Introduction

Between the start of 2010 and end of 2014, transportation of crude oil by railroad ("crude-by-rail") in the United States grew from less than forty thousand barrels per day (bpd) to nearly one million bpd. At its peak, crude oil shippers moved more than 10% of total domestic production by rail. This development has no recent precedent: the U.S. Energy Information Administration (EIA) only began tracking the crude-by-rail phenomenon in 2010, and statistics published by the American Association of Railroads (2017) suggest that crude-by-rail volumes between 2006-2010 were, at most, negligible. The rise of crude-by-rail is somewhat surprising in light of the fact that many early pipeline projects were developed as attempts to circumvent the railroad cartel orchestrated by Standard Oil (Granitz and Klein, 1996). Since that time, pipeline transportation has become the dominant form of overland long-distance crude oil transportation, owing to its low amortized per-barrel-mile cost and regulated pricing and access provisions. The sudden resurrection of crude-by-rail is thus puzzling in light of its relatively high amortized costs and provision by a rail industry widely believed to exercise market power in access pricing. In this paper, we try to understand the economic factors that drove the rise of crude-by-rail—emphasizing the flexibility it provides to crude oil shippers—and quantify how crude-by-rail has affected investment in crude oil pipelines.

The simplest and least controversial explanation for crude-by-rail's ascent is that rail transportation was, until recently, the only feasible way to move newly discovered tight oil resources to demand centers. The remarkable speed at which production grew from the Bakken, Niobrara, and other shale formations in the upper Midwest outpaced both the capacity of existing pipeline infrastructure and the ability of pipeline investors to adequately respond. In this story, crude-by-rail is a "stopgap" measure while the years required for pipeline permitting and construction pass. If this is the sole reason behind the ascent of crude-by-rail, it implies that crude-by-rail is merely a transitory phenomenon, driven by an unexpected boom in production in a new place, and that long-run investment in pipeline infrastructure has been unaffected. The delays experienced by recent pipeline projects such as the Dakota Access Pipeline (DAPL, completed in June, 2017) and the Keystone XL project (awaiting permits) are consistent with this story.

An alternative explanation for the rise of crude-by-rail, which is not necessarily exclusive of the "delay" story, is that crude-by-rail may be an attractive transportation option *in spite of* its higher costs. Because rail infrastructure already exists between the upper Midwest and nearly every major refining center in the country, crude-by-rail allows shippers to flexibly decide when and where to ship crude in response to changes in upstream and downstream prices. Industry observers often make this point: for instance, a 2013 *Wall Street Journal* article attributed the lack of shipper interest in a proposed crude oil pipeline from West Texas to California to a preference for the flexibility afforded by crude-by-rail transportation (Lefebvre, 2013). This flexibility is further underscored by the recent substantial fall in crude-by-rail volumes that has followed the late-2014 decrease in oil prices. If this "option value" explanation is sufficiently important, then crude-by-rail will have a durable impact on the U.S. oil industry as a substitute for long-run investment in pipeline infrastructure.

Our paper is aimed at quantifying the economic importance of the option value story by assessing the extent to which crude-by-rail has depressed investment in pipeline capacity. Specifically, we evaluate counterfactual pipeline investment in a world in which crude-by-rail is more expensive and therefore less attractive to crude oil shippers. Beyond shedding light on the economics driving one of the most significant developments in the U.S. crude oil industry in decades, our analysis also addresses policy questions stemming from the disparities in environmental damages that arise from transporting crude oil by pipeline versus railroad. Clay, Jha, Muller and Walsh (2017), for instance, estimates that air pollution damages associated with railroad transportation of crude oil from the Bakken to the East Coast are \$2.73 per barrel, on average, owing primarily to freight locomotives' NO_x emissions.[1] Overall, Clay et al. (2017) finds that air pollution damages from crude-by-rail are nearly twice those from pipeline transportation and are also much larger than estimated damages from spills and accidents. Our results inform how policies, such as emissions equipment regulation or emission pricing, that would address these damages and increase the cost of rail shipping would impact incentives to invest in pipeline capacity and thereby lead to a long-run shift away from railroad transportation.[2]

To quantify the impact of crude-by-rail on pipeline investment incentives, we develop a model in which crude oil shippers can use either a pipeline or a railroad to arbitrage oil price differences between an upstream supply source, where the volume of oil production is sensitive to the local oil price, and downstream markets, where the oil price is stochastic.[3] Pipeline transportation has large fixed costs with potentially significant economies of scale

[1]See table 2 in Clay et al. (2017), noting that there are 42 gallons in a barrel.

[2]In fact, a 2008 EPA rule requires a large reduction in emission rates from newly-built locomotives beginning in 2015 (Federal Register, 2008).

[3]An alternative, "reduced form" strategy for evaluating the impact of crude-by-rail on pipeline construction would be to collect data on pipeline investments and then run regressions to estimate how geographic or temporal variation in railroad transportation costs and railroad use have affected investment. This strategy is impractical, however, since: (1) pipeline investments are infrequent and lumpy (for instance, there have only been three *de novo* pipelines constructed out of the Bakken since the shale boom: Enbridge Bakken, Double H, and Dakota Access (North Dakota Pipeline Authority, 2017)); and (2) variation in railroad costs and utilization is driven by many of the same variables that impact pipeline investment (upstream and downstream crude oil prices, for instance) and is therefore endogenous.

and negligible variable costs. This cost structure is similar to many other "natural monopoly" industries, and as a result, pipeline investment and operations are tightly regulated, usually under cost-of-service rules with common carrier access. Shippers finance pipeline construction by signing long-term (e.g., 10 year) "ship-or-pay" contracts that commit them to paying a fixed tariff per barrel of capacity reserved, whether they actually use the capacity or not, thereby allowing the pipeline to recover its capital expense. Importantly, pipeline shippers must make this commitment knowing only the distribution of possible downstream prices that may be realized during the duration of the contract. If the realized downstream crude oil price is sufficiently high to induce enough upstream production to fill the line to capacity, the resulting wedge between the upstream and downstream prices is the pipeline shippers' reward for their commitment.

In our model, rail provides non-pipeline shippers with a means to arbitrage upstream versus downstream price differences without making a long-term commitment. Instead, rail shippers simply pay a variable cost of transportation (that exceeds the pipeline tariff) whenever they ship crude by rail. Crucially, railroad shippers only pay this cost when they decide to actually ship crude, which they can do after they observe the realized downstream price. This flexibility generates option value, which is further enhanced by the ability of railroads to reach multiple destinations (across which crude oil prices are imperfectly correlated), not just the destination served by the pipeline. In addition, the ability to arbitrage crude oil price differences via rail limits the returns that can be earned by the pipeline shippers, since spatial price differences become bounded by the cost of railroad transportation. Thus, the availability of the rail option reduces shippers' incentives to commit to pipeline capacity. We believe that our model is the first to demonstrate and provide intuition for this effect.

Pipeline capacity in our model is determined by an equilibrium condition in which the marginal shipper is indifferent between committing to the pipeline and relying on railroad transportation. Because the returns to pipeline investment are decreasing in the pipeline's capacity (since a larger pipeline is congested less frequently) the model yields a unique equilibrium level of capacity commitment. We show that the equilibrium capacity investment increases with the cost of railroad transportation, and we derive an expression that relates the magnitude of this key sensitivity to estimable objects such as the distribution of downstream oil prices, the upstream oil supply curve, the cost of pipeline investment, and the magnitude of railroad transportation costs.

To calibrate and validate our model, we collect data from a variety of sources. From Bloomberg and Platts, we obtain prices of crude oil at major refining centers (Brent on the East Coast, Alaska North Slope (ANS) on the West Coast, Louisiana Light Sweet (LLS) on the Gulf Coast, and West Texas Intermediate (WTI) at Cushing, OK) and the price of Bakken crude oil at Clearbrook, MN. We obtain data on crude-by-rail flows from the EIA, and we show that these flows are responsive to price differentials, albeit with a lag of several months to two years. We also obtain data on rail transportation costs from the U.S. Surface Transportation Board (STB) and Genscape; these data show that while railroad transportation charges are roughly constant over time, charges for rail car leases and possibly other complementary goods and services (such as terminal fees) increase with shipping volumes, consistent with the presence of scarcity rents or market power.[4]

With these data in hand, we calibrate our model to match, as best we can, market conditions in June 2014, when the Dakota Access Pipeline (DAPL) announced that it had received firm commitments from shippers to support a 470,000 bpd line. We use our historical data on downstream crude prices to estimate the future distribution of crude prices that shippers faced at that time, and we use the Genscape data to estimate railroad transportation costs as a function of railroad volumes shipped. To obtain a supply curve for oil produced from the Bakken, we combine supply elasticity estimates from the recent literature on shale oil and gas production (Hausman and Kellogg (2015), Newell, Prest and Vissing (2016), Newell and Prest (2017) and Smith and Lee (2017)) with a contemporary forecast that Bakken production would reach 1.5 million bpd within a few years at current oil prices (North Dakota Pipeline Authority, 2014). Given these inputs, we perform a final validation of our model by solving for the pipeline tariff that is implied by an equilibrium in which shippers choose to commit to the actual DAPL capacity (in addition to previously installed export capacity out of North Dakota). The implied tariff from our model is quite close to the actual published DAPL tariff.

Our calibrated model indicates that the development of crude-by-rail has likely had economically meaningful effects on investment in pipeline capacity. We find that a $1 per barrel increase in the cost of rail transportation, a change of about 9% relative to current levels, results in an increase in equilibrium pipeline capacity of between 29,400 and 73,700 bpd, relative to the actual DAPL capacity of 470 thousand bpd (and total Bakken local refining and pipeline export capacity of 1.323 million bpd). These effects are likely to be lower bounds, as they do not account for economies of scale in pipeline construction or for contracts between shippers and rail service providers that likely cause our data to understate the value of crude-by-rail.

Overall, our analysis suggests that crude-by-rail will be more than a "stopgap" transportation option in the U.S. crude oil market. Instead, our model of pipeline investment

[4]This result echoes results from Busse and Keohane (2007) and Hughes (2011) showing evidence of market power and price discrimination in the provision of railroad transportation services for coal and ethanol, respectively.

implies that the option value offered by rail transportation can durably erode the incentive to invest in pipeline capacity. An implication of this result is that policies affecting the relative cost of railroad transport—such as regulations that target rail's environmental externalities—can affect pipeline capacity investment and thereby induce economically significant long-run substitution between pipelines and rail.

The remainder of the paper proceeds as follows. Section 2 presents our model of pipeline investment in the presence of a crude-by-rail option. Then in section 3 we describe our dataset, followed by a discussion in section 4 of the empirical relationship between railroad flows of crude oil and variation in oil price differentials. Section 5 then discusses how we calibrate and validate our model. Section 6 presents our results, and section 7 concludes.

2 A model of pipeline investment in the presence of a rail option

This section presents a model that captures what we believe are the essential tradeoffs between pipeline and railroad transportation of crude oil. The central tension in the model is the balance between the low cost of pipeline transportation and the flexibility afforded by rail. Our aim is to capture how factors such as transportation costs and expectations about future prices for crude oil affect firms' decisions, on the margin, to invest in pipeline capacity versus rely on the railroads.

We begin by building intuition with a simple version of our model in which there is only a single destination that can be reached by pipeline or rail. We then expand the model to allow for the possibility that rail can be used to flexibly deliver crude to alternative destinations when those destinations yield higher netbacks.

2.1 Setup of single destination model

The simplest version of our model involves a single "upstream" destination that supplies crude oil and a single "downstream" destination where oil is demanded. Transportation decisions are made by shippers who purchase crude oil at the upstream location, pay for pipeline or railroad transportation service, and then sell the oil at the downstream location. The essential difference between the two modes of transportation is that construction of the pipeline—the cost of which is completely sunk and must be financed by pipeline shippers' commitments—must occur before the level of downstream demand is realized. Railroad shippers, on the other hand, can decide whether or not to use the railroad after observing downstream demand.

