Crude by Rail, Option Value, and Pipeline Investment

Thomas R. Covert
Ryan Kellogg*

March 19, 2018

Abstract

Because U.S. shale oil reserves are often located far from major oil demand centers, long-distance crude oil transportation has garnered increasing industry and public policy attention. Pipelines have dominated U.S. oil transport for decades, owing to their low average per-barrel cost, but they are inflexible once built. Railroads have emerged as an alternative with relatively high per-barrel costs but also flexibility that allows oil shipments to respond to oil price shocks. We examine the substitutability of these two technologies, motivated in part by the environmental and accident externalities associated with crude-by-rail that drive a wedge between its private and social cost. We develop a model that illustrates how private markets allocate capacity between flexible versus inflexible technologies and then calibrate it to crude oil transportation out of North Dakota. Under conservative assumptions on pipeline economies of scale, we find that the elasticity of pipeline capacity to railroad transportation costs lies between 0.23 and 0.56, depending on parameters such as the upstream oil supply elasticity. Our results imply that crude-by-rail is an economically significant long-run substitute for pipeline transportation and that increases to railroad transportation costs—such as those stemming from environmental or safety regulation—would substantially increase pipeline investment and decrease crude-by-rail flows.

*Covert: Booth School of Business, University of Chicago; and National Bureau of Economic Research; Thomas.Covert@chicagobooth.edu. Kellogg: Harris Public Policy, University of Chicago; and National Bureau of Economic Research; kelloggr@uchicago.edu. We are grateful to the Sloan Foundation for generous financial support, and we thank seminar and conference participants at the 2017 NBER Hydrocarbon Infrastructure and Transportation Conference, 2018 AEA Meeting, 2018 Winter Marketing-Economics Summit, Iowa State University, Montana State University, and UC Berkeley 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.

1 Introduction

From 2010 through 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 last large-scale use of crude-by-rail we are aware of dates back to World War II, prior to the advent of large-diameter, long-distance pipelines (Johnson (1967), p. 327). The resurrection of crude-by-rail is puzzling in light of its relatively high cost per barrel (relative to amortized per-barrel pipeline 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 behind 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.

Why would crude oil shippers elect to use railroads rather than pipelines, given their greater cost?[1] One simple explanation is that the remarkable oil production growth in the Bakken, Niobrara, and other shale formations in the upper Midwest outpaced the speed at which new pipeline capacity could be built, so that producers effectively had no choice. 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. Another explanation—not necessarily exclusive of the "delay" story—is that the flexibility offered by crude-by-rail makes it an attractive transportation option in spite of its higher cost. Because rail infrastructure already exists between the upper Midwest and nearly every major refining center in the country, the cost of shipping crude by rail is largely variable, so that rail shippers can avoid costs by choosing to not ship. Pipeline shippers do not have this ability, since pipeline construction costs are entirely sunk. Moreover, rail allows shippers to decide not just whether but where to move crude oil, so that they can actively respond to changes in upstream and downstream oil 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 paper examines the economic importance of crude-by-rail's flexibility by modeling and quantifying the extent to which crude-by-rail depresses incentives to invest in pipeline capacity. Specifically, we evaluate the substitutability of these two technologies by assessing how an increase in the cost of railroad transportation would affect equilibrium pipeline

[1] Throughout this paper, we follow transportation industry terminology by referring to pipeline and rail customers as shippers. The pipelines and railroads themselves are known as carriers, not shippers.

investment. We ultimately find effects that have economically meaningful magnitudes, consistent with an interpretation that crude-by-rail adds value to shippers beyond being merely a "stopgap" while pipelines are permitted and constructed.[2]

Beyond shedding light on the economics driving one of the most significant developments in the U.S. oil industry in decades, our analysis speaks to policy questions stemming from the environmental and safety externalities of crude-by-rail. Clay, Jha, Muller and Walsh (2017) 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.[3] Overall, Clay et al. (2017) estimates that air pollution damages from crude-by-rail are nearly twice those from pipeline transportation and are also much larger than damages from spills and accidents. Our results inform estimates of the impacts of policies that target these externalities—and thereby raise the cost of rail transport—on investments in pipeline capacity and on the volume of oil flow that will substitute away from rail to pipelines.[4]

To quantify the impact of crude-by-rail on pipeline investment incentives, we develop a model in which crude oil shippers can use either build a pipeline or use 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.[5] Pipeline transportation has large fixed costs with potentially significant

[2]In spite of the analysis presented here, there are several reasons why crude-by-rail did not occur at large scale between World War II and the shale boom. First, U.S. crude oil production declined in essentially all major producing basins from the early 1970s until 2008, so that existing pipeline capacity was sufficient to convey production until the shale boom. (Exceptions include the Alaska North Slope and the deepwater Gulf of Mexico, where rail infrastructure does not exist. See https://www.eia.gov/dnav/pet/pet_crd_crpdn_adc_mbblpd_a.htm and https://www.eia.gov/naturalgas/crudeoilreserves/top100/pdf/top100.pdf for data.) Second, while U.S. oil production steadily grew between WWII and 1970, a number of factors in that era strongly favored pipelines over rail. Policies such as oil import restrictions and state-level control of production levels (especially by the Texas Railroad Commission) actively worked to maintain U.S. oil price stability, reducing incentives to use flexible transportation (see p.68-9 of Cookenboo (1955); pp.377, 427-8, and 475 of Johnson (1967); and p.12-3 of Smiley (1993)). Pipelines were primarily owned by vertically integrated oil majors rather than independent carriers, since federal regulators interpreted common carry regulation as forbidding third-party capacity contracts (see pp.368-70, 462, 471, and 476 of Johnson (1967) and pp.113-4 of Makholm (2012)). Independent shippers did use oil pipelines at posted tariff rates; however, rate regulators during this area are widely believed to have let pipelines earn excess returns (see pp.99 and 111-12 of Cookenboo (1955); pp.407-12, 450, and 473 of Johnson (1967); pp.90-97 of Spavins (1979); and p.116 of Makholm (2012)). These excess returns would in turn generate incentives to invest in excess capacity (Averch and Johnson, 1962).

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

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

[5]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

economies of scale and negligible variable costs. This cost structure is similar to many other "natural monopoly" industries, and as a result the maximum tariff that pipelines can charge to shippers is regulated by the Federal Energy Regulatory Commission 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, which they can decide to do (or not) 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, not just the destination served by the pipeline. A key insight form our model is that the ability to arbitrage crude oil price differences via rail limits the returns that can be earned by pipeline shippers, since spatial price differences become bounded by the cost of railroad transportation. Thus, the availability of the rail option reduces shippers' incentive to commit to pipeline capacity.

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 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.

We use our model to quantify the impact an increase in rail transportation costs would have on pipeline investment, using the Dakota Access Pipeline (DAPL) as a case study. This

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.

pipeline was completed, with a capacity of 520,000 bpd, in June, 2017, and it first received firm commitments from shippers in June, 2014. We calibrate the model to conditions prevailing in June, 2014 and then ask how pipeline capacity commitments would have changed in a counterfactual in which crude-by-rail were more costly.

We calibrate our model using data from a variety of sources and previous work in the literature. We use data on downstream crude prices on the U.S. West, Gulf, and East Coasts to estimate the future distribution of crude prices that shippers faced. We obtain data on crude-by-rail flows from the U.S. Energy Information Administration and show that these flows are responsive to price differentials, albeit with a lag of several months to two years. This lag motivates a specification of our model in which rail movements use short-term contracts rather than flow freely on spot markets. We obtain railroad cost data from the U.S. Surface Transportation Board and from Genscape (a private industry intelligence firm) to estimate railroad transportation costs as a function of volumes shipped. These cost data show that rates charged for rail transportation, rail car leases, and possibly other services (such as terminal fees) co-vary with shipping volumes, consistent with the presence of scarcity rents or market power in these markets.[6] Finally, to obtain an upstream supply curve for Bakken crude oil, we use elasticity estimates from the literature on shale oil and gas (Hausman and Kellogg (2015), Newell, Prest and Vissing (2016), Bjørnland, Nordvik and Rohrer (2017), Newell and Prest (2017), and Smith and Lee (2017)).

Given these inputs, we validate 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. The implied tariffs from our model are quite close to the actual published DAPL tariff of \$5.50–\$6.25/bbl for 10-year committed shippers (Gordon, 2017).

Our calibrated model indicates that crude-by-rail has likely had economically meaningful effects on pipeline investment. We find that a \$2 per barrel increase in the cost of rail transportation, consistent with internalization of the externalities estimated in Clay et al. (2017), results in an increase in equilibrium pipeline capacity of between 53,000 and 131,000 bpd, relative to the actual DAPL capacity of 520,000 bpd (and total Bakken pipeline export capacity of 1.283 million bpd). These effects are likely to be lower bounds, as they do not account for economies of scale in pipeline construction. Moreover, we show that these capacity changes are associated with large decreases in rail shipments: the estimated elasticity of rail volumes to the cost of crude-by-rail ranges from -0.89 to -1.98 across specifications. In contrast, this elasticity only takes a value of roughly -0.2 when we hold pipeline capacity

[6]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.

fixed.

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 results imply that the option value offered by railroads can erode the incentive to invest in pipeline capacity. An implication of this result is that policies that increase the cost of rail transport—such as regulations targeting rail's environmental externalities—can increase pipeline investment and thereby induce economically significant long-run substitution from railroads to pipelines.

More broadly, the model we develop to understand the tradeoffs between pipeline and rail transportation captures intuition that is applicable to other settings in which a low-cost but inflexible technology substitutes with a high-cost, flexible technology. For instance, the logic by which relative costs and future demand uncertainty affect pipeline investment also applies to investments in infrastructure such as urban light rail (which substitutes with more flexible buses), natural gas distribution lines (which substitute with more flexible heating oil delivery), and baseload electric power stations (which substitute with more flexible gas-fired peaker units).[7]

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. Section 3 then describes our data and model calibration. Section 4 follows with a discussion of the empirical relationship between crude-by-rail flows and oil price differentials, and then describes how we use this relation to inform a version of our model in which rail shipments require short-term contracts. We discuss the validation of our model in section 5, and we present our main results in section 6. 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

[7]Prior work, such as Borenstein (2005), has studied the equilibrium allocation between baseload and peaker electric generation, though without building the intuition developed here.

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 3.3 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. In section 4.3 we discuss how we augment the model presented here to account for these contracts.

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 at the pipeline terminus (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 $[\underline{P}, \bar{P}]$.[8]

- 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.[9]

- 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 fixed 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 $P_d > P_r(K)$.

[8] 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.

[9] 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.

2.2 Equilibrium pipeline capacity in the single destination model

Consider a simple two-period version of our model. In period 1, prospective shippers decide whether to make ship-or-pay commitments to the pipeline. Then in period 2, the pipeline is completed with a capacity equal to the total commitment, P_d is realized, and shippers can decide whether to also ship crude by rail.

