Crude by Rail, Option Value, and Pipeline Investment

Thomas R. Covert
Ryan Kellogg*

December 8, 2018

Abstract

The U.S. shale boom has profoundly increased crude oil movements by both pipelines—the traditional mode of transportation—and railroads. This paper develops a model of how pipeline investment and railroad use are determined in equilibrium, emphasizing how railroads' flexibility allows them to compete with pipelines. We show that policies that address crude-by-rail's environmental externalities by increasing its costs should lead to large increases in pipeline investment and substitution of oil flows from rail to pipe. Similarly, we find that policies enjoining pipeline construction would cause 80–90% of the displaced oil to flow by rail instead.

*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 NBER IO Summer Institute, 2018 Winter Marketing-Economics Summit, Iowa State University, Montana State University, Oberlin College, Pitt-CMU, Wharton, UBC, 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

Since the mid 2000s, U.S. shale oil production has increased from essentially nothing to more than 6 million barrels per day (bpd) in 2018.[1] At present, shale formations are responsible for more than half of all U.S. production, vaulting the United States past Russia and Saudi Arabia to become the largest oil-producing nation (Energy Information Administration, 2018). Because the primary shale oil plays—the Eagle Ford, the Permian, and especially the Bakken—are located in areas far from refinery infrastructure or coastal export facilities, crude oil transportation has emerged as a critical issue for the industry. In practice, shale oil has moved not only by pipeline—the traditional mode for nearly all long-distance oil transmission since at least WWII (Johnson 1967, p. 327)—but also by railroad. The volumes and economic stakes are significant. For instance, more than 90% of the Bakken Shale's current production of 1.2 million bpd (worth more than $80 million per day) is transported by pipeline and rail (North Dakota Pipeline Authority 2018).

At the same time, both pipeline and rail transportation of crude oil have attracted substantial controversy and opposition. The Keystone XL and Dakota Access Pipelines have been the subject of prominent public protests. Neither pipeline was permitted under the Obama administration, and under the Trump administration only Dakota Access has been constructed. Crude-by-rail has meanwhile resulted in several high-profile accidents that have resulted in not only spills but also fatalities (such as the 2013 derailment and explosion of a Bakken crude oil train in Lac Megantic, Quebec, that killed 47 people).

The primary goals of this paper are to understand the economic forces driving crude oil shippers' transportation choices and to quantify how investment in pipeline infrastructure substitutes with use of crude-by-rail.[2] Despite the controversy, externalities, and private economic value at stake, we are not aware of other work that addresses these questions. Indeed, the recent rise of crude-by-rail is a puzzle, given its high private cost per barrel relative to the amortized cost per barrel for pipelines and its provision by a rail industry that is widely believed to exercise market power in access pricing (Busse and Keohane 2007, Hughes 2011, Preonas 2017). One common explanation for the rise of crude-by-rail is simply that the rapid rise in shale oil production outpaced the speed at which new pipeline capacity could be built, so that producers effectively had no choice but to turn to the railroads. In this paper, we instead demonstrate that these two technologies should co-exist in equilibrium and study how they substitute with one another, emphasizing how the flexibility that railroads

[1] Up-to-date shale play-level production data are available from the EIA at https://www.eia.gov/energyexplained/data/U.S.%20tight%20oil%20production.xlsx.

[2] 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.

provide to crude oil shippers allows railroads to successfully compete with pipelines despite their higher costs. 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 not to ship. Pipeline shippers do not have this ability, since pipeline construction costs are entirely sunk and require up-front commitment from shippers. 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 rail transport (Lefebvre, May 23, 2013).

Our interest in the rise of crude-by-rail is motivated in part by its substantial unpriced externalities. Clay, Jha, Muller and Walsh (2017) estimates that external damages associated with railroad transportation of crude oil from the Bakken to the Gulf Coast are \$2.02 per barrel—roughly 20% of the private transportation cost—owing primarily to freight locomotives' NO_x emissions.[3] Overall, Clay et al. (2017) estimates that air pollution damages from rail movements are nearly twice those from pipeline transportation and are also much larger than damages from spills and accidents. These estimates raise questions about how policies intended to reduce these externalities, by raising the cost of rail transport, would affect investments in pipeline capacity and the volume of oil flow that would substitute away from rail to pipelines.[4] They also raise the question of how, conversely, regulatory or political foreclosure of pipeline construction would impact crude-by-rail flows, exacerbating rail's externalities. One goal of our paper is to provide quantitative answers to these questions.