The model assumes that rail shippers use spot crude and transportation markets, so that rail volumes respond immediately to price variation. As we show in sections 3 and 4, however, rail flows in practice follow price movements with a lag of up to two years, owing to contracts among shippers and transporters. We discuss the implications of these contracts for the interpretation of our model in section 4.1.

We model shippers as atomistic, so that they are price takers in both the upstream and downstream crude oil markets, and in the market for transportation services. This assumption is motivated by the large number of potential parties who may act as shippers: upstream producers, downstream refiners, and speculative traders. The equilibrium level of pipeline investment is then governed by an indifference condition in which, on the margin, shippers' expected per-barrel return to committing to the pipeline equals the amortized per-barrel cost of the line (which is then the pipeline's tariff for firm capacity).

We now derive this equilibrium condition and examine the forces that govern it. Begin with the following definitions:

- Let $P(Q)$ denote the upstream inverse net supply curve for crude oil. Q denotes the total volume of oil exported from upstream to downstream. In the context of North Dakota, this curve represents supply of crude oil from the Bakken formation net of local crude demand. $P'(Q) > 0$. (For brevity, we henceforth refer to $P(Q)$ as the supply curve rather than "inverse net supply".) Let $P_u = P(Q)$ denote the upstream oil price.

- The downstream market (e.g., a coastal destination that can access the global waterborne crude oil market) is sufficiently large that demand is perfectly elastic at the downstream price P_d. P_d is stochastic with a distribution given by $F(P_d)$, with support $[\underline{P}, \bar{P}]$.[5]

- K denotes pipeline capacity. The cost of capacity is given by $C(K)$, with $C'(K) > 0$ and $C''(K) \leq 0$. Shippers that commit to the pipeline must pay, for each unit of capacity committed to, the average cost $C(K)/K$, thereby allowing the pipeline to recover its costs ($C(K)$ implicitly includes the regulated rate of return). Given capacity, the marginal cost of shipping crude on the pipeline up to the capacity constraint is zero.[6]

[5] Uncertainty about P_d is isomorphic to uncertainty about the intercept of the upstream supply function $P(Q)$. Thus, our model can in principle accommodate uncertainty about upstream supply.

[6] This zero marginal cost assumption reflects the fact that the marginal cost of pumping an additional barrel of oil per day through a pipeline is quite small relative to the amortized cost of constructing a marginal barrel per day of pipeline capacity.

- Let Q_p denote the volume of crude shipped by pipe, and let Q_r denote the volume of crude shipped by rail. $Q = Q_p + Q_r$.
- The marginal cost of shipping by rail is given by $r(Q_r)$, where $r_0 \equiv r(0) > 0$ and $r'(Q_r) \geq 0$.

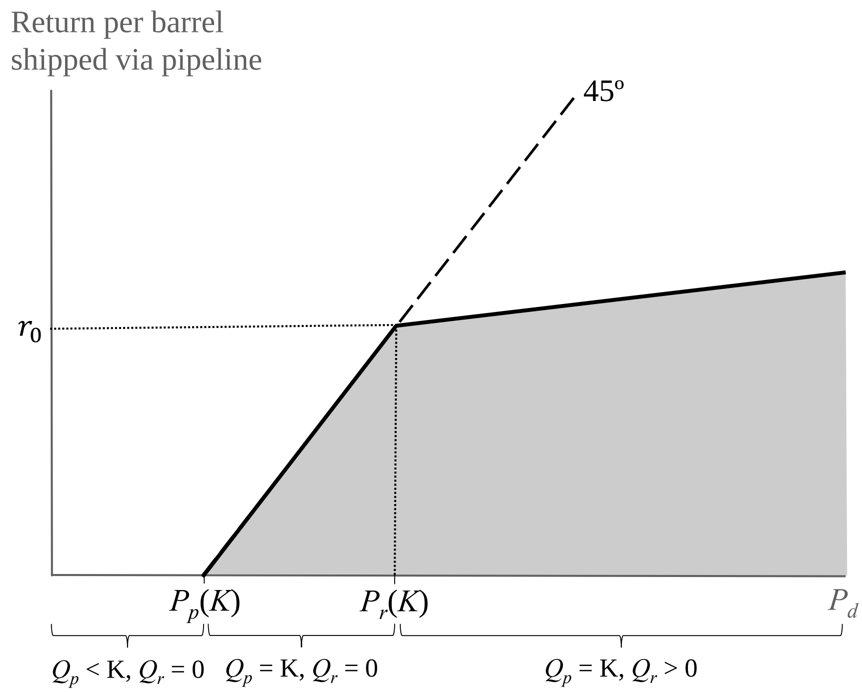
Given a pipeline capacity K, the pipeline and rail flows Q_p and Q_r are determined by the realization of the downstream price P_d. For very low values of P_d, little crude oil is supplied by upstream producers, and the pipeline is not filled to capacity ($Q = Q_p < K$). Arbitrage then implies that $P_u = P_d$. Because the upstream supply curve is strictly upward-sloping, increases in P_d lead to increases in quantity supplied, eventually filling the pipeline to capacity. Let $P_p(K) = P(K)$ denote the minimum downstream price such that the pipeline is full.

For downstream prices $P_d > P_p(K)$, no more oil can flow through the pipeline, but rail may be used. Crude oil volumes will move over the railroad only to the extent that the differential between P_d and P_u covers the marginal cost $r(Q_r)$ of railroad transport. Define $P_r(K)$ as the minimum downstream oil price such that railroad transportation is used. This price is defined by $P_r(K) = P(K) + r_0$. Thus, there is an interval of downstream prices, $[P_p(K), P_r(K)]$, for which pipeline flow $Q_p = K$, rail flow $Q_r = 0$, and the upstream price is constant at $P(K)$. For downstream prices that strictly exceed $P_r(K)$, railroad volumes will be strictly positive and determined by the arbitrage condition $P_u = P(K+Q_r) = P_d - r(Q_r)$. This arbitrage condition implies a function $Q_r(P_d)$ that governs how rail flows increase with P_d when $P_d > P_r(K)$.

2.2 Equilibrium pipeline capacity in the single destination model

Prospective shippers will be willing to make the up-front investment in the pipeline if the expected return from owning the right to use pipeline capacity meets or exceeds the investment cost. This cost, on the margin, is simply given by the average per-barrel cost $C(K)/K$. The expected return to pipeline capacity is given by the expected basis differential $P_d - P_u$. Figure 1 provides the intuition for how this expected return depends on capacity K, the rail cost function $r(Q_r)$, and the distribution $F(P_d)$. When the downstream price P_d is less than $P_p(K)$, the return to capacity is zero because the pipeline is not full and $P_u = P_d$. For $P_d \in [P_p(K), P_r(K)]$, the return $P_d - P_u$ falls on the 45° line, since rail flows are zero and P_u is therefore fixed at $P_p(K)$. Finally, for $P_d > P_r(K)$, the basis differential is simply equal to the cost of railroad transportation $r(Q_r)$, since arbitrage by rail shippers equates $P_d - P_u$ to $r(Q_r)$. When $P_d > P_r(K)$, the differential $P_d - P_u$ strictly increases with P_d, as shown in

Figure 1: Expected return achieved by pipeline shippers



Note: P_d denotes the downstream price, with distribution $F(P_d)$. Q_p and Q_r denote crude oil pipeline and rail flows, respectively. r_0 denotes the intercept of the rail marginal cost function $r(Q_r)$. The shaded area, probability-weighted by $F(P_d)$, represents the expected return to a pipeline with capacity K. See text for details.

figure 1, iff $r'(Q_r) > 0$.[7]

The expected return to pipeline shippers is then given by the shaded area in figure 1, weighted by the probability distribution $F(P_d)$. When prospective shippers must make pipeline commitments before knowing the realization of P_d, the equilibrium capacity K will balance this expected return (which decreases in K) against the pipeline investment cost of $C(K)/K$ for a marginal barrel per day of capacity. Figure 1 thereby illustrates how the presence of the option to use rail transportation weakens the incentive to increase pipeline capacity: absent rail, the expected return to a pipeline of capacity K would be the entire triangle between the horizontal axis and the 45° line, rather than the shaded trapezoid shown in the figure.

Formally, the condition that governs the equilibrium capacity level is given by equation (1), where the first term on the right-hand side captures returns to pipeline shippers when the pipeline is at capacity but rail is not used, and the second term captures returns when

[7] The relation between $P_d - P_u$ and P_d need not be affine, as shown in the figure.

P_d is sufficiently high that the pipeline is at capacity and rail flows are strictly positive:

$$\frac{C(K)}{K} = \int_{P_p(K)}^{P_r(K)} (P_d - P_p(K)) f(P_d) dP_d + \int_{P_r(K)}^{\bar{P}} r(Q_r(P_d)) f(P_d) dP. \tag{1}$$

Even though we assume that shippers are atomistic, the equilibrium pipeline investment implied by equation (1) will differ from the socially optimal investment if there are returns to scale in pipeline construction. Social welfare is maximized when the expected return to shipping crude oil via pipeline is equated to the marginal cost of construction $C'(K)$, not average cost $C(K)/K$. In the presence of scale economies, $C'(K) < C(K)/K$ so that the optimal pipeline capacity K^* is strictly greater than the equilibrium capacity from equation (1). This divergence between market and socially optimal investment in the presence of increasing returns to scale is driven by average-cost regulation of pipeline tariffs and is emblematic of rate regulation in many natural monopoly settings.

Our model illuminates the comparative static of primary interest in this paper: how does pipeline capacity investment respond to changes in the cost of rail transportation? Figure 2 provides the intuition. Consider a pipeline project that, facing a railroad cost curve with intercept r_0, would attract equilibrium commitments from shippers for a capacity of K. Now suppose that the rail cost intercept were instead $r_0' > r_0$. This increase in rail transportation cost increases the basis differential realized by pipeline shippers whenever rail transportation is used, as indicated in the upper, striped shaded area. This increase in expected return then increases shippers' willingness to commit to capacity, so that equilibrium pipeline capacity must increase. The new capacity level K' balances the increase in expected return when rail is used against the decrease in expected return caused by the decrease in the probability that the pipeline is fully utilized (represented by the lower shaded area in figure 2).