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, iff $r'(Q_r) > 0$.[10]

The expected return to pipeline shippers is then given by the shaded area in figure 1, weighted by the probability distribution $F(P_d)$. The equilibrium capacity K will balance this expected return (which decreases in K) against the pipeline investment cost of $C(K)/K$.

Figure 1 also 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 just the shaded trapezoid shown in the figure.

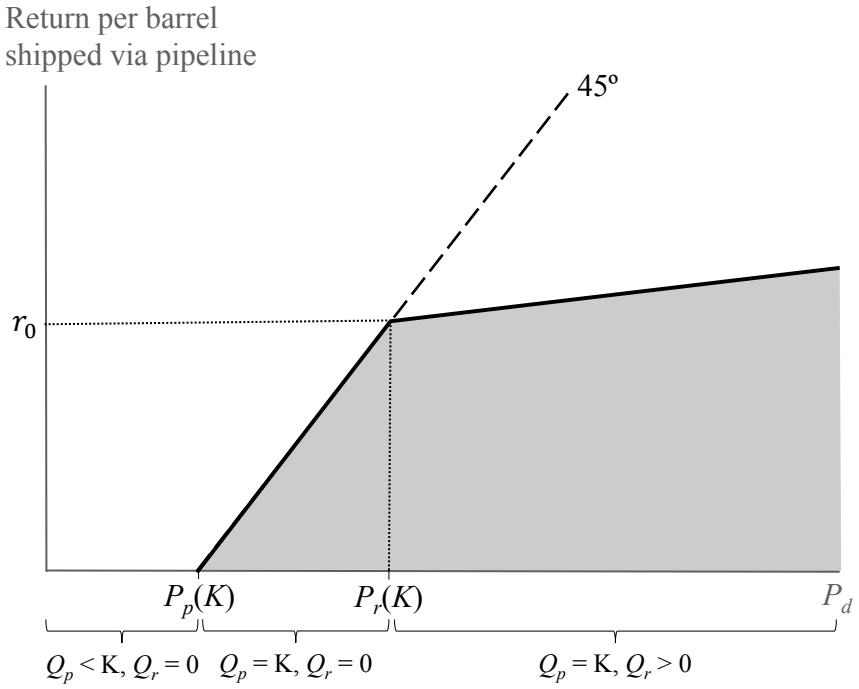
Formally, the condition that governs the equilibrium capacity level is given by equation (1), where the first term on the right-hand side captures returns to pipeline shippers when the pipeline is at capacity but rail is not used, and the second term captures returns when P_d is sufficiently high that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

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

Figure 1: Expected return achieved by pipeline shippers

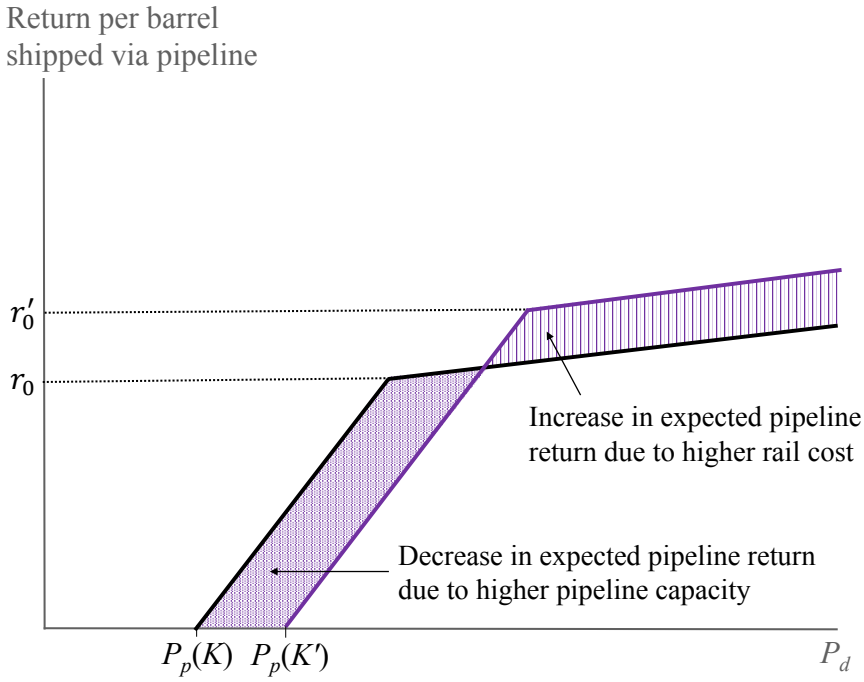


Note: P_d denotes the downstream price, with distribution $F(P_d)$. Q_p and Q_r denote crude oil pipeline and rail flows, respectively. r_0 denotes the intercept of the rail marginal cost function $r(Q_r)$. The shaded area, probability-weighted by $F(P_d)$, represents the expected return to a pipeline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

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 area. This increase in expected return then increases shippers' willingness to commit to capacity, so that equilibrium pipeline

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.

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).[11] 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:[12]

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

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

[11] 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 $Q_r(P_d)$.

[12] 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.

to scale, this expression 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, dK/dr_0 is then 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. Third, and 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.

Expression (2) also clarifies that the following information is required to obtain an estimate of dK/dr_0:

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 3 discusses our calibration of these parameters, which uses estimates from our own calculations as well as estimates from prior studies. Our numerical implementation of the model also models each month of the 10-year shipper commitment as a separate period, thereby allowing the distribution $F(P_d)$ to vary over the life of the commitment.[13] Our calculation of the expected return to pipeline shippers is then the average of the expected

[13] As discussed in section 3.2, from the perspective of shippers at the time of the commitment, P_d is more uncertain at the end of the commitment period than it is at the beginning of the commitment period.

returns over all 120 months of the commitment period.[14] As a consequence, we evaluate the derivative dK/dr_0 numerically rather than use equation (2) directly.

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). 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 its empirical validity in section 3.2.3.

Appendix B presents a derivation of the conditions that determine the equilibrium pipeline capacity K and its sensitivity to r_0 in this multiple rail destination model. Because rail transportation is more valuable when it can service destinations other than the pipeline terminus—and thereby reach prices that may be higher than P_d—equilibrium pipeline capacity is smaller than in the case in which rail can only serve a single destination. In addition, the sensitivity of pipeline capacity to the cost of railroad transportation, dK/dr_0, is larger than in the single destination model, since there is a reduced probability that the pipeline is full but rail is not used.

This model requires additional calibration relative to the single destination model. Beyond just needing the distribution of the pipeline downstream price $F(P_d)$, we also require

[14] We weight each period's expected return by its discount factor back to the date of the commitment.

the distribution of best rail price *conditional* on that price: $F(\widetilde{P} \mid P_d)$. We discuss the estimation of this distribution in section 3.2.3.

3 Data and calibration

To quantify the economic impact of crude-by-rail's flexibility on equilibrium pipeline capacity, we calibrate our model of pipeline investment to the recently constructed Dakota Access Pipeline (DAPL). Using the reported pipeline capacity and estimates of the distribution of future downstream prices anticipated by shippers, the cost function for crude-by-rail, and the upstream supply elasticity, we compute the average-cost tariff at which a marginal shipper would commit to the pipeline and then estimate the sensitivity of the pipeline's size to the cost of railroad transportation.

3.1 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 LP, 2014a). 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.[15] In addition, the North Dakota Pipeline Authority forecast that Bakken production would reach 1.5 million bpd within a few years at expected crude oil prices (North Dakota Pipeline Authority, 2014). Per Biracree (2016) and North Dakota Pipeline Authority (2017), existing local refining capacity was (and remains) 88 mbpd, and other Bakken export pipeline capacity was 763 mbpd (including a planned 24 mbpd expansion of the Double H pipeline that was completed in 2016).

DAPL was put into service in June, 2017 with a capacity of 520 thousand bpd (mbpd) and at a reported construction cost of \$4.78 billion (Energy Transfer Partners LP, 2017).[16] It is difficult to be certain, however, of the volume of capacity to which shippers committed in June, 2014. For one, the official June 2014 announcement of the successful open season stated a volume of 320 mbpd (Energy Transfer Partners LP, 2014a), though by September 2014 DAPL announced executed precedent agreements with shippers supporting a capacity of 450 mbpd (Energy Transfer Partners LP (2014b) and Phillips 66 (2014)). Second, back in 2012 a competing project, the Sandpiper Pipeline, had secured shipper commitments for a 225 mbpd line from the Bakken to Lake Superior (Enbridge Energy Partners LP (2012)

[15] 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.

[16] Construction cost includes \$1 billion for the Energy Transfer Crude Oil Pipeline (ETCO) from Patoka, IL to Nederland, TX on the Gulf Coast.

and Enbridge Energy Partners LP (2015)). This project was beset by environmental permitting delays in Minnesota and was postponed indefinitely in September, 2016 after Enbridge (Sandpiper's main owner) and Marathon (Sandpiper's "anchor shipper") invested in DAPL and cancelled their Sandpiper shipping agreement (Shaffer (2014), Marathon Petroleum Corporation (2016), and Darlymple (2016)). It is not clear to what extent Sandpiper's demise was foreseen in June, 2014, when shippers initially committed to DAPL. The reference case calibration of our model assumes a committed capacity of 520 mbpd, equal to the DAPL capacity actually constructed and approximately equal to the total DAPL plus Sandpiper capacity that had been announced by June, 2014 (320 mbpd for DAPL and 225 mbpd for Sandpiper).[17] As sensitivities, we will also examine results based on assumed capacities of 320, 450, and 570 mbpd.[18]

3.2 Crude oil prices

3.2.1 Price data

We obtained data on spot market crude oil prices from Bloomberg and Platts.[19] 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.

Figure 4 plots these spot price data, aggregated to the monthly level. Panel (a) shows that prices at the three coastal destinations are very tightly correlated, typically differing 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.[20] In addition, both panels of figure 4 clearly illustrate the substantial decrease in crude oil prices that occurred during the second half of 2014.

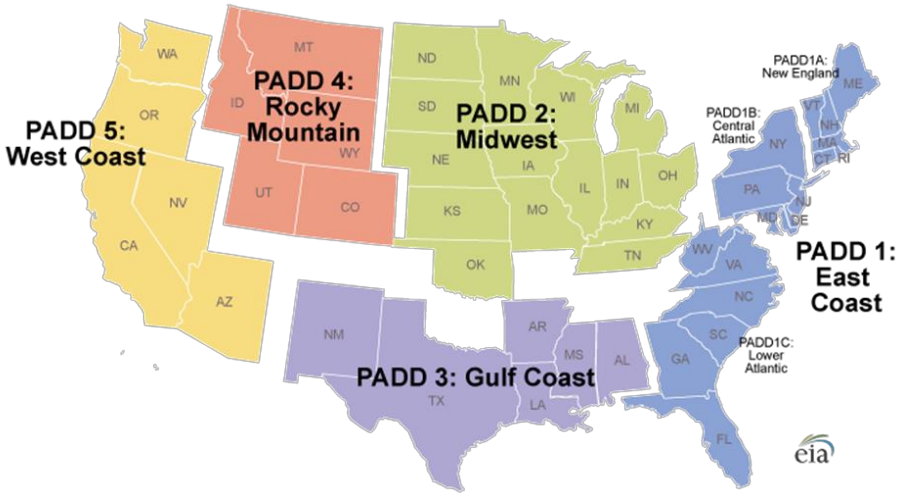
[17]The increase in DAPL capacity to 520 mbpd occurred following a supplemental open season held in early 2017 (Energy Transfer Partners LP, 2017).