We are further motivated by our belief that understanding substitution between pipeline and rail transportation is interesting in its own right, both because the rise of crude-by-rail is one of the most significant developments in the U.S. oil industry in decades, and because the model we develop captures intuition that is applicable to other settings in which a low-cost but inflexible investment can substitute for a high-cost, flexible one.[5] For instance, the

[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, June 30, 2008. Vol 73, No. 126, p. 37096.).

[5]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 grew steadily between WWII and 1970, a number of factors in that era favored pipelines over rail.

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), baseload electric power stations (which substitute with more flexible gas-fired peaker units), and zero marginal cost (but non-dispatchable) renewable power sources.[6] These tradeoffs between cost and flexibility also have a parallel in the finance literature that examines the relative returns of investments in illiquid versus liquid assets (see, for instance, Amihud and Mendelson 1986 and Pástor and Stambaugh 2003).

To quantify how changes in the costs and benefits of crude-by-rail impact pipeline investment, and how barriers to pipeline investment affect rail flows, we develop a model in which crude oil shippers can use pipelines or rail to physically arbitrage oil price differences between an upstream supply source and downstream markets, where the oil price is stochastic.[7] Pipeline transportation has large fixed costs with potentially significant 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

---

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). Finally, rail service prior to the 1980 Staggers Act was tightly federally regulated, with generally higher rates and lower-quality service than is the case today (Burton 1993, Winston 2005).

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

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

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 do) after they observe the realized downstream price. This flexibility generates option value, which is further enhanced by the ability of railroads to reach multiple destinations, not just the destination served by the pipeline. A key insight from 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 oil 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 marginal return to pipeline investment is 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 elasticity of upstream oil supply, and the cost of pipeline investment.

We use our model to quantify the impacts of crude oil transportation policies on pipeline and rail use, using the Dakota Access Pipeline (DAPL) as a case study. We calibrate the model using a variety of sources to match the economic conditions prevailing in June, 2014, when DAPL received firm commitments from its shippers (the line was completed in June, 2017, with a capacity of 520,000 bpd). 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 modestly with shipping volumes, consistent with the presence of scarcity rents or market power in these markets. 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, Newell and Prest 2017, and Smith and Lee 2017).

After calibrating our model using these inputs, we validate it 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), depending on the specification used and on the assumed parameters. This result gives us confidence that our model, stylized as it may be, captures the relevant economic forces that determine crude oil transportation choices.

In light of the unpriced externalities associated with crude-by-rail, we first use our model to ask how much larger DAPL would be in a world where rail shippers were forced to internalize these externalities. We find that a \$2 per barrel increase in the cost of rail transportation, consistent with the externalities estimated in Clay et al. (2017), results in an increase in equilibrium pipeline capacity of between 64,000 and 150,000 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 flows to rail per-barrel costs ranges from -0.9 to -2.2 across our specifications. In contrast, this elasticity takes a value of only -0.2 when we hold pipeline capacity fixed. Thus, policies that increase the cost of rail transport—such as regulations targeting rail's environmental externalities—may increase pipeline investment and induce economically significant long-run substitution from railroads to pipelines.

Next, we ask how much more rail transportation would be used in a world with stricter pipeline regulation. We find that, were construction of the 520,000 bpd DAPL prevented, crude-by-rail flows would increase by 303,000 to 417,000 bpd, depending on the specification. These increases in rail flows are between 82% and 91% of the decreases in pipeline flows.[8] Thus, we conclude that foreclosure of pipeline construction, in the presence of a rail option, leads to a substantial increase in rail use, even though the substitution is not perfectly one-for-one.

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

[8] Because the pipeline does not always flow at capacity, pipeline flows in our model decrease by 369,000 to 459,000 bpd rather than the full 520,000 bpd.

model calibration. Section 4 follows with a discussion of the empirical relationship between rail flows and oil price differentials, and it describes how we use this relationship 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. Section 6 then presents our results on the effects of crude oil transportation policies. 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 or rely on the railroads.

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

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 4.2 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 $S(Q)$ 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. $S'(Q) > 0$. (For brevity, we henceforth refer to $S(Q)$ as the supply curve rather than "inverse net supply".) Let $P_u = S(Q)$ 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 $[\underline{P}, \overline{P}]$.[9]

- 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 pipeline's regulated rate of return). Given capacity, the marginal cost of shipping crude on the pipeline up to the capacity constraint is zero.[10]

- 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 $r'(Q_r) \geq 0$.