Formally, we obtain the comparative static dK/dr_0 by applying the implicit function theorem to equation (1).[8] To simplify the problem, we assume that the railroad marginal cost function is affine: $r(Q_r) = r_0 + r_1 Q_r$. We then obtain:[9]

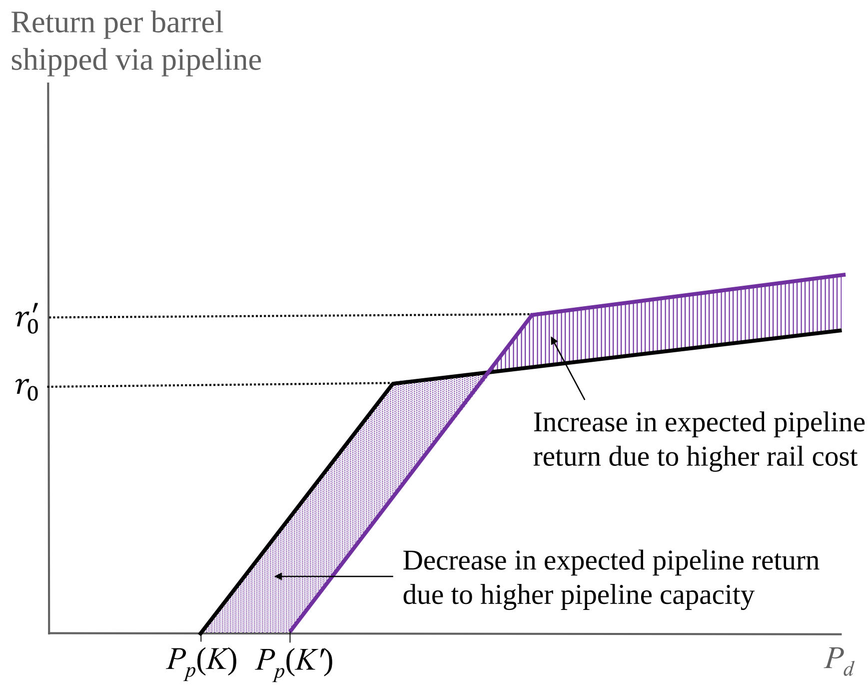
$$\frac{dK}{dr_0} = \frac{1 - F(P_r(K)) - \int_{P_r(K)}^{\bar{P}} \frac{r_1}{P'(K+Q_r(P_d))+r_1} f(P_d) dP_d}{\frac{d}{dK}\frac{C(K)}{K} + \int_{P_p(K)}^{P_r(K)} P'(K) f(P_d) dP_d + \int_{P_r(K)}^{\bar{P}} r_1 \frac{P'(K+Q_r(P_d))}{P'(K+Q_r(P_d))+r_1} f(P_d) dP_d}. \tag{2}$$

First, note that in the simple case in which $r_1 = 0$ and $C(K)$ exhibits constant returns

[8] To derive equation (2), we also apply the implicit function theorem to the arbitrage condition $P(K + Q_r) = P_d - r(Q_r)$ that defines the rail flow function $Q_r(P_d)$.

[9] Note that the term involving the derivative of $P_p(K)$ is equal to zero, and the terms involving the derivative of $P_r(K)$ cancel.

Figure 2: Comparative static: equilibrium capacity varies with rail cost intercept r_0



Note: P_d denotes the downstream price, with distribution $F(P_d)$. K denotes equilibrium pipeline capacity when the rail cost curve has intercept r_0; K' denotes equilibrium pipeline capacity when the rail cost curve has intercept r_0'. See text for details.

to scale, this expression just reduces to:

$$\frac{dK}{dr_0} = \frac{1 - F(P_r(K))}{\int_{P_p(K)}^{P_r(K)} f(P_d)dP_d} \frac{1}{P'(K)}. \tag{3}$$

In words, this is the ratio of the probability rail is used and the probability that the pipeline capacity constraint binds but rail is not used, multiplied by the inverse of the slope of the supply curve at K. The intuition for this expression is that: (1) if rail is likely to be used, shocks to r_0 are costly, so the optimal K will be sensitive to such shocks; (2) if there is a low probability that the pipe is full but no rail is being used, the returns to capacity do not rapidly diminish in K, so shocks to r_0 can yield large changes in the optimal K; and (3) if the upstream supply curve is steep, then the response of supply to transportation cost shocks is low, so that increases in r_0 do not call for large increases in K.

Second, note that the terms involving r_1 reduce the numerator of equation (2) and increase the denominator, so that if railroad transportation costs are sensitive to railroad transportation volumes, the sensitivity of K to r_0 is reduced. Intuitively, large values of r_1 reduce the option value of rail transportation, so that pipeline economics are then less sensitive to shocks to railroad transportation costs.

Finally, when there are increasing returns to scale, so that $d(C(K)/K)/dK < 0$, dK/dr_0 will be larger than in the case of constant returns. Thus, it is not possible to unambiguously rank the simple case in equation 3 and the general case in equation 2.

Expression (2) also clarifies the information required to obtain an estimate of dK/dr_0. Specifically, evaluating expression (2) requires values for:

1. The distribution $F(P_d)$ of downstream crude oil prices at the time that shippers make commitments.

2. The slope of the upstream crude oil supply curve $P(Q)$.

3. The parameters r_0 and r_1 that govern the railroad transportation cost function $r(Q_r)$.

4. The cost structure of pipeline construction; i.e., $C(K)$.

Section 5 discusses our calibration of the above parameters, which uses estimates from our own calculations as well as estimates from prior studies.

2.3 Modeling multiple railroad destinations

This section considers how the spatial option value afforded by railroads affects the tradeoff shippers face between pipeline and railroad transportation. We augment the model described above by allowing for multiple downstream destinations at which crude oil prices are imperfectly correlated with P_d, the price at the location served by the pipeline. Railroad shippers can deliver crude to these locations after observing the realized price at each, whereas shippers on the pipeline can only deliver crude to the pipeline destination.

Specifically, we make the following changes to the model presented in section 2.1:

- Let $\tilde{P}$ denote the maximum of the set of prices across all downstream locations (demand for Bakken crude is perfectly elastic at each location), and let $F(\tilde{P} \mid P_d)$ denote the distribution of $\tilde{P}$ conditional on P_d (where P_d again denotes the downstream price at the destination served by the pipeline). In the case of Bakken crude transportation—where the downstream destinations are the East Coast, Gulf Coast, Cushing, and the West Coast, $\tilde{P}$ and P_d will be highly correlated and often identical. By construction, $\tilde{P} \geq P_d$.

- Assume that the cost of shipping by rail to any location is identical and given by $r(Q_r)$ as described in section 2.1. This assumption implies that rail shippers will send all rail volumes to the downstream location with the highest price and thereby obtain $\tilde{P}$. We discuss violations of this prediction in our data in section 4.

- Assume that $\tilde{P} - P_d$ is bounded above by r_0. This assumption implies that there will be no railroad shipments to any location whenever the pipeline does not operate at full capacity. This assumption substantially simplifies the model. We discuss the empirical validity of this assumption in section 5.1.2.

As in section 2.1, define $P_p(K)$ as the minimum P_d such that the pipeline is full. Again, we have $P_p(K) = P(K)$. Similar to section 2.1, define $P_r(K)$ as the minimum $\tilde{P}$ such that rail is used. Again, we have $P_r(K) = P(K) + r_0$.

We now derive the equilibrium relationship governing the pipeline capacity built. As before, the marginal committed shipper must pay $C(K)/K$ regardless of the realization of P_d or $\tilde{P}$. By committing, the pipeline shipper can again earn returns in two states of the world: (1) the pipeline is full but no rail is used; and (2) the pipeline is full and rail is being used.

If the pipeline is full but rail is not used, the upstream price is simply given by $P(K)$, and shippers earn $P_d - P(K)$. This situation occurs when $P_d \geq P_p(K)$ and $\tilde{P} \leq P_r(K)$. Thus, the expected value for committed shippers that accrues when the pipe is full but rail is not used is given by:

$$\int_{P_p(K)}^{\bar{P}} \left[\int_{P_d}^{P_r(K)} (P_d - P(K)) f(\tilde{P}|P_d) d\tilde{P} \right] f(P_d) dP_d. \tag{4}$$

When rail is used (and the pipeline is therefore full), the upstream price is given by $P(K + Q_r(\tilde{P}))$, and pipeline shippers earn $P_d - P(K + Q_r(\tilde{P}))$. This situation occurs when $P_d \geq P_p(K)$ and $\tilde{P} \geq P_r(K)$. Thus, the expected value for committed shippers that accrues when rail is used is given by:

$$\int_{P_p(K)}^{\bar{P}} \left[\int_{P_r(K)}^{\bar{P}} (P_d - P(K + Q_r(\tilde{P}))) f(\tilde{P}|P_d) d\tilde{P} \right] f(P_d) dP_d. \tag{5}$$

Equilibrium capacity K is therefore given by equation (6), which collapses to the single-destination equilibrium equation (1) if $\tilde{P}$ is always equal to P_d.

$$\frac{C(K)}{K} = \int_{P_p(K)}^{\bar{P}} \left[\int_{P_d}^{P_r(K)} (P_d - P(K)) f(\tilde{P}|P_d) d\tilde{P} \right] f(P_d) dP_d$$
$$+ \int_{P_p(K)}^{\bar{P}} \left[\int_{P_r(K)}^{\bar{P}} (P_d - P(K + Q_r(\tilde{P}))) f(\tilde{P}|P_d) d\tilde{P} \right] f(P_d) dP_d. \quad (6)$$

For a given capacity K, each term on the right-hand side of equation (6) will be smaller than the corresponding term on the right-hand side of equation (1). Thus, the equilibrium pipeline capacity in the presence of multiple rail destinations will be smaller than the case in which rail can only serve a single destination. An objective of our empirical work is to calculate the magnitude of this difference in equilibrium pipeline capacity and thereby understand the economic importance of the spatial option value afforded by railroads.

We now apply the implicit function theorem to determine dK/dr_0:[10]

$$\frac{dK}{dr_0} = \frac{\int_{P_p(K)}^{\bar{P}} \left[\int_{P_r(K)}^{\bar{P}} (1 - \frac{r_1}{P'(K+Q_r(\tilde{P}))+r_1}) f(\tilde{P}|P_d) d\tilde{P} \right] f(P_d) dP_d}{\frac{d(C(K)/K)}{dK} + \int_{P_p(K)}^{\bar{P}} \left[\int_{P_d}^{P_r(K)} P'(K) f(\tilde{P}|P_d) d\tilde{P} + \int_{P_r(K)}^{\bar{P}} \frac{r_1 P'(K+Q_r(\tilde{P}))}{P'(K+Q_r(\tilde{P}))+r_1} f(\tilde{P}|P_d) d\tilde{P} \right] f(P_d) dP_d} \quad (7)$$

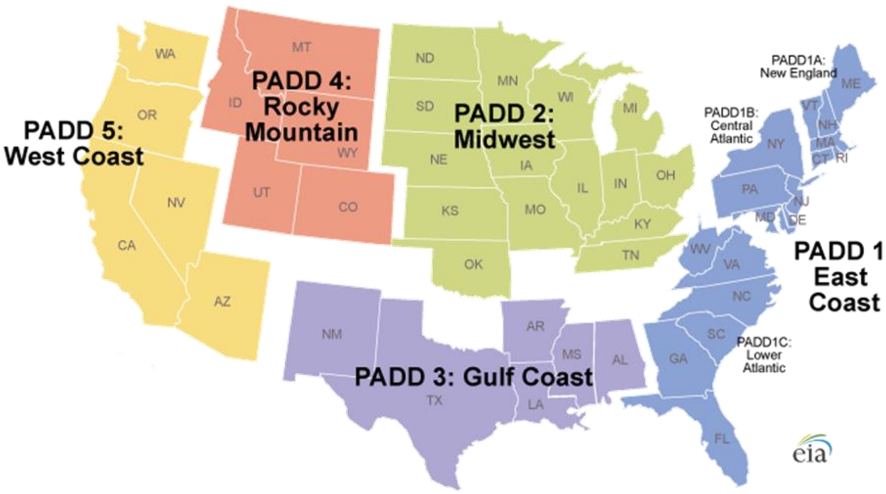
The most important difference between equations (7) and (2) is that the presence of multiple rail destinations decreases the probability that the pipe is full but rail is not used. This change causes the second term in the denominator of (7) to be smaller than the corresponding term in (2), thereby increasing the sensitivity dK/dr_0 of pipeline capacity to the cost of rail transport.[11]

As in the case with a singe rail destination, we can calibrate this relationship using assumptions about the upstream supply curve, the parameters governing costs and the distribution of downstream prices. However, in this case, instead of just needing the unconditional distribution of the "pipeline" downstream price $F(P_d)$, we now additionally require the distribution of best rail prices *conditional* on that price: $F(\widetilde{P} \mid P_d)$.

[10] Note that the terms involving $P_p(K)$ are equal to zero, and the terms involving $P_r(K)$ cancel.