[18]DAPL was built with the ability to expand to 570 mbpd and initiated an open season on the incremental 50 mbpd in March, 2018 (Darlymple (2016) and Energy Transfer Partners LP (2018)).

[19]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.

[20]Note that the Clearbrook price series does not begin until May 4, 2010.

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



Source: EIA

Because Bakken crude oil is quite light relative to ANS, the ANS benchmark may understate the value of Bakken crude on the West Coast.[21] Because LLS and Heavy Louisiana Sweet (HLS) are comparable to Bakken and ANS, respectively,[22] we add the historic price difference between LLS and HLS, equal to \$0.53/bbl, to the ANS price series in all of our calculations.[23]

3.2.2 Calibration of downstream oil price distribution for pipeline shippers

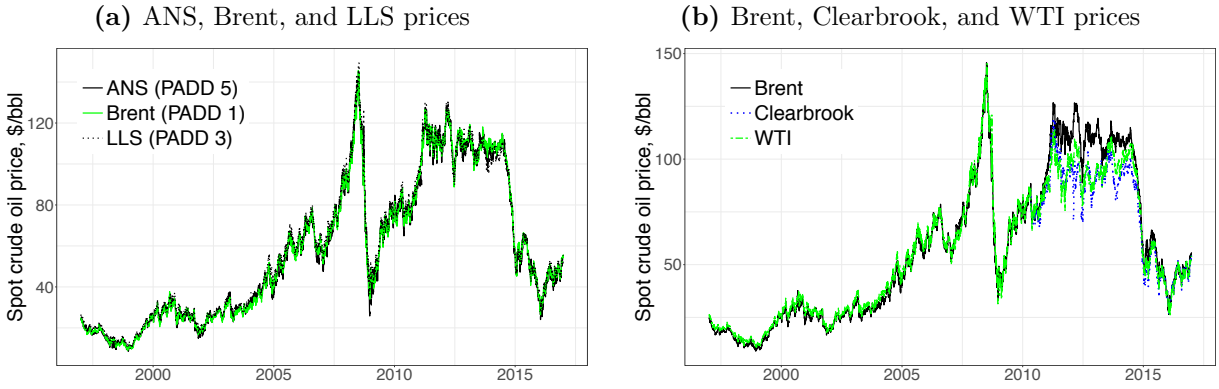
To calculate the expected return to pipeline shippers, our model requires an estimate of the distribution $f(P_d)$ of future downstream prices that these shippers face over the duration of their ship-or-pay commitments, which extend 13 years into the future (covering the 10-year commitments and a 3-year construction period). We assume a lognormal distribution for $f(P_d)$, which requires us to specify a mean and variance for each month of shippers' commitment period. 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 throughout their commitment period, since there is no LLS futures market and since the Brent and LLS prices have historically been

[21] ANS crude is 32 API and 0.96% sulfur, and Bakken is 43.3 API and 0.07% sulfur (S&P Global Platts, 2017).

[22] LLS crude is 38.4 API and 0.388% sulfur, and HLS crude is 33.4 API and 0.416% sulfur (S&P Global Platts, 2017).

[23] This price difference is calculated over the entire period for which both prices are available from Bloomberg: January 6, 1988 to March 31, 2017.

Figure 4: Crude oil spot 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.

quite close (figure 4).[24]

To obtain the variance of $f(P_d)$ for each month of shippers' commitment, 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).[25] To do so, we estimate the standard deviation of long differences in logged monthly Brent prices. For instance, the standard deviation of 37-month differences in log(price) yields the expectation, taken at the time of commitment (June, 2014), of price volatility in the first month of pipeline service. 13-year differences yield expected price volatility in the month in which the 10-year shipping commitments expire.

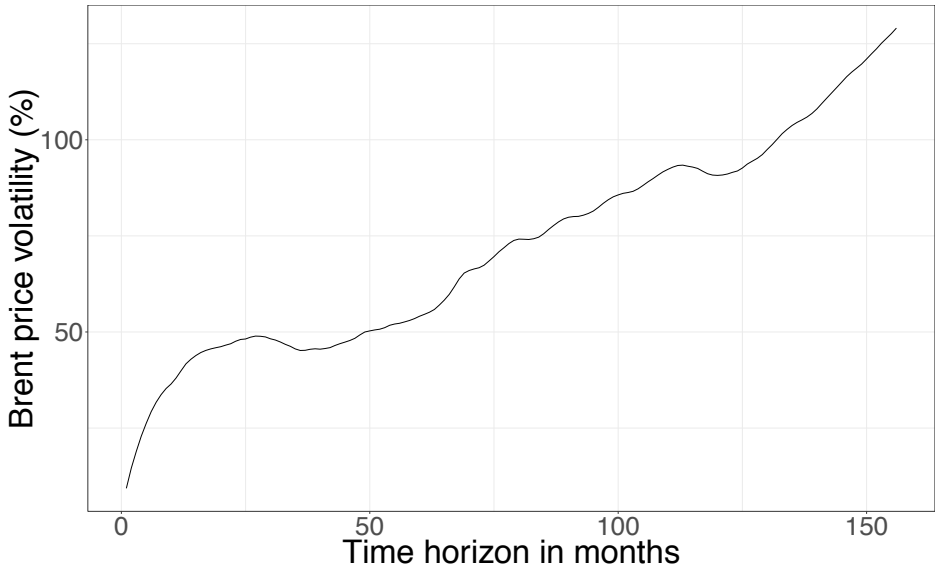
We calculate price volatilities for all horizons using Brent price data from May, 1996 (the first observation for which a 13-year lag is available in the data) through May, 2014 (the last full month before shippers committed to DAPL). These volatilities are presented in figure 5, which shows that uncertainty over the future price of Brent increases substantially over the 3-year to 13-year horizon, from a volatility of 45% at 37 months to 129% at 13 years.[26]

[24]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).

[25]An alternative approach to calculating expected future price uncertainty would be to use the prices of oil futures options, as in Kellogg (2014). However, futures options are not liquid beyond a two-year time horizon. Thus, we project average historic price volatility as our measure of expected future volatility.

[26]Figure 5 is not strictly monotonic due to finite sample variation in the Brent price history. Noise in the relationship between volatility and time horizon is effectively averaged out when the pipeline investment

Figure 5: Historic Brent price volatility over 3-year to 13-year time horizons



Note: Volatilities are calculated from standard deviations of long differences in historic logged Brent crude prices. Volatilities in percent are calculated for each horizon by exponentiating the standard deviation, subtracting one, and multiplying by 100. See text for details.

3.2.3 Calibration of 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.[27] Though the model assumes that shipments to all rail destinations face a common cost r_0, our cost data, discussed in section 3.4, 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. We then 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.

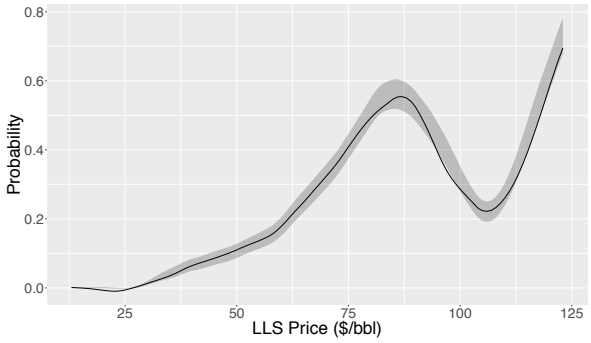
By construction, D has a point mass at 0. We therefore non-parametrically estimate the conditional distribution of D in two steps. First, we estimate $Z(P_d) = \Pr(D = 0 \mid P_d)$, the

model calculates the expected return to pipeline shippers over the full duration of their 10-year commitments.

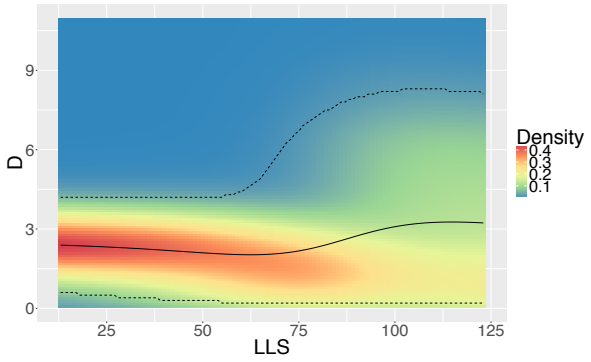
[27] We exclude those 8 days from the estimation procedure.

Figure 6: 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.

probability that P_d is the best downstream rail destination.[28] Second, we estimate $f(D \mid P_d, D > 0)$.[29] 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 6 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, though there is a departure from monotonicity when LLS prices are between \$90–\$105 per barrel. LLS prices in this range occurred frequently during 2011–2014, and rarely before. For the 20 or so years prior to 2011, Brent usually traded at discount to LLS of a few dollars per barrel. Starting in 2011, however, this discount frequently became a premium, sometimes by as much as \$10 per barrel.[30] Thus, despite high LLS prices, 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.

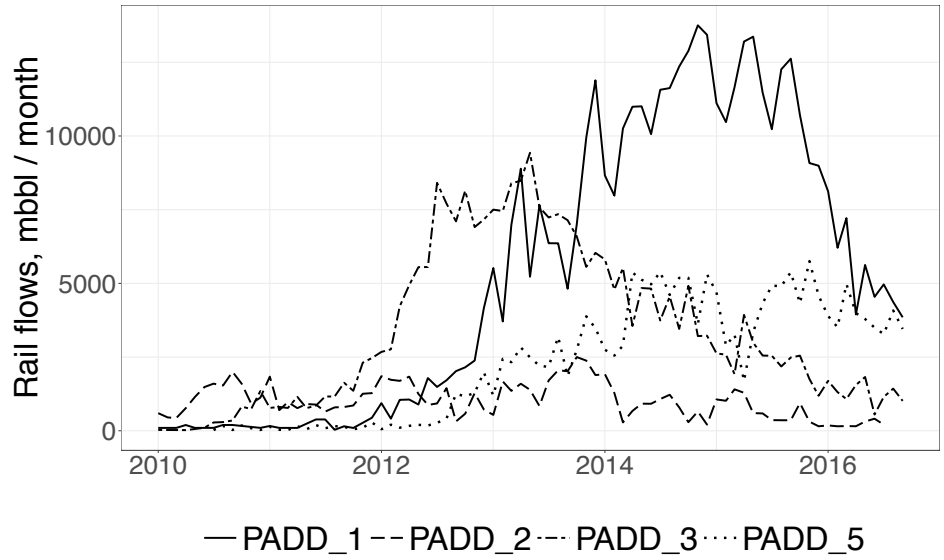
Panel (b) of 6 shows the density of D, conditional on $D > 0$ and P_d. For most values of P_d, this density is concentrated around just a few dollars per barrel, with a downward trend

[28] We estimate $\Pr(D = 0 \mid P_d)$ by running a locally quadratic regression of an indicator for whether D is zero onto P_d. We use the **nprobust** R package, provided by Calonico, Cattaneo and Farrell (2017).