[9] In some specifications we implement, we also allow for uncertainty in upstream supply. See section 3.5.

[10] 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.

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 $P_p(K) = S(K)$ denote the minimum downstream price such that the pipeline is full.

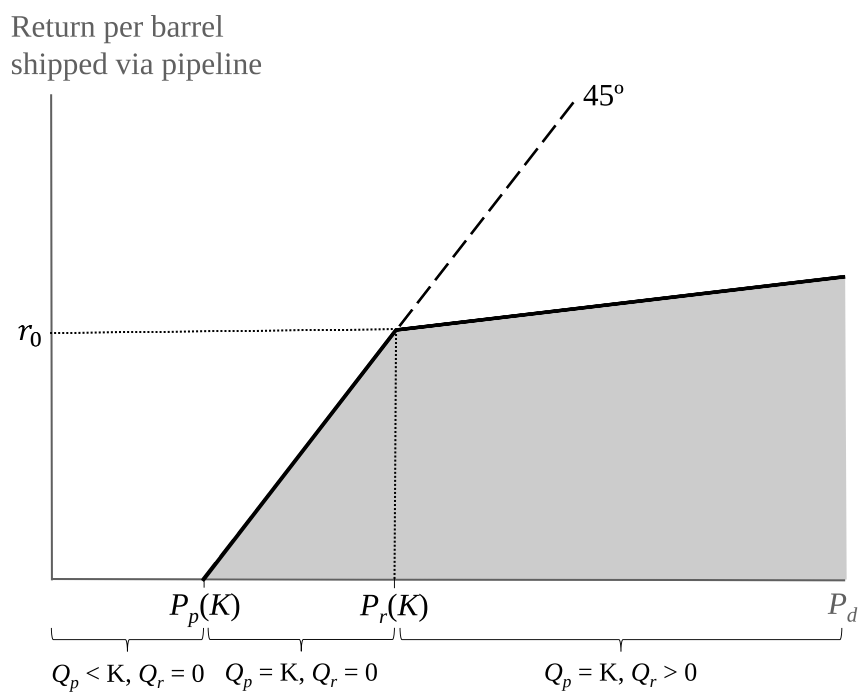
For downstream prices $P_d > P_p(K)$, no more oil can flow through the pipeline, but rail may be used. Crude oil volumes will move over the railroad only to the extent that the differential between P_d and P_u covers the marginal cost $r(Q_r)$ of railroad transport. Define $P_r(K)$ as the minimum downstream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

2.2 Equilibrium pipeline capacity in the single destination model

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

Figure 1: Expected return achieved by pipeline shippers



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

figure 1, iff $r'(Q_r) > 0$.[11]

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's average 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

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

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

$$\frac{C(K)}{K} = \int_{P_p(K)}^{P_r(K)} (P_d - P_p(K))dF(P_d) + \int_{P_r(K)}^{\overline{P}} r(Q_r(P_d))dF(P_d). \tag{1}$$

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

Our model illuminates the comparative static of primary interest in this paper: how does pipeline capacity investment respond to changes in the cost of rail transportation? Figure 2 provides the intuition. Consider a pipeline project that, facing a railroad cost curve with intercept r_0, would attract equilibrium commitments from shippers for a capacity of K. Now suppose that the rail cost intercept were instead $r_0' > r_0$. This increase in rail transportation cost increases the basis differential realized by pipeline shippers whenever rail transportation is used, as indicated in the upper, striped area. This increase in expected return then increases shippers' willingness to commit to capacity, so that equilibrium pipeline capacity must increase. The new capacity level K' balances the increase in expected return when rail is used against the decrease in expected return caused by the reduced 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).[12] 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:[13]