[11] When $r_1 > 0$ the impact of multiple rail destinations on the final terms in the numerator and denominator of equation (7) is ambiguous, so that the overall comparison of dK/dr_0 between equations (7) and (2) is also ambiguous. Nonetheless, we find in practice that dK/dr_0 is roughly 10% larger when we evaluate our model allowing for crude-by-rail to flow to multiple destinations.

Figure 3: Map of EIA Petroleum Administration for Defense Districts (PADDs)



Source: EIA

3 Data

3.1 Crude oil price data

We obtained data on spot market crude oil prices from Bloomberg and Platts.[12] We use the price of Bakken crude at Clearbrook, MN as the "upstream" market price, and we use prices for Brent, Louisiana Light Sweet (LLS), and Alaska North Slope (ANS) as benchmark prices for U.S. East Coast, Gulf Coast, and West Coast "downstream" destinations, respectively. These downstream pricing points are located in the EIA's "Petroleum Area for Defense Districts" (PADDs) 1, 3, and 5, respectively, as shown in figure 3. Finally, we also use the price of West Texas Intermediate (WTI) at Cushing, OK as another destination. Both Cushing and Clearbrook are located in PADD 2.

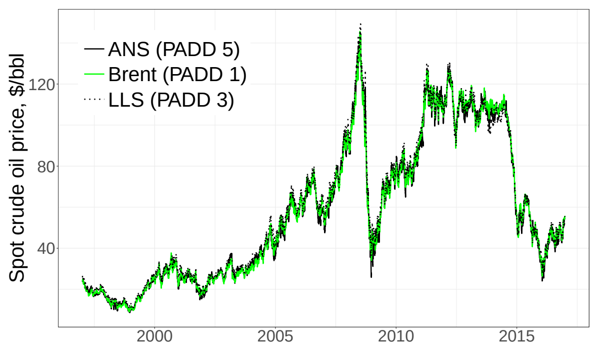
Because Bakken crude oil is quite light, the Brent and ANS benchmarks may understate the value of Bakken crude on the East and West Coasts. In future work, we hope to use data on refined product prices and crude assays to correct the Brent and ANS prices for grade differences. For now, using the Bloomberg prices directly will cause us to understate the spatial option value of crude-by-rail.

Figure 4 plots our spot price data, aggregated to the monthly level. Panel (a) shows that prices at the three coastal destinations are very tightly correlated, typically differing

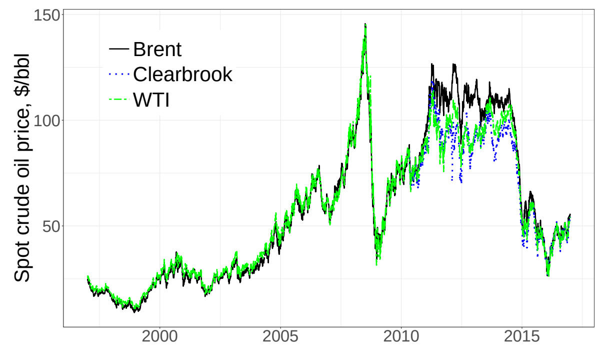
[12] We use Bloomberg prices for Brent, WTI, ANS and LLS and we use Platts prices for Clearbrook. Though Bloomberg does publish a Clearbrook price series, the Clearbrook series from starts 5 months earlier than the series from Bloomberg.

Figure 4: Crude oil spot prices

(a) ANS, Brent, and LLS prices



(b) Brent, Clearbrook, and WTI prices



Source: Bloomberg. "WTI" refers to West Texas Intermediate, "ANS" refers to Alaska North Slope, and "LLS" refers to Louisiana, Light, Sweet. See text for details.

by no more than a few $/bbl over the last 20 years. Moreover, no single destination has a consistent price advantage over another. In contrast, panel (b) shows that the PADD 2 pricing locations at Clearbrook and Cushing were substantially discounted relative to coastal destinations prior to mid-2014.[13] In addition, both panels of figure 4 clearly illustrate the substantial decrease in crude oil prices that occurred during the second half of 2014.

3.2 Crude oil flow data

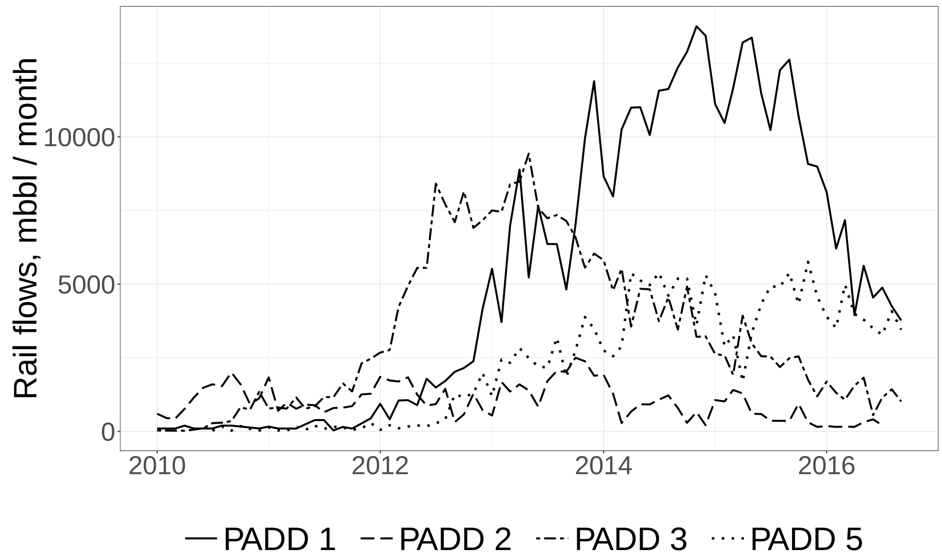
Data on monthly PADD-to-PADD flows of crude-by-rail were obtained from the EIA.[14] Figure 5 presents data on crude oil shipments from PADD 2 to PADDs 1, 2, 3, and 5.[15] Volumes are dominated by shipments to the coastal destinations rather than intra-PADD 2 shipments, according with both the depressed WTI price early in the sample and the fact that rail transport to Cushing competed with pipelines, whereas transportation to the coasts did not. Shipments to the coasts rise substantially beginning in 2012, plateau in late 2014, and then begin to fall substantially in late 2015. The rise and fall of crude-by-rail is consistent with the rise and fall in spatial price differentials shown in panel (b) of figure 4, though changes in rail volumes appear to follow changes in price differentials with a substantial lag. We examine the relationship between oil prices and crude-by-rail flows in detail in section 4.

[13] Note that the Clearbrook price series does not begin until May 1 2010.

[14] We used the EIA's API to obtain the crude-by-rail data available online at https://www.eia.gov/dnav/pet/PET_MOVE_RAILNA_A_EPC0_RAIL_MBBL_M.htm.

[15] Shipments from PADD 2 to PADD 4 are zero, according with the fact that PADD 4 only exports crude.

Figure 5: Crude-by-rail monthly volumes from PADD 2, by destination PADD



Source: EIA

3.3 Railroad transportation cost data

We obtained data on the cost of transporting crude by rail from the U.S. Surface Transportation Board (STB) and from Genscape, a private industry intelligence firm. The STB is the United State's regulator of interstate railroads, and we were able to obtain a data use agreement for the STB's restricted-access waybill sample datasets for 2009–2014. These data contain detailed information on volumes and carrier revenues for a sample of shipments of crude oil and many other commodities.[16]

We use the STB data to examine the time series variation in transportation rates charged to shippers of crude oil. For each month of our sample,[17] we calculate the total revenue (across all shipments originating in PADD 2 and delivered to PADDs 1, 3, or 5) earned by railroad carriers and divide by the total number of bbl-miles of crude oil shipped.[18] Figure 6 presents the resulting time series of average revenue per bbl-mile.[19] While noisy, the data

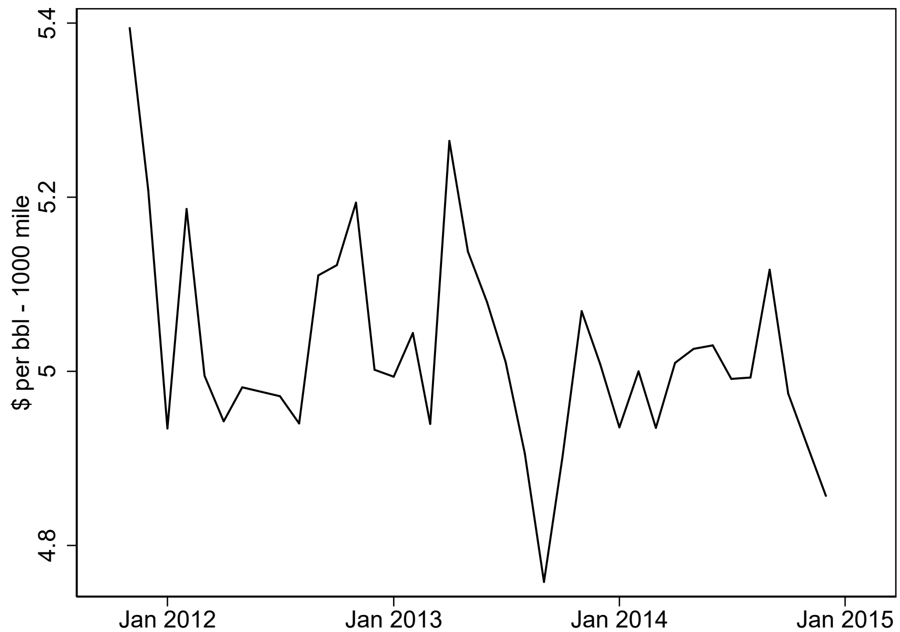
[16]To isolate the waybill sample to crude oil shipments, we only keep shipments with a Standard Transportation Commodity Code (STCC) of 1311110.

[17]To assign a movement date (and therefore a month) to each shipment, we follow the procedure described in Energy Information Administration (2017) to convert waybill dates to movement dates.

[18]The revenue measure we use is the sum of total freight line-haul revenue with fuel surcharges. We obtain total revenue and bbl-miles across all shipments each month using expansion factors that account for variation in sampling rates for shipments of different sizes.

[19]To protect the confidentiality of individual carriers' rates, we are unable to present results that are disaggregated (either spatially or temporally) beyond the monthly level. We are also unable to present data from before November, 2011.

Figure 6: STB average revenue per bbl-mile shipped



Note: Data cover sampled waybills from PADD 2 to PADDs 1, 3, and 5. Data from June 2012 are omitted to protect the confidentiality of carrier's rates; the May 2012 and July 2012 observations are connected by a line. See text for details.

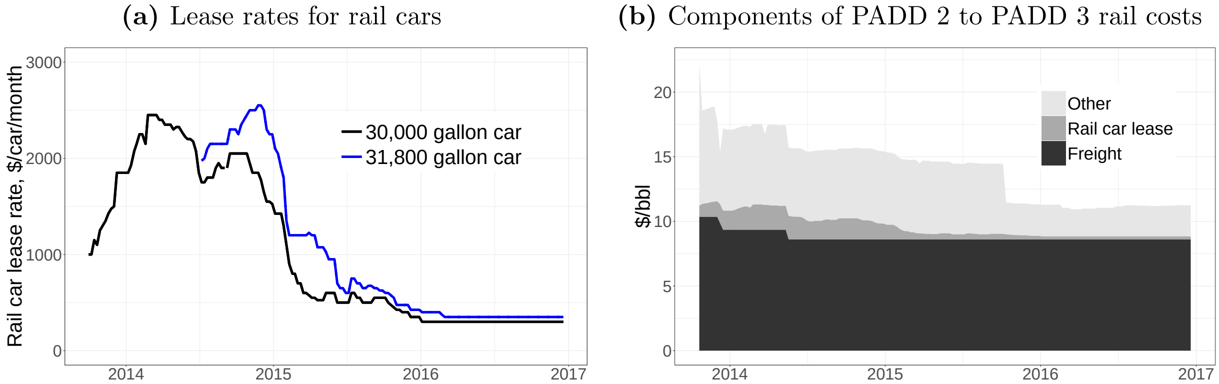
indicate that railroad transportation rates were roughly constant from 2012–2014, despite substantial variation in crude basis differentials and railroad transportation volumes during this period.