[29] We estimate $f(D \mid P_d, D > 0)$ using the **hdrcde** R package, which implements locally linear conditional density estimation methods in Hyndman and Yao (2002).

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

Figure 7: Crude-by-rail monthly volumes from PADD 2, by destination PADD



Source: EIA

in the mode except at values of P_d in excess of roughly \$80/bbl.

3.3 Crude-by-rail flows

Data on monthly PADD-to-PADD flows of crude-by-rail were obtained from the EIA.[31] Figure 7 presents data on crude oil shipments from PADD 2 to PADDs 1, 2, 3, and 5.[32] 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 follow changes in price differentials with a non-trivial lag. This lag is consistent with the presence of contracting in the crude-by-rail market; section 4 will use these data to make inferences about the average contract duration and then develop a version of our pipeline investment model that accounts for crude-by-rail contracts.

[31] 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.

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

3.4 Railroad transportation costs

3.4.1 Railroad 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 States'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 2009–2015. These data contain detailed information on volumes and carrier revenues for a sample of shipments of crude oil and many other commodities.[33]

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,[34] we calculate the total revenue (across all shipments originating in PADD 2 and delivered to PADDs 1, 2, 3, or 5) earned by railroad carriers and divide by the total number of bbl-miles of crude oil shipped.[35] Figure 8, panel (a), presents the resulting time series of average revenue per bbl-mile.[36] This figure shows that railroad transportation rates were roughly constant around $5 per thousand bbl-mile from 2011–2014 before falling by approximately $1 per thousand bbl-mile in 2015. This decrease in transportation rates follows the sharp drop in crude oil prices that began in late 2014 (figure 4) and coincides with a decrease in crude-by-rail volumes (figure 7). The 2015 rate decrease does not appear to merely reflect a compositional change in shipments: figure 8, panel (b) shows that it persists after controlling for destination PADD, distance travelled, the number of carrying railroads, and the number of cars.

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 87.0% of crude oil moves on contract rates and that contract shipments enjoy an average discount of $0.52 per 1000 bbl-mile relative to tariff rates.[37]

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

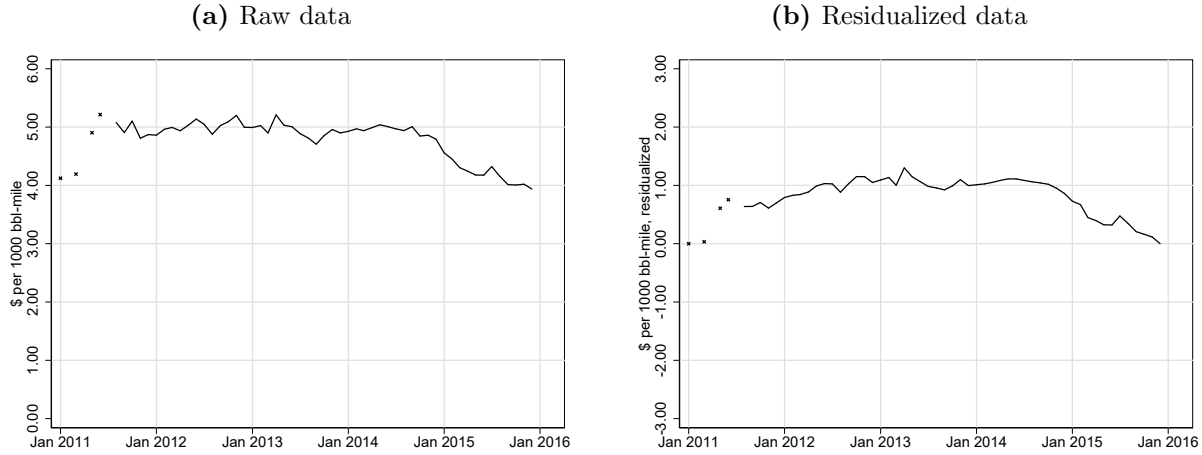
[34] To assign a movement date (and therefore a month) to each shipment, we follow the procedure described in Energy Information Administration (2017) to convert waybill dates to movement dates.

[35] 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.

[36] 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 prior to 2011.

[37] We calculate both of these numbers using the same sample of waybills as was used to generate figure 8. To obtain the contract discount, we regress revenue per 1000 bbl-mile on a tariff vs contract dummy variable

Figure 8: STB average revenue per bbl-mile shipped



Note: Data cover sampled waybills from PADD 2 to PADDs 1, 2, 3, and 5. Data from February, April, and July 2011 are omitted to protect the confidentiality of carriers' rates; data from months before July 2011 are therefore plotted as points rather than a line. Panel (b) plots date fixed effects (relative to the January, 2011 observation) from a waybill-level regression of revenue per 1000 bbl-mile on date FE, PADD FE, distance shipped, the number of carrying railroads, and the number of cars, weighted by bbl-miles. See text for details.

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. 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 begins in October, 2013.

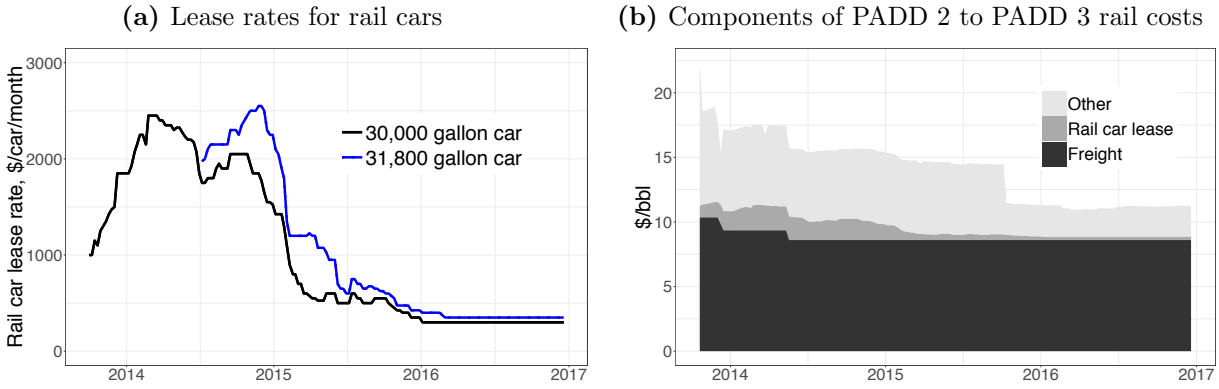
Panel (a) of figure 9 presents Genscape's assessments of leasing rates for rail cars. Lease rates rise in the first part of the sample, when the oil price is high and transportation volumes are growing, and then fall late in the sample, when the oil price is low and transportation volumes are falling. This pattern is consistent with scarcity rents or market power during the "boom" period that then dissipated when oil prices fell.[38]

Figure 9, 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

and on month-of-sample fixed effects, distance travelled, the number of carrying railroads, and the number of cars, while weighting the regression by bbl-miles. $0.52 is the regression coefficient on the tariff vs contract dummy variable. If we do not include any controls variables, we obtain a $0.71 contract discount.

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

Figure 9: 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.

costs (primarily terminal fees) come together to form the total cost of shipping crude by rail, using the PADD 3 destination as an example. Genscape's freight cost assessment is constant for most of the sample at a level that is roughly consistent with the STB data shown in figure 8, given an average trip distance to the gulf coast of about 1,900 miles (Clay et al. (2017)). Panel (b) of figure 9 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 data and reflects the removal of gathering costs from Genscape's assessments rather than an actual change in costs.

3.4.2 Calibration of the crude-by-rail cost function

We use the Genscape rail cost dataset to calibrate values of r_0 and r_1 for our model of pipeline investment, acknowledging the limitation that these data reflect cost assessments rather than hard private transaction data (which we do not believe are systematically collected by any entity). We estimate $r_{0,i}$ for each destination i as the minimum all-in rail cost for transportation between the Bakken and 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 had roughly halved since 2014. We select the minimum reported cost because these costs coincide with both fully constructed rail capacity and limited use of that capacity, which we view as a reasonable analogue to the model's notion

of r_0.[39]

We ultimately use several values for r_1 in our model. At the low end, we assume $r_1 = 0$, consistent with the facts that rail loading and unloading capacity now far exceeds rail flows of oil and that tanker cars should be viewed as commodities in the medium to long term. At the high end, we assume $r_1 = \$6$/bbl per million barrels per day (mmbpd), which fits Genscape's highest rail costs of \$20-25 per bbl that were realized in October 2013. These high costs, however, likely reflect short-run constraints in rail terminals and tank cars that would be unlikely to persist over the multi-year horizon relevant to the pipeline customers considered by our model. We therefore view $r_1 = \$6$/bbl per mmbpd as an upper bound and focus most of our attention on calibrations of the model that use an intermediate value of $r_1 = \$3$/bbl per mmbpd.

3.5 Upstream oil supply elasticity

Our model of pipeline investment includes a static supply curve for Bakken crude oil production. This construct is inherently strained given that oil is an exhaustable resource; we adopt it here both to maintain our model's tractability and because estimation of a dynamic model of Bakken drilling and production has not been done in the literature and is beyond the scope of this paper. Due in part to this shortcoming, we ultimately calibrate our model's crude oil supply curve using a range of elasticities.

Rather than estimate the Bakken supply elasticity ourselves, we rely on the literature that has endeavored to estimate supply curves for U.S. shale oil and gas. Bjørnland et al. (2017) finds that monthly Bakken production does not respond essentially at all to contemporaneous price shocks, consistent with Anderson, Kellogg and Salant (forthcoming)'s work showing that changes in oil supply come from drilling of new wells rather than changes in production from existing wells (which experience production declines that are invariant to price). We therefore turn to estimates of the drilling elasticity of U.S. shale wells. However, drilling elasticity estimates are likely to be substantial over-estimates of the impact of oil prices on upstream production over the 13-year time horizon relevant to shippers in our pipeline model, for two reasons. First, production from newly drilled wells is pooled with production from pre-existing wells, so that the percentage change in production after a price shock is substantially smaller than the percentage change in drilling for many years, until production from pre-existing wells declines to nearly zero. Second, as explained by Smith and Lee (2017), wells drilled following an increase in the price of oil will tend to be less productive than wells drilled previously.