The STB dataset also includes information on whether each shipment was under a "tariff" or "contract" rate. Tariff shipments are charged a publicly-posted tariff that is available to any shipper under common-carry regulation. Contract shipments are under negotiated rates that may include volume commitments or discounts, and typically have a term of 1-2 years, according to industry participants. The STB data indicate that 88.0% of crude oil moves on contract rates and that contract shipments enjoy an average discount of $0.65 per 1000 bbl-mile relative to tariff rates.[20]

Our data from Genscape complement the STB data by providing information on the cost of leasing rail cars—which are provided by third parties, not the railroads themselves—as well as other elements of crude oil transportation, such as loading and unloading terminal fees.

[20]Both of these numbers use the same sample of waybills used to generate figure 6. To obtain the contract discount, we regress revenue per 1000 bbl-mile on a tariff vs contract dummy variable and on month-of-sample fixed effects, while weighting the regression by bbl-miles. $0.65 is the regression coefficient on the tariff vs contract dummy variable. If we do not weight the regression, we obtain a $0.73 contract discount.

Figure 7: Cost assessments from Genscape



Source: Genscape. Rail car lease rates are not PADD-specific. PADD 2 to PADD 3 decomposition assumes that rail cars are 30,000 gallons and complete 1.75 round trips per month. See text for details.

Unlike the STB data, Genscape's data are cost assessments rather than actual transaction data: each week, Genscape surveys shippers to determine their estimates of the cost of making a spot crude shipment to a particular destination. The Genscape data series begin in October, 2013.

Panel (a) of figure 7 presents Genscape's assessments of leasing rates for rail cars. Unlike the STB data on railroad transportation fees, the cost of rail car rental is not constant over time. Lease rates rise in the first part of the sample, when the oil price is high and transportation volumes are growing, and then fall late in the sample, when the oil price is low and transportation volumes are falling. This pattern is consistent with scarcity rents during the "boom" period that then dissipated when oil prices fell.[21]

Figure 7, panel (b) shows how Genscape's assessments of freight costs (monies paid to the railroad, equivalent in principle to the STB revenue data), rail car lease costs, and other costs (primarily terminal fees) come together to form the total cost of shipping crude by rail, using the PADD 3 destination as an example. First, it is clear that Genscape's freight cost assessments are rarely updated, as the freight cost is constant for much of the sample. Nonetheless, the level of freight costs is consistent with the STB data shown in figure 6, given an average trip distance to the gulf coast of roughly 1,900 miles (Clay et al. (2017)). Panel (b) of figure 7 also indicates that charges such as terminal fees are economically significant, exceeding $2/bbl even at the end of the sample when the price of crude oil is low. The substantial jump in "Other" costs in late 2015 is observed in all PADDs in the Genscape

[21] See Tita (2014) and Arno (2015) for discussions of the boom and bust in railcar lease rates.

data and reflects the removal of gathering costs from Genscape's assessments rather than an actual change in costs.

4 Rail flows and price differentials

The model from section 2 assumes that rail shippers will (a) select the destination with the highest upstream price, net of transportation costs; and (b) adjust the magnitude of these shipments as upstream prices rise and fall. In practice, however, crude-by-rail shipments violate both of these assumptions: figure 5 shows that every destination has positive rail flows in every month, starting in 2012, and figures 5 and 4 together suggest that rail volumes follow price movements with a lag. This section examines these departures from the model and discusses their implications for the interpretation of the model's results.

A likely primary driver of the deviations of rail flows from assumptions (a) and (b) is the presence of contracts between shippers, railroads, end users, and logistics providers. These contracts are frequently mentioned in industry press and publications, such as Hunsucker (2015), and may contain provisions that guarantee minimum volumes or provide volume discounts, over time horizons of several months to more than a year. These provisions are consistent with rail flows responding to lagged rather than contemporaneous prices, and can also rationalize why crude flows to multiple destinations simultaneously (if there has been recent variation in which destination is "best" in the recent past). Another potential explanation for multi-destination flow is that shippers to each rail destination face an upward sloping supply curve that is specific to that destination. In equilibrium, market forces would then spread out shipments across destinations so that the net returns to shipping to each destination were equalized. A variety of forces could give rise to such destination-specific increasing costs: congestion along rail routes, congestion at unloading facilities, and local scarcity in the rail car leasing market. Destination-specific downward sloping demand curves for Bakken crude provide a related, but distinct explanation. Refiners in destination markets, particularly those markets with limited demand for light oil (like the Gulf Coast), might only be able to process so much light oil before their refining infrastructure was operating inefficiently.

Unfortunately, we lack the data to directly test or estimate models involving contracts or diminishing returns to specific destinations. Private shipping contracts are not publicly disclosed, though the STB data do indicate that, on average, contracted freight services have lower rates than do spot freight shipments. Moreover, as described in section 3, the destination-specific cost data that we do have is either incomplete (i.e., the STB data is just for freight services) or updated at such a low frequency that they are unlikely to be useful in estimation (Genscape).

Instead, we focus on measuring overall departures from modeling assumptions (a) and (b) by estimating the correlation of the destination shares of Bakken crude with the contemporaneous and lagged oil prices at those destinations. To compute the destination shares, we combine data from the North Dakota Pipeline Authority on the monthly share of each transportation mode (local consumption, crude by rail, pipeline, and truck) for Bakken crude oil production and data from the EIA on the monthly share of each rail destination for crude-by-rail originating in PADD 2. Because our data end in 2016, well in advance of the completion of DAPL, we assume that the pipeline share corresponds to pipeline shipments that reach Cushing, OK and not the Gulf Coast.[22] We divide the total rail share from the NDPA data into destination-specific crude-by-rail shares using the EIA data. The combined data give the share of North Dakota crude oil production that is refined locally, transported by truck to Canada, transported by pipeline to Cushing, OK, and transported by rail to each of PADDs 1, 2, 3, and 5.

Because refined consumption is empirically at or just below the reported capacity for refineries in North Dakota during the entire time period, we subtract it from total production and focus on the destinations that appear to be unconstrained. We assume shippers to PADD 1 earn the Brent price for their cargoes, shippers to PADD 2 earn WTI, shippers to PADD 3 earn LLS, and shippers to PADD 5 earn ANS. There is no publicly available light oil benchmark for any central Canadian trading hub, so we assume that truck shipments earn the local Clearbrook price.

To measure the correlation of destination shares with destination prices, we estimate a multinomial logit choice model. Formally, let u_{ijt} be the indirect utility that an atomistic shipper i experiences when shipping to destination j during month t. We assume that u_{ij} is a linear combination of current and lagged prices at destination j, a fixed effect for that destination, a time-varying unobserved mean utility shock specific to j and t, and an iid type-1 extreme value "taste" shock specific to i, j and t:

$$u_{ij} = \sum_{l=0}^{L} \beta_{t-l} p_{j,t-l} + \delta_j + \xi_{jt} + \epsilon_{ijt}$$

Under the assumption that shippers choose the destination with the highest indirect utility, the fraction of crude oil production that is shipped to destination j during period t is given

[22] There were at least two distinct pipeline routes to Cushing: the Enbridge Mainline and Spearhead systems, which travel east into Minnesota and Illinois and then southwest into Cushing, and the Butte pipeline system, which connects to the Guernsey, Wyoming trading hub, which in turn is connected to the Platte, Pony Express and White Cliffs pipeline systems that connect to Cushing. We are not aware of any pipeline routes to other major pricing centers.

by the standard logit formula:

$$s_{jt} = \frac{\exp\left(\sum_{l=0}^{L} \beta_{t-l} p_{j,t-l} + \delta_j + \xi_{jt}\right)}{\sum_k \exp\left(\sum_{l=0}^{L} \beta_{t-l} p_{k,t-l} + \delta_k + \xi_{kt}\right)}$$

To estimate this model, we treat pipeline transportation as the "outside good" and use the Berry (1994) logit inversion formula to correlate the log odds ratios of the destination shares with the destination-specific prices:

$$\log s_{jt} - \log s_{0,t} = \sum_{l=0}^{L} \beta_{t-l} \left(p_{j,t-l} - p_{0,t-l}\right) + \delta_j - \delta_0 + \xi_{jt} - \xi_{0t}$$

We estimate the model using OLS, noting that our interpretation of the β's is complicated by standard supply/demand endogeneity concerns.

Table 1 presents estimates of the above model using contemporaneous destination prices and price lags of order 3 to 24 months. The estimated coefficient on contemporaneous prices in column 1 is roughly 0, and contemporaneous prices explain practically none of the variation in destination shares, providing initial evidence that rail shipment decisions are not especially sensitive to spot pricing conditions. The remaining columns successively add lags of destination-specific crude oil prices, and in the process provide more explanatory power.

In general, the coefficient on the oldest price realization has the most positive value and is more precisely estimated than any other coefficient in a given specification. This pattern holds for specifications including lags of up to 18 months (column 7) and is consistent with a model in which a large fraction of shippers sign contracts that require or otherwise provide incentives for consistent shipments over several months. For example, if all shippers signed T-month contracts, then the price vector driving month t's destination shares would be p_{t-T}. Among the specifications considered in table 1, the model with lags up to 18 months fits the data the best, both in the raw and adjusted R^2 sense, therefore suggesting that contracts with terms as long as 18 months are common in crude-by-rail shipping.

Another pattern in Table 1 is the negative and precisely estimated coefficients on contemporaneous prices. It is hard to tell a rational story in which shippers actively avoid destinations that currently have higher prices, all else equal. However, it is not hard to blame this result on the endogeneity problem standard in all supply/demand models like this one. If ξ_{jt} constitutes a supply shock (e.g., unobserved opening of a new loading or unloading facility), contemporaneous prices at destination j will be *negatively* correlated with the shock, so the coefficient on contemporaneous prices will be biased *downwards*.

Though we make no claims that this empirical model identifies "true" shipper preferences,

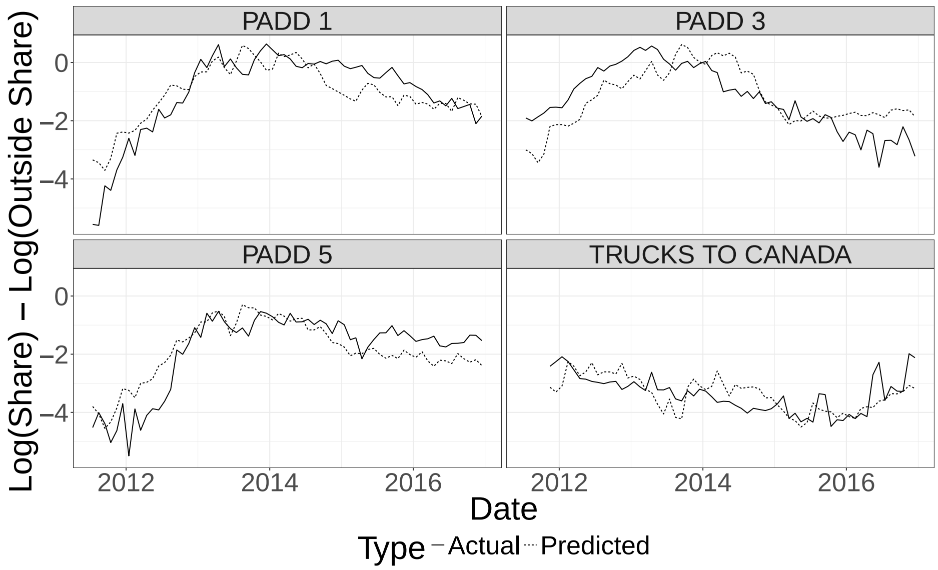
Table 1: Multinomial Logit Shipment Destination Share Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
P_t	-0.00	-0.07^{***}	-0.08^{***}	-0.07^{***}	-0.09^{***}	-0.08^{***}	-0.06^{***}	-0.03^{**}	-0.01
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
P_{t-3}		0.09^{***}	0.02	0.00	0.02	-0.00	-0.01	-0.01	-0.01
		(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
P_{t-6}			0.09^{***}	0.02	0.01	0.03	0.01	-0.00	-0.01
			(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
P_{t-9}				0.09^{***}	0.01	0.00	0.02	0.01	0.00
				(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
P_{t-12}					0.10^{***}	0.03^{**}	0.02^{*}	0.03^{***}	0.02^{**}
					(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
P_{t-15}						0.09^{***}	0.03^{***}	0.03^{**}	0.03^{**}
						(0.01)	(0.01)	(0.01)	(0.01)
P_{t-18}							0.08^{***}	0.05^{***}	0.04^{***}
							(0.01)	(0.01)	(0.01)
P_{t-21}								0.04^{***}	0.04^{***}
								(0.01)	(0.01)
P_{t-24}									0.02^{**}
									(0.01)
R^2	0.00	0.07	0.14	0.23	0.34	0.41	0.46	0.43	0.37
Adj. R^2	-0.01	0.06	0.13	0.21	0.33	0.39	0.44	0.40	0.34
N	400	397	386	371	356	341	326	311	296

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Multinomial logit regressions of shipment destination shares onto current and lagged destination prices. PADD 2 pipeline shipments are the outside good. Shipments by truck to Canada are priced using Bakken Clearbrook, whose price series starts in May, 2010. Thus, in specifications with longer lags of price, there are fewer observations for shipments by truck to Canada than for other destinations. All specifications include destination fixed effects. Newey-West standard errors in parentheses. R^2 values are calculated "within" each destination.

Figure 8: Actual vs Predicted Rail Shares Under 18 Month Lagged Model



Note: PADD 2 pipeline transportation is the outside option. Because predicted PADD 2 rail shipments are constant at the estimated fixed effect, their actual and predicted values are excluded from this figure.

it is the case that it matches data reasonably well. Figure 8 plots actual and fitted values of the log odds ratios for each destination over the time period covered by our data. There two patterns to match: rail shipments increase to parity with pipeline shipments by early 2014 and then decline, and truck shipments gradually diminish until they begin growing again in 2016. The empirical model is able to match both of these patterns, bolstering the fundamental concept underpinning our model that rail flows do respond to oil price shocks, even if those responses occur with a non-trivial lag.

4.1 Implications for interpretation of the model

The divergence between the pipeline flow data and our model primarily derives from the fact that most real-world crude-by-rail flows are tied to contracts, but the model assumes that crude-by-rail operates on spot markets. Our model therefore omits mechanisms by which contracts add value to shippers: contracts with rail providers result in lower transportation fees (as revealed in the STB data in section 3.3), and contracts with downstream refiners provide refiners with medium-run certainty over their crude mix. Because our model does not account for this value and because the oil price and Genscape transportation cost data that we use to calibrate our model reflect spot oil markets and spot shipment costs rather than contract prices, our model is biased toward under-valuing crude-by-rail.[23]

In principle, we could augment our pipeline investment model by allowing rail shippers to sign one or two year contracts with transportation providers. However, we believe that this approach is impractical for at least two reasons. First, and most importantly, shippers' contracts are private and not observable. We therefore lack a basis for quantifying the pricing and transportation cost benefits that these contracts confer. Second, allowing for these contracts would substantially complicate our model, since we would need to track state variables for contracted rail volumes over the life of the pipeline. Rather than take this step, we find it preferable to use our model as described in section 2, even though it abstracts away from crude-by-rail contracts, knowing that this approach causes us to under-value the option value generated by railroad transportation.

5 Model calibration and validation

To quantify the economic impact of crude-by-rail's flexibility on equilibrium pipeline investment, we calibrate the model defined in Section 2 to the recently constructed Dakota Access Pipeline (DAPL). Using reported construction costs and various assumptions about the long run elasticity of local supply, estimates of the cost structure of crude-by-rail transportation, and the likely distribution of future downstream crude oil prices, we compute the average-cost tariff that would finance the pipeline's construction and then estimate the sensitivity of the pipeline's size to the cost of railroad transportation. Because the DAPL investment decision, like all major pipeline investments, involves a long-term commitment, each of these inputs must be quantified with the long run in mind.

5.1 Calibration details

5.1.1 Dakota Access Pipeline facts

We calibrate our model to fit market conditions in June, 2014, when DAPL received firm commitments from its eventual customers (Energy Transfer Partners, 2014). At this time, the Brent crude oil price was \$111.87/bbl and expected to remain high: the three-year Brent futures price was \$99.19/bbl.[24] Completion of DAPL, with a capacity of 470,000

[23] By ignoring contacting, we believe it is likely that we are also under-valuing the effect of rail prices on equilibrium pipeline investment. Under the "simplified" model described in equation 3, the equilibrium pipeline size is concave in rail costs. If the effect of contracts on the value of crude-by-rail is equivalent to a decrease in rail costs, the concavity of the relationship between pipeline investment and crude-by-rail costs implies that $\frac{dK}{dr_0}$ will be larger when crude-by-rail moves primarily on contracts.

[24] We use futures price data from Quandl, downloaded from https://www.quandl.com/collections/futures/ice-brent-crude-oil-futures. Contracts were not actively traded at horizons beyond three years.

bpd, was expected to bring the sum of Bakken pipeline export capacity and local refining capacity to 1.323 million bpd (Biracree (2016), Arno (2016)).[25] This combined capacity falls modestly short of the North Dakota Pipeline Authority's contemporaneous forecast that Bakken production would reach 1.5 million bpd within a few years at expected crude oil prices (North Dakota Pipeline Authority, 2014). The pipeline reportedly cost \$3.8 billion to construct.

5.1.2 Downstream oil price distribution for pipeline shippers

To calculate the expected return to pipeline shippers, our model requires an estimate of the distribution $f(P_d)$ of future downstream prices that these shippers face in the long run. The definition of "long-run" that we adopt is a 10 year time horizon, thereby aligning our estimated price distribution with the 10-year commitments made by DAPL shippers. Although DAPL sends oil to the U.S. Gulf Coast, where the relevant price is LLS, we use the three-year Brent (East Coast) futures price of \$99.19/bbl to measure the expected price $E[P_d]$ faced by DAPL shippers, since there is no LLS futures market and since the Brent and LLS prices have historically been quite close (figure 4).[26]

To obtain the variance of $f(P_d)$, we calculate the historic long-run volatility of the Brent crude spot price (we continue to use Brent rather than LLS to be consistent with our use of futures prices for $E[P_d]$ and because Brent is historically the most liquidly traded waterborne crude price). To do so, we estimate the standard deviation of 10-year differences in logged monthly prices. As of June, 2014, this standard deviation over the entire price history was 0.759.[27] We assume that the distribution of future prices faced by pipeline shippers is lognormal, with this standard deviation, so that a 95% confidence interval for prices 10 years in the future covers a decrease of 77% to an increase of 343% relative to the expected price of \$99.19/bbl.[28]

[25] See especially Biracree (2016) for a breakout of Bakken-area local refining capacity and export pipeline capacity, though note that the capacity of DAPL is 470,000 bpd rather than 450,000 bpd as stated in the article.

[26] We use three-year futures price to measure the long-run expected price 10 years in the future, rather than extrapolate or forecast long-run price changes, because the literature on long-run oil-price forecasting suggests that the martingale assumption typically leads to the smallest forecast errors (Alquist, Kilian and Vigfusson, 2013).

[27] The full Brent price history on Bloomberg goes back to May, 1983, so that the first observation for which 10-year differences exist is May, 1993.

[28] The 95% confidence interval for future price changes is obtained by $e^{0.759*1.96}$ and its reciprocal.

5.1.3 Downstream oil price distribution for rail shippers

We use the history of daily prices for Brent, LLS, WTI and ANS to compute $\widetilde{P}$ (the maximum of the downstream prices accessible by rail) and its empirical conditional density $f(\widetilde{P} \mid P_d)$, now treating P_d as the LLS price. Recall that our model assumes that committed pipeline shippers would never prefer to use rail instead of the pipeline. In our 20 years of spot price data, this assumption holds for all but 8 days, so our conditional density estimates do not directly impose it.[29] Though the model assumes that shipments to all rail destinations face a common cost r_0, the data do show some differences on the order of a few dollars per barrel. To incorporate these differences in shipping costs, we revise our definition of what rail shippers earn to $\widetilde{P} = \max_i P_i - r_{0,i}$. Thus, $\widetilde{P}$ is the best *netback* (as opposed to best *downstream price*) a rail shipper could receive. Then, we define the pricing difference $D = \widetilde{P} - (P_d - r_{0,d})$, which is the amount by which the best rail netback exceeds a rail netback to LLS. By construction, D will have a point mass at 0. To deal with the statistical issues associated with point masses in density estimation, we non-parametrically estimate the conditional distribution of D in two steps. First, we estimate $Z(P_d) = \Pr(D = 0 \mid P_d)$, the probability that P_d is the best downstream rail destination. Second, we estimate $f(D \mid P_d, D > 0)$. Subsequently, in our simulations we compute realizations of $\widetilde{P}$, conditional on P_d, by taking binomial draws with probability $Z(P_d)$. If a draw equals zero, we assign rail shipments to LLS and pay them the LLS netback $P_d - r_{0,LLS}$. If a draw equals one, we take a draw of D from $f(D \mid P_d, D > 0)$ and pay rail shipments that draw plus the LLS netback.

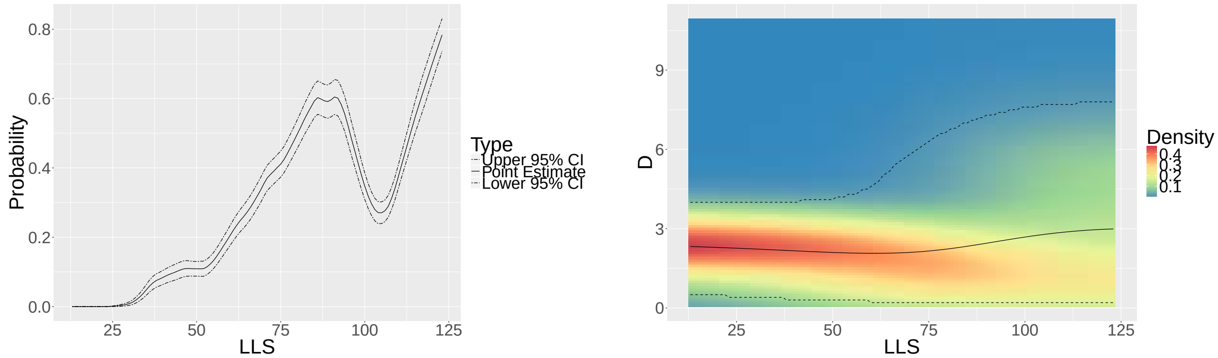
Figure 9 shows the estimated conditional distribution of D given P_d. Panel (a) shows that the probability that P_d is the best rail destination is generally increasing in P_d. However, there is a clear departure from monotonicity when LLS prices are between \$95-105 per barrel, which occurred frequently during 2011-2014, and rarely before. These price levels occurred at the same time as a significant departure in the historical Brent LLS relationship. For the 20 or so years prior to 2011, Brent had traded at discount to LLS of a few dollars per barrel. Starting in 2011, this discount frequently became positive, sometimes by as much as \$10 per barrel.[30] Thus, despite high LLS price levels, prices on the East coast were even higher, and by an amount that exceeded the difference in shipping costs. This pattern largely stopped With the relaxation of the U.S. crude oil export ban in January 2015. Panel (b) of 9 shows the density of D, conditional on $D > 0$ and P_d. For most values of P_d, this density is concentrated around just a few dollars per barrel, with a downward trend in the mode. However, at higher levels of P_d, the average D begins to shift upwards, consistent with the data in panel (a).