[39] 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.

Newell and Prest (2017) find a drilling elasticity of roughly 1.6 for all U.S. shale oil, using a quarterly panel of drilling rates and crude oil prices.[40] For shale gas, Hausman and Kellogg (2015) estimates a price elasticity of about 0.8, using instrumental variables regressions of gas drilling rates on lagged gas prices. Newell et al. (2016) uses similar methods and estimates an elasticity of about 0.7. For our model, we vary roughly halve these drilling elasticity estimates to obtain a range of Bakken supply elasticities between between 0.4 and 0.8.[41]

3.6 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 (10). As a conservative baseline, we simply assume that the construction of DAPL has constant returns to scale, 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.[42] The elasticity implied by Soligo and Jaffe (1998) is 0.59, indicating substantial scale economies.[43] 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 use the elasticity of 0.59 from Soligo and Jaffe (1998).

4 Crude-by-rail flows and contracting

The model from section 2 assumes that rail shippers will, in every month: (a) select the destination with the highest upstream price, net of transportation costs; and (b) fluidly ad-

[40] Newell and Prest (2017)'s model includes three quarterly lags of oil prices and instruments for the oil price using prices of other commodities.

[41] A gross minimum for this elasticity would be about 0.3: this value comes from the decrease in North Dakota production from a peak of 1,178 mbpd in September 2014 to a trough of 940 mbpd in December 2016 (just more than two years later), relative to an oil price that roughly halved (data from EIA at https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=MCRFPND2&f=M; accessed March 18, 2018). Use of an elasticity higher than 0.8 would cause our model to estimate responses of pipeline capacity to crude-by-rail costs that are even larger than those presented in section 6.

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

[43] 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.

just the magnitude of these shipments as upstream prices rise and fall. In practice, however, crude-by-rail shipments violate both of these assumptions: figure 7 shows that every destination has positive rail flows in every month, starting in 2012, and figures 7 and 4 together suggest that rail volumes follow price movements with a lag.

A likely 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.[44] Although we do not have access to individual private shipping contracts, we know from the STB data (discussed in section 3.4) that most crude-by-rail shipments are on contracts, typically at a discount to spot rates. In this section, we first endeavor to learn about average contract durations by empirically examining the aggregate response of rail flows to price differentials, using two empirical strategies. We then develop an alternative version of our pipeline investment model that requires rail shipments to use contracts rather than operate on a spot basis.

4.1 Descriptive relationship between rail flows and crude prices

We begin 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 (NDPA) on the monthly share of each transportation mode (local consumption, crude-by-rail, pipeline, and truck) with data from the EIA on the monthly share of each rail destination for shipments originating in PADD 2. Because our data end in 2016, 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.[45] We divide the total rail share from the NDPA data into destination-specific crude-by-rail shares using the EIA data. The combined data yield 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

[44]Simultaneous multi-destination flow may also be explained in part by: (1) destination-specific marginal shipping costs that are strictly increasing; or (2) destination-specific demand for Bakken crude that is strictly decreasing.

[45]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 and Double H pipeline systems, which connect 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.

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_{ijt} 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 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:

$$\log s_{jt} - \log s_{0,t} = \sum_{l=0}^{L} \beta_{t-l} \left(p_{j,t-l} - p_{0,t-l}\right) + \delta_j - \delta_0 + \xi_{jt} - \xi_{0t} \tag{4}$$

Table 1 presents OLS estimates of equation (4) using contemporaneous destination prices and price lags of order 3 to 24 months. In general, the coefficient on the oldest price realization in each specification has the most positive value. 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 one to two years.[46]

[46]For example, if all shippers signed T-month contracts, then the price vector driving month t's destination

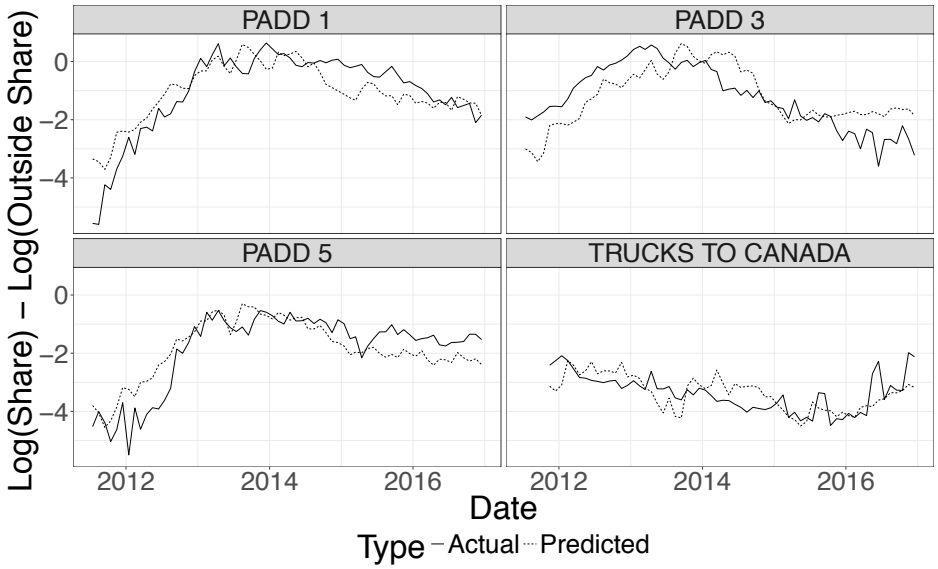
Table 1: Multinomial Logit Shipment Destination Share Regressions

	(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

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

Note: Table presents multinomial logit regressions of monthly shipment destination shares onto current and lagged monthly 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 10: 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.

Another pattern in table 1 is the negative and precisely estimated coefficients on contemporaneous prices. These negative coefficients likely reflect standard supply/demand endogeneity: if ξ_{jt} constitutes a supply shock (e.g., the unobserved opening of a new loading or unloading facility) and local crude demand is downward-sloping, contemporaneous prices at destination j will be negatively correlated with the shock. Thus, the coefficient on contemporaneous prices will be biased downwards.

Despite the simplicity of this destination-share model, it fits the data reasonably well. Figure 10 plots actual and fitted time series of the log odds ratios for each destination. 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, suggesting an important role for contracting in determining rail flows and 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.

shares would be p_{t-T}.

4.2 A destination share model with explicit rail contracts

To better infer average rail contract durations, we modify the destination share model from section 4.1 by explicitly modeling rail contracts. Let $R_{i,t}$ be the flow of crude to destination i during month t, and let M_t be the total output of North Dakota crude during month t. Let $q_{i,t}$ be the share of "uncontracted" crude oil at month t that goes under contract to location i, and let the parameter s denote the share of previously contracted crude oil that is still under contract in the next month. $R_{i,t+1}$ can then be decomposed into *new flows*, which we hypothesize respond to contemporaneous price shocks, and *old flows*, or previously contracted flows, which should be the result of previous price shocks:

$$R_{i,t+1} = \underbrace{sR_{i,t}}_{\text{old flows still contracted}} + q_{i,t+1} \underbrace{\left(M_{t+1} - s \sum_j R_{j,t} \right)}_{\text{new flows available to be contracted}}$$

If we view individual contracts as infinitesimal claims to the total flow M_t that do not retire between periods t and $t+1$ with probability $s \in (0,1)$, then the average contract length will be $\frac{1}{1-s}$. Our goal is to estimate s and therefore $\frac{1}{1-s}$.

We must first specify a model for $q_{i,t+1}$, the share of available production that goes under contract to destination i in period $t+1$. We adopt a random utility choice model. Atomistic shippers observe a vector of crude price differentials $X_{i,t+1}$ and destination fixed effects δ_i that account for shipment cost differences or other fixed facors. They then choose the location with the highest value of indirect utility $X_{i,t+1}\beta + \delta_i + \epsilon_{i,t+1}$, where $\epsilon_{i,t+1}$ is an *i.i.d.* logit shock. Following algebra that we ensconce in Appendix C, we can then re-write the contracting model as equation (5), in which $R_{0,t+1}$ denotes flow to the outside good (pipeline to Cushing):

$$R_{i,t+1} = sR_{i,t} + R_{0,t+1} \exp\left(X_{i,t+1}\beta + \delta_i\right). \tag{5}$$

Our data on $R_{i,t}$ are likely measured with error, since they are computed by the EIA from the STB's sample of waybills, in which sampling rates can be below 10%.[47] To account for this issue, let $\mu_{i,t}$ be an iid measurement error that is orthogonal to $R_{i,t}$, so that observed

[47] This measurement error precludes using a standard log transformation to estimate the equation $\log\left(R_{i,t+1} - sR_{i,t}\right) = \log R_{0,t+1} + X_{i,t+1}\beta + \delta_i + \xi_{i,t+1}$. In this specification, s must be sufficiently small such that the term inside the left-hand-side logarithm is strictly positive for all observations. With measurement error in the $R_{i,t}$, however, especially in the early part of the sample when rail flows are small, we have found that s then cannot be greater than roughly 0.1, implying an average contract length of only a month or two.

flows are $R^*_{i,t} = R_{i,t} + \mu_{i,t}$. A feasible version of our estimating equation is then:

$$R^*_{i,t+1} = sR^*_{i,t} + R_{0,t+1} \exp\left(X_{i,t+1}\beta + \delta_i\right) + \mu_{i,t+1} - s\mu_{i,t}. \tag{6}$$

Because the measurement error $\mu_{i,t}$ is mechanically correlated with $R^*_{i,t}$, we instrument for $R^*_{i,t}$ using the first three lags of destination price differences, in levels, squares and cubes. Lagged prices should influence $R^*_{i,t}$ through contracting forces but will be uncorrelated with sampling-based measurement error in the EIA data. The instrument set also includes $R_{0,t+1}$ and destination dummy variables.[48]

The results of estimating equation (6) are in presented in table 2. The point estimate on s is 0.95, which implies an average contract length of about 18 months. The coefficient on downstream prices is positive, consistent with the latent shipping choices being driven by variation in netbacks, and the point estimates on the location dummies are all negative, consistent with rail transportation being more expensive than pipeline transportation.[49] Though the estimated destination dummies are economically large relative to the estimated price coefficient $\hat{\beta}$, they imply economically plausible effects on shippers' choice among rail destinations. For example, the difference in downstream dummies between PADD 1 and PADD 5 imply that shippers prefer PADD 1 until the downstream price at PADD 5 exceeds the PADD 1 price by \$4.50/bbl.

4.3 Embedding rail contracting into the pipeline investment model

To address the importance of contracting in the crude-by-rail (CBR) market, we develop an alternative version of the model from section 2 by requiring rail shippers to use contracts rather than a spot market for rail services. Because we do not observe individual rail contracts, we do not know the extent to which these contracts allow the shippers to retain flexibility in when to ship (as in the case of volume discount agreements) or completely foreclose flexibility (firm ship-or-pay commitments). As a bounding exercise on the importance of CBR contracts, we examine how the incentives of both rail and pipeline shippers change when rail shipments require firm ship-or-pay commitments.

Informed by our estimated average contract length from section 4.2, we impose that rail shippers must sign 20-month ship-or-pay commitments in order to access the railroad.[50]

[48] We do not use contemporaneous price differences in order to avoid bias from standard supply/demand endogeneity of the price differences $X_{i,t+1}$.

[49] The magnitude of these dummies, relative to the estimated price coefficient $\hat{\beta}$, is very large relative to the difference between the cost of rail and the cost of the outside good (pipeline to Cushing), which is at most \$15/bbl. The revealed preference for pipeline shipments in these estimates may reflect long-term ship-or-pay commitments made by shippers on crude oil pipelines to Cushing.

[50] We choose 20 rather than 18 months because 20 month contract periods divide evenly into the 10-year

Table 2: GMM estimates of the rail contract model

	Estimates
s	0.95 (0.05)***
β	0.08 (0.04)**
δ_1	−4.51 (1.46)***
δ_2	−5.58 (1.05)***
δ_3	−5.19 (1.68)***
δ_5	−4.87 (1.20)***
δ_T	−5.36 (0.68)***
Criterion function	2318.24
J-Test	8.93
Num. obs.	385

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Note: Table presents estimates of equation (6). s denotes the share of contracted flows that do not expire each month, and β is the coefficient on destination i's oil price differential relative to Cushing. The δ_i's are rail destination fixed effects. δ_T refers to shipments to Canada via truck. The outside good is shipment to Cushing by pipeline. Two-step GMM HAC standard errors in parentheses. See text for details.

During the contract period, rail shippers then hold the option to ship crude volumes up to the committed capacity level, at no additional marginal cost. Thus, every 20 months, beginning with the in-service date of the pipeline, prospective rail shippers face a problem similar to that faced by pipeline shippers: they must decide whether to commit to rail capacity, even though there is uncertainty regarding the future downstream crude price over the life of the contract.

Once their 20-month contracts expire, rail shippers can decide the quantity of rail capacity to renew, based on the current downstream price. Over the 10-year commitment period for firm pipeline shippers, there are six rail contracting "cycles". This modeling approach therefore diminishes the flexibility of CBR. Instead of being able to freely adjust volumes shipped on a monthly basis in response to downstream price shocks, rail shippers may only adjust their capacity limit every 20 months. In addition, we assume that the 20-month ship-or-pay commitments signed by rail shippers are location-specific, so that they can no longer take advantage of transitory price differences across downstream locations. Thus, rail shippers in the contracting version of the model can only realize the downstream price P_d at the pipeline terminus.

The fact that rail shippers cannot immediately increase shipment volumes in response to an increase in P_d means that pipeline shippers enjoy higher shipping margins following

commitment period for pipeline shippers.

large realizations of P_d, relative to the baseline model.[51] Thus, pipeline shippers are willing to commit to more pipeline capacity in the rail contracting version of the model, all else equal. Because ship-or-pay contracts are maximally restrictive (relative to volume discounts or rebates), we believe that this rail contracting model generates a lower bound on the value of CBR and an upper bound on rail shippers' willingness to pay for capacity.

Finally, we do allow rail shippers to enjoy a reduced shipping rate relative to the baseline model. Per our estimates from the STB data (section 3.4), we reduce the value of r_0 in the contracting model by \$0.52 per 1000 bbl-miles relative to the baseline model to account for contract discounts.[52]

5 Validation

Given the inputs discussed above, we validate our model by calculating the expected return for shippers who signed ten-year firm shipping agreements on DAPL, given the pipeline's actual capacity of 520 mbpd (which increased total Bakken pipeline export and local refining capacity to 1,371 mbpd). We then compare this expected return—which in equilibrium should equal the average per-bbl cost of the pipeline—to the actual DAPL tariff for long-term shippers of \$5.50–\$6.25/bbl (Gordon, 2017).

Table 3 presents the implied average per-bbl cost of DAPL from our model, covering a range of assumptions on the upstream supply elasticity, the responsiveness of crude-by-rail (CBR) costs to rail flow (r_1), and the spatial and temporal flexibility of CBR (note that the implied cost is not affected by assumptions about pipeline returns to scale). Costs implied by our baseline (single-destination, spot CBR) model are given in the third column of table 3. These costs naturally vary with the assumed values for the supply elasticity and for r_1, but they are generally in the neighborhood of the \$5.50–\$6.25/bbl tariff.

The implied cost generally decreases 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 so far as to decongest the pipeline, thereby leaving pipeline shippers with zero return, is therefore increasing in the supply elasticity.[53] The implied cost increases with r_1 because higher

[51] Pipeline shippers also suffer from lower margins following low realizations of P_d, since rail shippers will ship at margins less than the (sunk) cost of rail. However, margins cannot fall below zero (for pipeline or rail) because shippers are not obligated to ship. This convexity of returns as a function of P_d leads to greater expected returns for pipeline shippers in the contracting model.

[52] Given the 1,900 mile distance from the Bakken to the Gulf Coast, the overall contract discount is roughly \$1/bbl.

[53] More precisely, 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. For values of $r_1 > 0$, the implied cost can increase with the

Table 3: Model validation

Input parameters		Implied average cost per bbl		
Supply elasticity	r_1 $/bbl per mmbpd	Baseline model	Multiple dest. model	CBR contract model
0.4	0	$5.61	$4.94	$6.76
0.6	0	$5.03	$4.42	$6.34
0.8	0	$4.74	$4.18	$6.12
0.4	3	$6.01	$5.35	$7.31
0.6	3	$5.56	$4.98	$7.09
0.8	3	$5.44	$4.90	$7.10
0.4	6	$6.39	$5.75	$7.81
0.6	6	$6.08	$5.51	$7.75
0.8	6	$6.10	$5.58	$7.92

Note: The actual DAPL tariff for ten-year committed shippers is between $5.50/bbl and $6.25/bbl. All rows assume $r_0 = \$11$/bbl. r_1 is in units of $/bbl per million barrels per day (mmbpd). The baseline model allows rail to reach a single destination (Gulf Coast) and allows rail transport to proceed on a spot market. The multiple destination model allows rail to reach the Gulf, East, or West coasts, using the spot market. The crude-by-rail (CBR) contract model allows rail to reach a single destination and constrains rail to use 20-month ship-or-pay contracts. See text for details.

values of r_1 increase the margin earned by pipeline shippers when CBR volumes are large.

In the model that allows rail shippers to enjoy spatial option value, shown in the fourth column of table 3, the implied DAPL cost is roughly 10% lower than in the baseline model. This fall in implied cost is driven by the decrease in shipping margin that occurs when rail shippers increase their volumes—commensurately increasing upstream production and the upstream price—in response to realized high prices at downstream locations other than the pipeline terminus. The modest difference in implied cost between the spatial and baseline models is consistent with small historical price differences across the U.S. Gulf, West, and East Coasts (see figure 4).

In contrast, when CBR shipments require 20-month ship-or-pay contracts, shown in the right-most column of table 3, the value of CBR diminishes and the expected margin to pipeline shippers increases. The expected margin, and thus the implied average cost of DAPL, is nearly $8/bbl in specifications with $r_1 = \$6$/bbl per million bpd (mmbpd). We believe that the 20-month ship-or-pay contracts we model are overly restrictive relative to actual CBR contracts, so that the "true" model for expected shipper returns lies between the fully flexible spatial model and the CBR contract model.

supply elasticity because large CBR volumes increase the return to pipeline shippers, and the magnitude of this effect can outweigh that from the increased probability that the pipeline is uncongested.

Finally, table 6 in the appendix presents implied costs using the baseline model with alternative values for DAPL's capacity. For capacities of 450 mbpd or 570 mbpd, implied costs are similar to those shown in table 3 (which uses the actual installed DAPL capacity of 520 mbpd), but costs are substantially higher (up to \$9.03/bbl) if one assumes that only 320 mbpd of new Bakken export capacity were committed to in June, 2014.

6 Results: the sensitivity of pipeline investment to rail costs

Table 4 presents results from the baseline model, with no spatial option value or crude-by-rail (CBR) contracting, and assuming constant returns to scale in pipeline construction. The top third of this table uses $r_1 = \$0$/bbl per mmbpd, consistent with an assumption that, over the long run, CBR exhibits constant returns to scale. Under this assumption, our estimates of dK/dr_0 range from 32.7 thousand bpd (mbpd) per \$/bbl (for an upstream supply elasticity of 0.4) to 54.9 mbpd per \$/bbl (for an elasticity of 0.8). Given the actual DAPL capacity of 520 mbpd (and total Bakken pipeline export capacity K of 1,283 mbpd), these values translate to elasticities of total pipeline export capacity K with respect to r_0 that range from 0.28 to 0.47. Because estimated air pollution externalities from rail transportation exceed \$2/bbl (Clay et al. (2017)), these estimates imply substantial long-run substitution from rail to pipeline capacity in the event that CBR's externalities are addressed through regulation.

The middle and bottom sections of table 4 display results when we use values for r_1 of \$3 and \$6/bbl per mmbpd, respectively. Our estimates of dK/dr_0 decrease modestly with r_1, and our smallest estimate is $dK/dr_0 = 26.6$ mbpd per \$/bbl (with an upstream elasticity of 0.4 and $r_1 = \$6$/bbl per mmbpd), corresponding to an elasticity of K with respect to r_0 of 0.23. Because this estimate neglects railroad spatial option value, railroad contracting, and pipeline economies of scale, we view it as a lower bound on the effect of the cost of railroad transportation on pipeline capacity.

Appendix table 7 shows how our estimates vary with the assumed size of the June, 2014 pipeline capacity commitment K. The results do not substantially deviate from baseline for values of K between 450 mbpd and 570 mbpd. For $K = 320$ mbpd and an upstream supply elasticity of 0.4, however, the estimate of dK/dr_0 is nearly double its value in the baseline model.[54]

[54]Under these assumptions, the implied cost of DAPL is \$9.03/bbl (appendix table 6), so that the cost of pipeline shipment is similar to the cost of rail shipment ($r_0 = \$11$/bbl). Pipeline and rail transport are therefore close substitutes, inflating the estimate of dK/dr_0.