[29]We exclude those 8 days from the estimation procedure.

[30]See, for example Blas (2013) and Hunsucker (2013)

Figure 9: Conditional Distribution of Downstream Prices

(a) Probability that $D = 0$, Conditional on LLS **(b)** Density of D Conditional on LLS, $D > 0$



Source: Bloomberg, Platts, Genscape. D is the difference between the highest rail netback and the LLS rail netback. Solid lines indicate conditional means. In panel (a), dashed lines indicate 95% confidence intervals. In panel (b), dashed lines indicate the 2.5% and 97.5% quantiles of the estimated conditional density.

5.1.4 Upstream oil supply elasticity

Rather than estimate the supply function for Bakken crude ourselves, we rely on the recent literature that has endeavored to estimate supply curves for U.S. shale oil and gas. Smith and Lee (2017) estimate a long-run supply curve for the Bakken that has a price elasticity which declines from 0.5 to 0.2 as the oil price increases from \$20/bbl to \$100/bbl. They obtain this supply curve by modeling heterogeneity across the Bakken in the "breakeven" price necessary to trigger drilling and therefore oil production. Hausman and Kellogg (2015) estimate price elasticities for shale gas of about 0.8, using instrumental variables regressions of gas drilling rates on lagged gas prices. Newell et al. (2016) use similar methods and estimate a shale gas elasticity of about 0.7. The same authors apply their approach to shale oil in the United States Newell and Prest (2017), estimating an elasticity (cumulatively over 3 quarters of lagged prices) of 1.3.

One interpretation of these very divergent estimates is that the concept of a supply elasticity for an exhaustible resource is inherently strained. Smith & Lee's (2016) elasticity does implicitly account for exhaustion, but as a result the elasticity concerns the price-responsiveness of total ultimate production (in bbls) rather than the responsiveness of produced flow (bbls per day). The other papers find large medium-run responses of drilling—which ultimately results in oil flow—to prices, but do not account for depletion or resource heterogeneity. In particular, if high prices induce the drilling of lower-quality reserves, the production elasticity will be less than the drilling elasticity. Overall, given the uncertainty over the price-responsiveness of upstream supply, we calibrate our model with a range of elasticities spanning 0.2 to 0.7.[31]

5.1.5 Railroad transporation costs

We use the Genscape rail cost data to calibrate values of r_0 and r_1. The Genscape data is constructed from telephone surveys of market participants who are asked at what price they believe they could conduct spot market transactions. As such, these prices are not true marginal costs for rail transportation, and in Section 3.3 we discuss a variety of issues with it. Nevertheless, this data can be reported at a disaggregated level and includes relevant costs like loading, unloading and tanker car rental, unlike the STB data. Acknowledging the limitations of the source data, we estimate $r_{0,i}$ as the minimum all-in rail cost in the Genscape data for transportation between the Bakken and destination i. For most destinations, this minimum price occurs in the spring of 2016, after (a) millions of bbls/day of loading and unloading capacity had been constructed, (b) rail car lease prices had fallen as a glut of new tanker cars became available and (c) global oil prices reached the lowest level in more than a decade. We select the minimum reported cost because these costs coincide with both fully constructed capacity and limited use of that capacity, which we view as a reasonable analogue to the model's notion of r_0.[32]

Lacking a credible means to estimate the full supply curve for rail services, we calibrate the model using various values of r_1. At the low end, we assume $r_1 = 0$, which would be consistent with the fact that rail loading/unloading capacity now far exceeds even optimistic estimates of the peak level of Bakken production, and that tanker cars should be viewed as commodities in the medium to long term. At the middle and higher ends, we assume $r_1 = 3$ and $r_1 = 6$, to match the fact that Genscape's highest rail costs of $20-25 per bbl were realized in October, 2013, when aggregate crude-by-rail activity was about 300 thousand bbls/day higher than when prices were lowest.

5.1.6 Pipeline economies of scale

Finally, we consider the extent to which there are increasing returns to scale in pipeline construction; i.e., we calibrate the $d(C(K)/K)/dK$ term in equation (7). As a conservative baseline, we simply assume that the construction of DAPL has constant returns to scale,

[31] We have also experimented with a supply elasticity of 1.3; however, this large elasticity causes the denominator of equation (7) to be negative, which taken seriously implies that the observed DAPL size is an unstable equilibrium (and that the stable equilibrium involves a substantially larger pipeline) Under constant returns to scale, our calculated values for dK/dr_0 are 40% to 90% larger than those shown in table 2 for an elasticity of 0.7.

[32] These costs are: $13.00/bbl for PADD 1, $8.54 for PADD 2, $10.94 for PADD 3, and $9.23 for PADD 5.

so that this derivative is zero. As an alternative, we assume that the pipeline's cost is a constant elasticity function of capacity, and we obtain an elasticity estimate from Soligo and Jaffe (1998)'s study of Caspian Basin oil export pipelines.[33] The elasticity implied by Soligo and Jaffe (1998) is 0.59, consistent with substantial scale economies.[34] Other work on U.S. natural gas pipelines suggests an even lower elasticity: Rui, Metz, Reynolds, Chen and Zhou (2011) obtains a sample of U.S. natural gas pipeline costs and regresses log(cost) on the log of pipeline diameter (along with controls for length and geography). That paper obtains an elasticity of pipeline cost with respect to diameter of 0.49, but because pipeline capacity is convex in diameter (the cross-sectional area of a pipe increases with diameter squared), the elasticity with respect to capacity is even lower. To be conservative, we elect to use the elasticity of 0.59 from Soligo and Jaffe (1998).

5.2 Validation

Given the inputs discussed above, we validate our model by solving for the DAPL average cost that is consistent with an equilibrium in which shippers committed to the actual DAPL capacity of 470 thousand bpd (which increased total Bakken pipeline export and local processing capacity to 1.325 million bpd). We then compare this implied cost—which per equation (6) is equal to the model's expected return to pipeline shippers—to the actual DAPL tariff for long-term shippers of \$5.50/bbl (Gordon, 2017).

Our model's implied costs are presented in column (3) of table 2 (note that the implied cost is not affected by assumptions about pipeline returns to scale). Our results naturally vary with the assumed values for the supply elasticity and for r_1, but are generally within 20% of the actual \$5.50/bbl tariff.

The implied cost increases with r_1 because higher values of r_1 decrease crude-by-rail volumes, thereby increasing the expected margins earned by pipeline shippers. It tends to decrease with the Bakken supply elasticity because, given oil price expectations in June, 2014, total Bakken production was expected to exceed total pipeline export and local refining capacity. The risk that oil prices might fall sufficiently far as to decongest the pipeline, thereby leaving pipeline shippers with zero return, is increasing in the supply elasticity.[35] Thus, the expected return, and therefore the implied cost of DAPK, decreases with the supply elasticity.[36]

[33]We were not able to find a study of scale economies in U.S. oil pipelines.

[34]We obtain this elasticity using the engineering cost estimates presented in table 2 of Soligo and Jaffe (1998). We regress log(cost) on log(capacity) and route fixed effects.

[35]For a larger supply elasticity, the oil price does not have to decrease as far below the expected price in order to decongest the line.

[36]For the case in which r_1 is assumed to be large (\$6/bbl per mmbpd), the implied cost actually increases

6 Results: the sensitivity of pipeline investment to rail costs

Column (4) of table 2 presents our calculations, following equation (7), of the counterfactual increase DAPL capacity in a world where crude-by-rail were \$1/bbl more expensive: dK/dr_0. As a baseline, the upper half of the table assumes that pipeline construction has constant returns to scale. Over supply elasticities ranging from 0.2 to 0.7, and over values of r_1 ranging from zero to \$6/bbl per million bpd, our estimates of dK/dr_0 range from 29.4 thousand bpd (mbpd) per \$/bbl to 73.7 mbpd per \$/bbl, relative to the actual DAPL capacity of 470 mbpd (and total Bakken local refining and pipeline export capacity of 1.323 million bpd). Thus, a 9.1% increase in rail transportation costs increases total investment in Bakken local refining and pipeline export capacity by between 2.2% and 5.6%, so that the elasticity of pipeline capacity to rail costs is between 0.24 and 0.61. Given air pollution externalities from rail transportation that exceed \$2/bbl (Clay et al. (2017)), these estimates imply substantial long-run substitution to pipeline capacity in the event that these externalities are priced or otherwise addressed through regulation.

The value of dK/dr_0 depends on assumed parameters in intuitive ways. Large values of the upstream supply elasticity increase the responsiveness of pipeline capacity to the rail transportation cost intercept r_0. And large values for r_1, which depress the value of rail transportation for high downstream price realizations, decreases the sensitivity of pipeline capacity to r_0.

In the bottom half of the table that allows for increasing returns to scale, we show that capacity increases are amplified, particularly in cases involving a large supply elasticity and a small value for r_1. With an elasticity of 0.7 and $r_1 = 0$, we actually obtain a value of dK/dr_0 that is nearly 1 million bpd per \$/bbl. This set of assumptions, combined with increasing returns to scale, is quite favorable to both the overall value of rail transportation (as evidenced by the quite low implied DAPL cost of \$4.15/bbl) and to the sensitivity of pipeline capacity to rail, so this result is more illustrative than real-world predictive.

Table 2 also presents, in column (5) the effect on pipeline capacity of removing crude-by-rail's spatial option value. For each parameter vector, we calculate column (5) by solving for the capacity such that the expected return to pipeline shippers, when rail can only access the Gulf Coast (PADD 3), equals the implied per-bbl cost from column (3). In most cases, removing spatial option value is equivalent to increasing the cost of crude-by-rail by modestly

when the elasticity changes from 0.5 to 0.7. This reversal occurs because at high prices, the large increase in supply for the high elasticity case causes a large increase in rail transportation costs, which increases the returns to pipeline shippers.