Table 4: Sensitivity of pipeline capacity and expected rail flow to the cost of crude-by-rail: results from baseline model

Input parameters		Impacts of a change in r_0 on K and $E[Q_r]$				
Supply elasticity	r_1	dK/dr_0	Elasticity of K w.r.t. r_0	$dE[Q_r]/dr_0$	Elasticity of $E[Q_r]$ w.r.t. r_0	Elasticity of $E[Q_r]$ w.r.t. r_0, K fixed
0.4	0	32.7	0.28	-17.6	-1.45	-0.20
0.6	0	43.5	0.37	-21.3	-1.29	-0.20
0.8	0	54.9	0.47	-25.7	-1.17	-0.20
0.4	3	29.4	0.25	-15.8	-1.32	-0.20
0.6	3	37.7	0.32	-18.5	-1.14	-0.20
0.8	3	45.7	0.39	-21.4	-1.01	-0.20
0.4	6	26.6	0.23	-14.3	-1.21	-0.20
0.6	6	33.2	0.28	-16.3	-1.02	-0.20
0.8	6	39.2	0.34	-18.4	-0.89	-0.20

Note: All rows assume constant returns to scale in pipeline construction and that r_0 = \$11/bbl. Capacity K and expected rail flows $E[Q_r]$ are in thousands of barrels per day (mbpd). r_1 is in units of \$/bbl per mmbpd. The baseline model allows rail to reach a single destination (Gulf Coast) and allows rail transport to proceed on a spot market. See text for details.

The fifth column of table 4 shows how expected rail flows, $E[Q_r]$, are impacted by a change in r_0. Across the specifications, $dE[Q_r]/dr_0$ has a smaller magnitude than does dK/dr_0, since for low downstream oil price realizations, rail flows will be zero regardless of the value of r_0.[55] Expressed as an elasticity, however, the estimated impact of rail costs on expected rail flows is large, ranging from -0.89 to -1.45.[56] The right-most column of table 4 compares these elasticities to those from counterfactual calculations of $dE[Q_r]/dr_0$ that hold pipeline capacity K constant. Absent the response of pipeline capacity, CBR volumes are quite insensitive to changes in rail transportation costs: the elasticity of $E[Q_r]$ to r_0 is only about -0.20. Thus, our results indicate that the main channel by which environmental regulation (or Pigouvian taxation) will reduce CBR volumes is substitution to investment in pipeline capacity.

Table 5 presents results from models allowing for multiple rail destinations, rail contracting, and pipeline economies of scale, focusing on the central case with r_1 = \$3/bbl per mmbpd. When CBR can flow to multiple destinations, the sensitivity of both pipeline capacity and expected rail flows to changes in r_0 is uniformly greater (by roughly 13–21%,

[55]Moreover, $dE[Q_r]/dr_0$ is not very sensitive to the upstream supply elasticity because this elasticity has two countervailing effects on $dE[Q_r]/dr_0$. First, large supply elasticities increase dK/dr_0, which will tend to increase the magnitude of $dE[Q_r]/dr_0$. Second, large supply elasticities increase the probability that $E[Q_r] = 0$, which will tend to decrease the magnitude of $dE[Q_r]/dr_0$.

[56]The elasticity of $E[Q_r]$ with respect to r_0 decreases in the upstream supply elasticity because $E[Q_r]$ increases with the upstream supply elasticity.

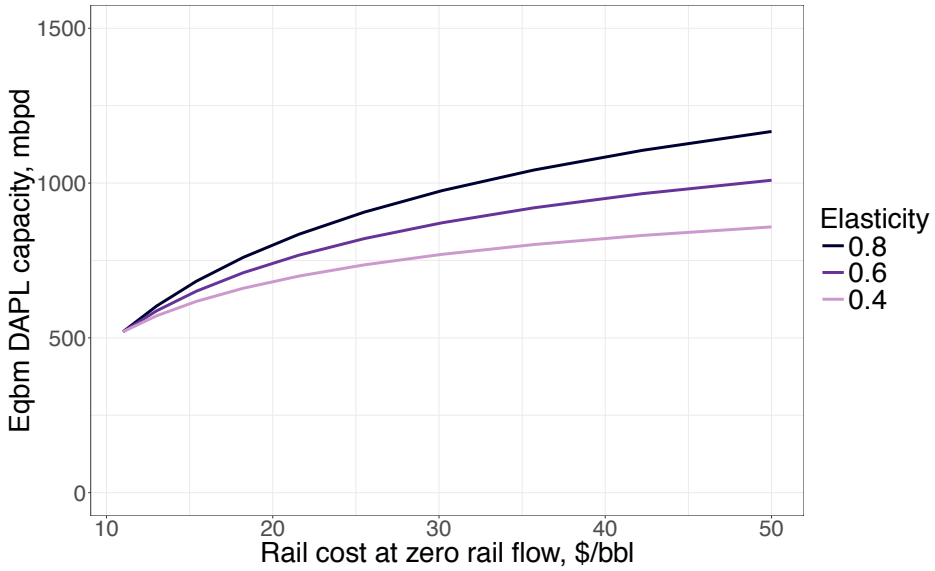
Table 5: Sensitivity of pipeline capacity and expected rail flow to the cost of crude-by-rail: alternative models for rail flexibility and pipeline economies of scale

Input parameters		Impacts of a change in r_0 on K and $E[Q_r]$				
Supply elasticity	r_1	dK/dr_0	Elasticity of K w.r.t. r_0	$dE[Q_r]/dr_0$	Elasticity of $E[Q_r]$ w.r.t. r_0	Elasticity of $E[Q_r]$ w.r.t. r_0, K fixed
Baseline model						
0.4	3	29.4	0.25	-15.8	-1.32	-0.20
0.6	3	37.7	0.32	-18.5	-1.14	-0.20
0.8	3	45.7	0.39	-21.4	-1.01	-0.20
Multiple rail destination model						
0.4	3	33.1	0.28	-17.9	-1.45	-0.20
0.6	3	44.4	0.38	-21.9	-1.31	-0.21
0.8	3	55.3	0.47	-26.0	-1.19	-0.21
CBR contracting model						
0.4	3	46.0	0.39	-23.9	-1.98	-0.23
0.6	3	56.0	0.48	-26.5	-1.63	-0.23
0.8	3	65.3	0.56	-29.6	-1.40	-0.24
Baseline model, increasing returns to scale						
0.4	3	42.4	0.36	-21.7	-1.81	-0.20
0.6	3	64.0	0.55	-29.1	-1.79	-0.20
0.8	3	96.0	0.82	-40.3	-1.90	-0.20

Note: Baseline model allows rail to reach a single destination (Gulf Coast) and allows rail transport to proceed on a spot market. Multiple destination model allows rail to reach the Gulf, East, or West coasts, using the spot market. Crude-by-rail (CBR) contract model allows rail to reach a single destination and constrains rail to use 20-month ship-or-pay contracts. Increasing returns to scale model assumes pipeline cost has a constant elasticity of 0.59 with respect to capacity. Capacity K and expected rail flows $E[Q_r]$ are in thousands of barrels per day (mbpd); r_1 is in units of \$/bbl per mmbpd. All rows assume $r_0 = \$11$/bbl. See text for details.

depending on specification) than in the single-destination baseline model, reflecting the intuition discussed in section 2.3. When we require CBR to use 20-month ship-or-pay contracts, our estimates are substantially larger than in the baseline model: the estimated dK/dr_0 ranges from 46.0 to 65.3 mbpd per \$/bbl, depending on the upstream supply elasticity. These increased magnitudes reflect the fact that requiring CBR to use ship-or-pay contracts makes CBR more similar to pipeline transportation, from the point of view of shippers. Thus, the effective elasticity of substitution between these two technologies is large in this model, at least locally to $r_0 = \$11$/bbl. Recall that the ship-or-pay contracting framework for CBR used in this model is likely too restrictive for shippers relative to rail contracts that are used in practice. Thus, we view our results from the CBR contracting model as an upper bound, and the results from the baseline model as a lower bound, on the effects of increasing r_0, conditional on the assumed parameter values and constant returns to scale in pipeline

Figure 11: Counterfactual DAPL capacity vs rail cost intercept r_0.



Note: "Elasticity" refers to upstream crude oil supply elasticity. Calculations assume constant returns to scale in pipeline construction, r_0 = \$11/bbl, and r_1 = \$3/bbl per million bpd (mmbpd). Capacity is in thousands of barrels per day (mbpd). All calculations use the baseline model that allows rail to reach a single destination (Gulf Coast) and allows rail transport to proceed on a spot market. See text for details.

construction.

The bottom-most section of table 5 presents results that allow for increasing returns to scale in pipeline construction, using the baseline mode for CBR flows. Here, the values of dK/dr_0 and $dE[Q_r]/dr_0$ are amplified by 44–110% relative to the constant returns case, especially when the upstream supply elasticity for crude oil is high.

Finally, we explore counterfactuals in which rail becomes prohibitively expensive, even for large crude oil price spreads.[57] Figure 11 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. For each assumed supply elasticity, capacity K is concave in r_0, a consequence of the fact that as 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 0.4 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

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

520 mbpd to 858 mbpd. With a supply elasticity of 0.8, DAPL's capacity reaches 1167 mbpd. These extrapolations should be interpreted with caution because increases in K at large values of r_0 are governed by supply responses at very high oil prices, where the supply elasticity is likely lower than that within the span of the data (Smith and Lee, 2017).

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 53,000 bpd larger than its actual 520,000 bpd capacity. Under more aggressive but plausible assumptions on input parameters, the impact could have been greater than 130,000 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 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 framework that can be used to evaluate and intuitively understand how tradeoffs between these technologies 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.

Anderson, Soren T., Ryan Kellogg, and Stephen W. Salant, "Hotelling Under Pressure," *The Journal of Political Economy*, forthcoming.

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].

Averch, Harvey and Leland L. Johnson, "Behavior of the Firm Under Regulatory Constraint," *The American Economic Review*, 1962, *52* (5), 1052–1069.

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.

Bjørnland, Hilde C., Frode Martin Nordvik, and Maximilian Rohrer, "Supply Flexibility in the Shale Patch: Evidence from North Dakota," Working Paper, Norges Bank Research 2017.

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.

Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell, "nprobust: Nonparametric Kernel-Based Estimation and Robust Bias-Corrected Inference," Working Paper 2017.

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.

Cookenboo, Leslie Jr., *Crude Oil Pipelines and Competition in the Oil Industry*, Cambridge, MA: Harvard University Press, 1955.

Darlymple, Amy, "Sandpiper pipeline future unclear as Enbridge buys into Dakota Access," *DL-Online*, 2016.

Enbridge Energy Partners LP, "Enbridge to Invest $6.2 Billion in Light Oil Market Access Program," 2012. [Online; `http://www.marketwired.com/press-release/enbridge-to-invest-62-billion-in-light-oil-market-access-program-tsx-enb-1734840.htm`; accessed 18 March, 2018].

_ , "SEC 10-K Filing," February 2015. [Online; `https://www.enbridgepartners.com/Investor-Relations/EEP/Financial-Information/Filing-Details.aspx?filingId=10082267&docId=222781`; accessed 18 March, 2018].

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 LP, "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].

_ , "Energy Transfer Commences Binding Expansion Open Season for Bakken Pipeline Transport," 2014. [Online; `http://ir.energytransfer.com/phoenix.zhtml?c=106094&p=irol-newsArticle&ID=1969920`; accessed 18 March, 2018].

_ , "Energy Transfer Announces the Bakken Pipeline is in Service Transporting Domestic Crude Oil from the Bakken/Three Forks Production Areas," 2017. [Online; `http://ir.energytransfer.com/phoenix.zhtml?c=106094&p=irol-newsArticle&ID=2278014`; accessed 13 March, 2018].

_ , "Announcement of Expansion Open Season," 2018. [Online; https://www.energytransfer.com/ops_bakken.aspx; accessed 13 March, 2018].

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.

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.

Hyndman, Rob J and Qiwei Yao, "Nonparametric estimation and symmetry tests for conditional density functions," *Journal of nonparametric statistics*, 2002, *14* (3), 259–278.

Johnson, Arthur M., *Petroluem Pipelines and Public Policy, 1906–1959*, Cambridge, MA: Harvard University Press, 1967.

Kellogg, Ryan, "The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling," *American Economic Review*, 2014, *104* (6), 1698–1734.

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

Makholm, Jeff D., *The Political Economy of Pipelines: A Century of Comparative Institutional Development*, Chicago, IL: University of Chicago Press, 2012.

Marathon Petroleum Corporation, "Marathon Petroleum Corporation agrees to equity participation in Bakken Pipeline system," 2016. [Online; http://ir.marathonpetroleum.com/phoenix.zhtml?c=246631&p=irol-newsArticle_pf&ID=2192194; accessed 18 March, 2018].

Newell, Richard G. and Brian C. Prest, "The Unconventional Supply Boom: Aggregate Price Response from Microdata," Working Paper, National Bureau of Economic Research 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].

Phillips 66, "Phillips 66 Becomes Joint Venture Partner with Energy Transfer to Build Bakken Crude Oil Pipelines," 2014. [Online; `http://investor.phillips66.com/financial-information/news-releases/news-release-details/2014/Phillips-66-Becomes-Joint-Venture-Partner-with-Energy-Transfer-to-Build-Bakken-Crude-default.aspx`; accessed 18 March, 2018].

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

Shaffer, David, "Minnesota Regulators Order a Broader Look at Pipeline Route," *Minneapolis Star Tribune*, 2014.

Smiley, Gene, *The American Economy in the 20th Century*, South-Western Publishing Company, 1993.

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.

S&P Global Platts, "Methodology and Specifications Guide: Crude Oil," 2017. [Online; `https://www.platts.com/IM.Platts.Content/MethodologyReferences/MethodologySpecs/Crude-oil-methodology.pdf`; accessed 18 March, 2018].

Spavins, Thomas C., "The Regulation of Oil Pipelines," in Edward J. Mitchell, ed., *Oil Pipelines and Public Policy*, Washington, D.C.: American Enterprise Institute for Public Policy Research, 1979.

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

A Additional tables

Table 6: Expected margin for pipeline shippers: baseline model with alternative values for committed DAPL capacity K_d

Input parameters		Implied average cost per bbl			
Supply elasticity	r_1 $/bbl / mmbbl/d	$K_d = 320$	$K_d = 450$	$K_d = 520$	$K_d = 570$
0.4	3	$9.03	$7.10	$6.01	$5.25
0.6	3	$7.83	$6.33	$5.56	$5.04
0.8	3	$7.22	$6.03	$5.44	$5.04

Note: The actual DAPL tariff for ten-year committed shippers is between $5.50/bbl and $6.25/bbl. All rows assume $r_0 = \$11$/bbl. r_1 is in units of $/bbl per million bpd (mmmbpd). Baseline model allows rail to reach a single destination (Gulf Coast) and allows rail transport to proceed on a spot market. See text for details.

Table 7: Sensitivity of pipeline capacity and expected rail flow to the cost of crude-by-rail: baseline model with alternative values for committed DAPL capacity K_d

Input parameters		Impacts of a change in r_0 on K and $E[Q_r]$				
Supply elasticity	r_1	dK/dr_0	Elasticity of K w.r.t. r_0	$dE[Q_r]/dr_0$	Elasticity of $E[Q_r]$ w.r.t. r_0	Elasticity of $E[Q_r]$ w.r.t. r_0, K fixed
Baseline model, DAPL capacity = 320 mbbl/d						
0.4	3	50.3	0.51	-38.9	-1.74	-0.19
0.6	3	49.0	0.50	-32.9	-1.31	-0.20
0.8	3	52.7	0.54	-32.2	-1.11	-0.20
Baseline model, DAPL capacity = 450 mbbl/d						
0.4	3	34.5	0.31	-21.7	-1.43	-0.20
0.6	3	40.8	0.37	-22.6	-1.19	-0.20
0.8	3	47.7	0.43	-24.7	-1.04	-0.20
Baseline model, DAPL capacity = 520 mbbl/d (main results in paper)						
0.4	3	29.4	0.25	-15.8	-1.32	-0.20
0.6	3	37.7	0.32	-18.5	-1.14	-0.20
0.8	3	45.7	0.39	-21.4	-1.01	-0.20
Baseline model, DAPL capacity = 570 mbbl/d						
0.4	3	26.6	0.22	-12.6	-1.25	-0.20
0.6	3	36.0	0.30	-16.1	-1.11	-0.20
0.8	3	44.6	0.37	-19.4	-0.99	-0.20

Note: All rows assume constant returns to scale in pipeline construction and that $r_0 = \$11$/bbl. Capacity K and expected rail flows $E[Q_r]$ are in thousands of barrels per day (mbpd). r_1 is in units of $/bbl per mmbpd. Baseline model allows rail to reach a single destination (Gulf Coast) and allows rail transport to proceed on a spot market. See text for details.

B Equilibrium in the pipeline investment model when rail can flow to multiple destinations

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:

$$\int_{P_p(K)}^{\bar{P}} \left[\int_{P_d}^{P_r(K)} (P_d - P(K)) f(\tilde{P}|P_d) d\tilde{P} \right] f(P_d) dP_d. \tag{7}$$

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:

$$\int_{P_p(K)}^{\bar{P}} \left[\int_{P_r(K)}^{\bar{P}} (P_d - P(K + Q_r(\tilde{P}))) f(\tilde{P}|P_d) d\tilde{P} \right] f(P_d) dP_d. \tag{8}$$

Equilibrium capacity K is therefore given by equation (9), which collapses to the single-destination equilibrium equation (1) if $\tilde{P}$ is always equal to P_d.

$$\begin{aligned} \frac{C(K)}{K} = & \int_{P_p(K)}^{\bar{P}} \left[\int_{P_d}^{P_r(K)} (P_d - P(K)) f(\tilde{P}|P_d) d\tilde{P} \right] f(P_d) dP_d \\ & + \int_{P_p(K)}^{\bar{P}} \left[\int_{P_r(K)}^{\bar{P}} (P_d - P(K + Q_r(\tilde{P}))) f(\tilde{P}|P_d) d\tilde{P} \right] f(P_d) dP_d. \end{aligned} \tag{9}$$

For a given capacity K, each term on the right-hand side of equation (9) 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.

We now apply the implicit function theorem to determine dK/dr_0:[58]

$$\frac{dK}{dr_0} = \frac{\int_{P_p(K)}^{\bar{P}} \left[\int_{P_r(K)}^{\bar{P}} (1 - \frac{r_1}{P'(K+Q_r(\tilde{P}))+r_1}) f(\tilde{P}|P_d) d\tilde{P} \right] f(P_d) dP_d}{\frac{d(C(K)/K)}{dK} + \int_{P_p(K)}^{\bar{P}} \left[\int_{P_d}^{P_r(K)} P'(K) f(\tilde{P}|P_d) d\tilde{P} + \int_{P_r(K)}^{\bar{P}} \frac{r_1 P'(K+Q_r(\tilde{P}))}{P'(K+Q_r(\tilde{P}))+r_1} f(\tilde{P}|P_d) d\tilde{P} \right] f(P_d) dP_d} \tag{10}$$

The most important difference between equations (10) 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 (10) 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.[59]

C Derivation of the destination share model with rail contracting

Recall that the contract model is:

$$R_{i,t+1} = \underbrace{sR_{i,t}}_{\text{old flows still contracted}} + q_{i,t+1} \underbrace{\left(M_{t+1} - s \sum_j R_{j,t} \right)}_{\text{new flows available to be contracted}}$$

To turn this into an estimable expression, first define an accounting identity for $R_{0,t+1}$,

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

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

the quantity of "outside" flows (via pipeline to Cushing) at time $t+1$:

$$\begin{aligned}
R_{0,t+1} &= M_{t+1} - \sum_i R_{i,t+1} \\
&= M_{t+1} - \sum_i \left(sR_{i,t} + q_{i,t+1}(M_{t+1} - s\sum_j R_{j,t}) \right) \\
&= M_{t+1} - \sum_i \left(sR_{i,t} + q_{i,t+1}(M_{t+1} - s\left(M_t - R_{0,t}\right)) \right) \\
&= M_{t+1}\left(1 - \sum_i q_{i,t+1}\right) - s\sum_i \left(R_{i,t} - q_{i,t+1}\left(M_t - R_{0,t}\right)\right) \\
&= M_{t+1}\left(1 - \sum_i q_{i,t+1}\right) - s\sum_i R_{i,t} + s\left(M_t - R_{0,t}\right)\sum_i q_{i,t+1} \\
&= M_{t+1}\left(1 - \sum_i q_{i,t+1}\right) - s\left(M_t - R_{0,t}\right) + s\left(M_t - R_{0,t}\right)\sum_i q_{i,t+1} \\
&= M_{t+1}\left(1 - \sum_i q_{i,t+1}\right) - s\left(M_t - R_{0,t}\right)\left(1 - \sum_i q_{i,t+1}\right) \\
&= \left(M_{t+1} - s\left(M_t - R_{0,t}\right)\right)\left(1 - \sum_i q_{i,t+1}\right)
\end{aligned}$$

This expression relates outside flows in period $t+1$, $R_{0,t+1}$, to "available" production and the share of it that is not contracted.

Also recall that we have assumed shares are given by a logit random utility model: $q_{it} = \frac{\exp(X_{it}\beta + \delta_i)}{1+\sum_j \exp(X_{jt}+\delta_j\beta)}$. By the standard properties of these models, the outside share in the expression for outside flows above is $(1 - \sum_j q_{j,t+1}) = \frac{q_{i,t+1}}{\exp(X_{i,t+1}\beta+\delta_i)}$ for any choice i.

We can use these relationships to write an estimable version of the contracting model:

$$\begin{aligned}
R_{i,t+1} &= sR_{i,t} + q_{i,t+1}(M_{t+1} - s\sum_j R_{j,t}) \\
&= sR_{i,t} + q_{i,t+1}(M_{t+1} - s\left(M_t - R_{0,t}\right)) \\
&= sR_{i,t} + q_{i,t+1}\frac{R_{0,t+1}}{1 - \sum_j q_{j,t+1}} \\
&= sR_{i,t} + R_{0,t+1}\exp\left(X_{i,t+1}\beta + \delta_i\right)
\end{aligned}$$