Table 2: Model validation and results

Input parameters		Implied		ΔK if no
Supply elasticity	r_1 $/bbl per mmbpd	avg. cost $/bbl	dK/dr_0 mbpd per $/bbl	spatial option value (mbpd)
Pipeline costs assumed to have constant returns to scale				
0.2	0	$6.28	36.28	31.06
0.35	0	$5.03	44.67	38.05
0.5	0	$4.51	56.01	48.88
0.7	0	$4.15	73.70	64.35
0.2	3	$6.67	32.46	28.27
0.35	3	$5.53	38.92	33.48
0.5	3	$5.13	47.18	41.42
0.7	3	$5.01	59.00	51.79
0.2	6	$7.05	29.37	25.92
0.35	6	$6.01	34.49	29.85
0.5	6	$5.74	40.75	35.87
0.7	6	$5.82	49.19	43.23
Pipeline costs assumed to have increasing returns to scale				
0.2	0	$6.28	51.35	42.40
0.35	0	$5.03	82.35	59.65
0.5	0	$4.51	157.09	90.55
0.7	0	$4.15	889.25	153.64
0.2	3	$6.67	44.16	38.44
0.35	3	$5.53	65.34	52.27
0.5	3	$5.13	105.46	76.90
0.7	3	$5.01	241.03	127.81
0.2	6	$7.05	38.74	35.10
0.35	6	$6.01	54.15	46.44
0.5	6	$5.74	79.37	66.70
0.7	6	$5.82	139.40	109.46

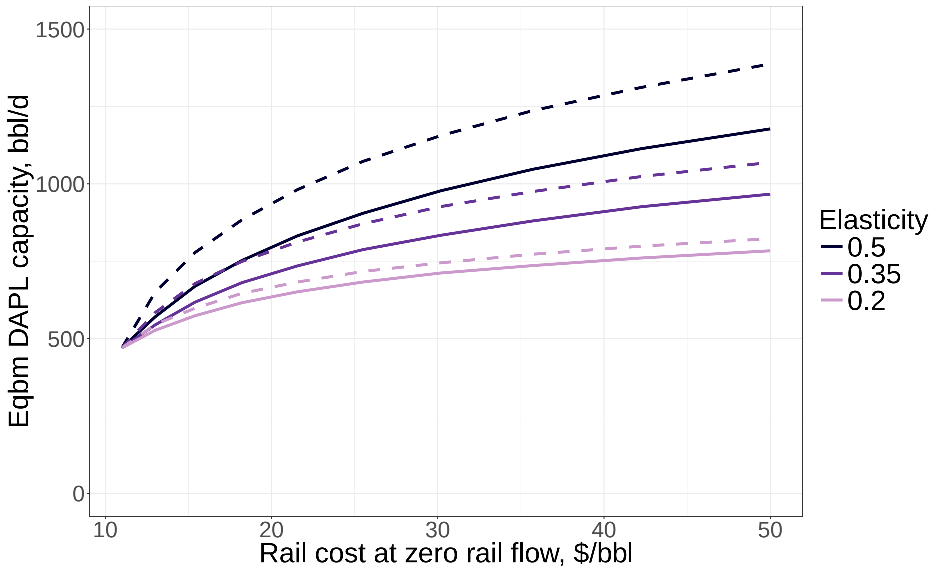
Note: Increasing returns to scale specifications assume pipeline cost has a constant elasticity of 0.59 with respect to capacity. All rows assume $r_0 = \$11$/bbl. mbpd denotes thousands of barrels per day, and mmbpd denotes millions of barrels per day. See text for details.

less than $1/bbl.[37] This outcome is consistent with the fact that price differences between coastal destinations have historically diverged, on average, by less than $1/bbl when prices have been sufficiently high to induce rail transport (figure 4).

Finally, we explore counterfactuals in which rail becomes prohibitively expensive, even

[37] In the cases involving increasing returns to scale, a large supply elasticity, and low values of r_1, removing spatial option value has a much smaller impact on pipeline capacity than does increasing r_0 by $1/bbl. This result comes from the intuition that the response of pipeline capacity to decreases in the returns to rail is more concave under increasing returns to scale (see figure 10). Because the benefits of spatial option value to rail shippers are stochastic, while the benefits of low costs are deterministic, this concavity implies that the relative impact of rail cost reductions is greater under increasing returns to scale.

Figure 10: Counterfactual DAPL capacity vs rail cost intercept r_0.
Solid line = constant returns to scale; dashed line = increasing returns to scale



Note: "Elasticity" refers to upstream crude oil supply elasticity. Increasing returns to scale specification assumes pipeline cost has a constant elasticity of 0.59 with respect to capacity. For all counterfactuals, r_1 = \$0/bbl per million bbl/d. See text for details.

for large price spreads.[38] Figure 10 presents counterfactual capacities for DAPL as r_0 is increased from its baseline value of \$11/bbl to an extremely high value of \$50/bbl.[39] For each assumed supply elasticity, the derivative dK/dr_0 decreases with r_0, a consequence of the fact that as capacity K increases, larger and larger downstream prices are required to congest the pipeline, but the probability of these price realizations becomes lower and lower. Nonetheless, effectively removing rail from the option set has substantial effects on pipeline investment. Even when the supply elasticity is just 0.2 and there are no scale economies in pipeline construction, increasing the rail cost intercept to $r_0 = \$50$/bbl causes DAPL's capacity to increase from 470 mbpd to more than 750 mbpd. Allowing for economies of scale results in even larger increases in pipeline capacity, especially in cases involving relatively high oil supply elasticities.

[38]These counterfactuals implicitly assume that, as rail becomes prohibitively expensive, other alternative transportation modes such as trucks also become prohibitively expensive.

[39]We obtain figure 10 by calculating, for each value of r_0, the capacity that would cause DAPL shippers to expect the same \$/bbl return that they would have expected at the baseline value of $r_0 = \$11$/bbl.

7 Conclusions

The development of the Bakken shale was associated with an unprecedented boom in crude-by-rail transportation. Since the late-2014 fall in oil prices, however, crude-by-rail volumes have fallen substantially. One interpretation of these recent shifts is that crude-by-rail was merely a transitory phenomenon, and that pipelines will henceforth convey nearly all overland crude oil flows. This paper emphasizes an alternative view of these events. We see the rise and fall of crude-by-rail volumes as underscoring the option value provided by rail transportation: rail enables shippers to vary shipment volumes and destinations in response to crude oil price shocks. This flexibility contrasts with pipelines that require long-term, binding ship-or-pay contracts in order to underwrite their large up-front costs. The model of pipeline investment versus railroad shipping that we develop in this paper illuminates why, even though rail volumes may ebb and flow over time, the existence of the rail option can durably reduce the incentive to invest in pipeline capacity. Morevoer, calibration of our model to recent oil market data suggests that this effect is economically significant.

Even under conservative assumptions, our model implies that policies that increase crude-by-rail's cost by addressing its environmental and safety externalities will have substantial implications for pipeline investment. Clay et al. (2017) finds that railroad air pollution externalities alone exceed \$2 per barrel shipped. Our results imply that, had policies caused rail transporters to internalize a \$2/bbl externality at the time of DAPL's investment decision, DAPL's capacity would have been at least 59,000 bpd larger than its actual 470,000 bpd capacity. Under more aggressive but plausible assumptions on input parameters, the impact could have been greater than 100,000 bpd.

Finally, we believe that our model is the first to illustrate how the presence of a costly but flexible transportation option—crude-by-rail—adversely affects investment in transportation infrastructure that requires large up-front commitments but has a relatively low amortized cost. The intuition and basic structure of our model readily apply to other settings involving tradeoffs between technologies that differ in the extent to which their costs are sunk versus variable. For instance, urban transportation planners must often choose whether to invest in dedicated light rail lines, which have large sunk costs that can translate to low per-passenger costs given sufficient ridership, or flexible bus networks. As another example, electricity is generated by both "baseload" plants (such as nuclear plants that have nearly zero marginal cost) and "peaker" plants that have low sunk but high marginal costs and can help serve stochastic electricity loads (Borenstein (2005)). Our model provides a conceptual framework that can be used to evaluate and intuitively understand how tradeoffs between sunk and flexible investments in these settings are affected by factors such as relative costs, scale economies, and demand uncertainty.

References

Alquist, Ron, Lutz Kilian, and Robert J Vigfusson, "Forecasting the price of oil," *Handbook of economic forecasting*, 2013, *2*, 427–507.

American Association of Railroads, "Freight Rail Traffic Data," 2017. [Online; `https://www.aar.org/Pages/Freight-Rail-Traffic-Data.aspx`; accessed 11 May, 2017].

Arno, David, "Tank-Car Lease Rates Plummet on Weak Crude-By-Rail Demand, Low Crude Prices," *Genscape Blog*, 2015. [Online; `http://www.genscape.com/blog/tank-car-lease-rates-plummet-weak-crude-rail-demand-low-crude-prices`; accessed 13 September, 2017].

_ , "North Dakota Crude-by-rail Players Adjust Strategies as Dakota Access Crude Pipe Work Continues," Technical Report 2016.

Berry, Steven T, "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 1994, pp. 242–262.

Biracree, Tom, "Bakken Pipeline Takeaway Capacity: Dakota Access to Close the Gap," *Oil and Gas Financial Journal*, 2016.

Blas, Javier, "US oil glut shifts to Gulf of Mexico," *The Financial Times*, 2013.

Borenstein, Severin, "The long-run efficiency of real-time electricity pricing," *The Energy Journal*, 2005, pp. 93–116.

Busse, Meghan R and Nathaniel O Keohane, "Market effects of environmental regulation: coal, railroads, and the 1990 Clean Air Act," *The RAND Journal of Economics*, 2007, *38* (4), 1159–1179.

Clay, Karen, Akshaya Jha, Nicholas Muller, and Randall Walsh, "The External Costs Of Transporting Petroleum Products By Pipelines and Rail: Evidence From Shipments Of Crude Oil From North Dakota," Working Paper, National Bureau of Economic Research 2017.

Energy Information Administration, "U.S. Energy-by-Rail Data Methodology," 2017. [Online; `https://www.eia.gov/petroleum/transportation/`; accessed 12 April, 2017].

Energy Transfer Partners, "Energy Transfer announces crude oil pipeline project connecting Bakken supplies to Patoka, Illinois and to Gulf Coast markets," 2014. [Online; `http://ir.energytransfer.com/phoenix.zhtml?c=106094&p=irol-newsArticle&ID=1942689`; accessed 13 September, 2017.

Federal Register, "Control of Emissions of Air Pollution From Locomotive Engines and Marine Compression-Ignition Engines Less Than 30 Liters per Cylinder; Republication," 2008.

Gordon, Meghan, "Dakota Access, ETCO Oil Pipelines to Start Interstate Servce May 14," *Platts News*, 2017.

Granitz, Elizabeth and Benjamin Klein, "Monopolization by "Raising Rivals' Costs": The Standard Oil Case," *The Journal of Law and Economics*, 1996, *39* (1), 1–47.

Hausman, Catherine and Ryan Kellogg, "Welfare and Distributional Implications of Shale Gas," *Brookings Papers on Economic Activity*, 2015, pp. 71–125.

Hughes, Jonathan E, "The higher price of cleaner fuels: Market power in the rail transport of fuel ethanol," *Journal of Environmental Economics and Management*, 2011, *62* (2), 123–139.

Hunsucker, Bridget, "LLS discount to Brent could be permanent on light sweet crude invasion," *Platts Oilgram News*, 2013.

_ , "Crude-By-Rail Evolves Into Permanent Midstream Fixture," *Genscape White Paper*, 2015.

Lefebvre, Ben, "Trains Leave Kinder Morgan Pipeline in Lurch," *The Wall Street Journal*, 2013.

Newell, Richard G. and Brian C. Prest, "Is the US the New Swing Producer?: The Price- Responsiveness of Tight Oil," Technical Report 2017/09/08/ 2017.

_ , _ , and Ashley Vissing, "Trophy Hunting vs. Manufacturing Energy: The Price-Responsiveness of Shale Gas," Working Paper, National Bureau of Economic Research 2016.

North Dakota Pipeline Authority, "Website Slide Deck, North Dakota Pipeline Authority," 2014. [Online; `https://ndpipelines.files.wordpress.com/2012/04/ndpa-slides-april-1-20141.pdf`; accessed 7 May, 2017].

— , "US Williston Basin Crude Oil Export Options - June 1, 2017," 2017. [Online; `https://northdakotapipelines.com/oil-transportation-table/`; accessed 24 August, 2017].

Rui, Zhenhua, Paul A. Metz, Douglas B. Reynolds, Gang Chen, and Xiyu Zhou, "Regression Modesl Estimate Pipeline Construction Costs," *Oil & Gas Journal*, 2011.

Smith, James L and Thomas K Lee, "The price elasticity of US shale oil reserves," *Energy Economics*, 2017, *67*, 121–135.

Soligo, Ronald and Amy Myers Jaffe, "The Economics of Pipeline Routes: the Conundrum of Oil Exports from the Caspian Basin," *Unlocking the Assets: Energy and the Future of Central Asia and the Caucasus*, 1998.

Tita, Bob, "Railcar Shortage in U.S. Pushes Up Lease Rates," *The Wall Street Journal*, 2014.