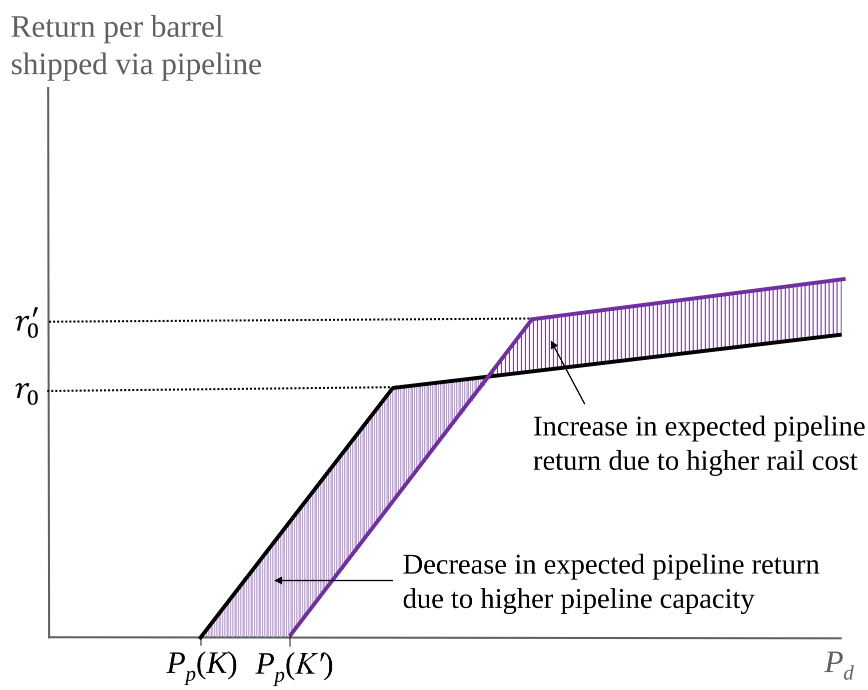
$$\frac{dK}{dr_0} = \frac{1 - F(P_r(K)) - \int_{P_r(K)}^{\overline{P}} \frac{r_1}{P'(K+Q_r(P_d))+r_1} dF(P_d)}{\frac{d}{dK}\frac{C(K)}{K} + \int_{P_p(K)}^{P_r(K)} S'(K)dF(P_d) + \int_{P_r(K)}^{\overline{P}} r_1 \frac{S'(K+Q_r(P_d))}{S'(K+Q_r(P_d))+r_1} dF(P_d)}. \tag{2}$$

In the simple case in which $r_1 = 0$ and $C(K)$ exhibits constant returns to scale, this

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

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

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



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

expression reduces to:

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

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.

When $r_1 > 0$, the terms involving r_1 reduce the numerator of equation (2) and increase the denominator, so that the sensitivity of K to r_0 is reduced. Intuitively, large values of r_1 reduce the option value of rail transportation, so that pipeline economics are then less sensitive to shocks to railroad transportation costs. Finally, when there are increasing returns to scale, so that $d(C(K)/K)/dK < 0$, dK/dr_0 will be larger than in the case of constant returns.

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 (or elasticity) of the upstream crude oil supply curve $S(Q)$.

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 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. Our calculation of the expected return to pipeline shippers is then the average of the expected 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). Appendix E provides details on the computation of the model.

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, $\tilde{P} \geq P_d$.

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

- 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 D derives 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, 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 an estimate of $F(\widetilde{P} \mid P_d)$. We discuss the estimation of this distribution in section 3.2.3. Computational details are provided in appendix E.2.

3 Data and calibration

To quantify the amount of substitution between pipelines and railroads, we calibrate our model to the recently constructed Dakota Access Pipeline (DAPL). This section summarizes the key facts on DAPL's construction and our procedures for estimating the distribution of future downstream prices anticipated by shippers, the cost function for crude-by-rail, and the upstream oil supply elasticity. Additional detail is provided in appendix B.

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 \$112/bbl and expected to remain high: the three-year Brent futures price was \$99/bbl.[15] Per Biracree (October 18, 2016) and North Dakota

[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 beyond a three year horizon.

Pipeline Authority (2017), existing local refining capacity was (and remains) 88 thousand bpd (mbpd), and other Bakken export pipeline capacity was 763 mbpd.[16]

DAPL was put into service in June, 2017 with a capacity of 520 mbpd and a reported construction cost of $4.78 billion (Energy Transfer Partners LP, 2017).[17] 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 shippers' commitments 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 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 September 12, 2014, Marathon Petroleum Corporation 2016, and Darlymple August 7, 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).[18] As sensitivities, we will also examine results based on assumed capacities of 320, 450, and 570 mbpd.[19]

3.2 Crude oil prices

3.2.1 Price data

We obtained data on spot market crude oil prices from Bloomberg and Platts.[20] 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

[16]This figure includes a planned 24 mbpd expansion of the Double H pipeline that was completed in 2016.

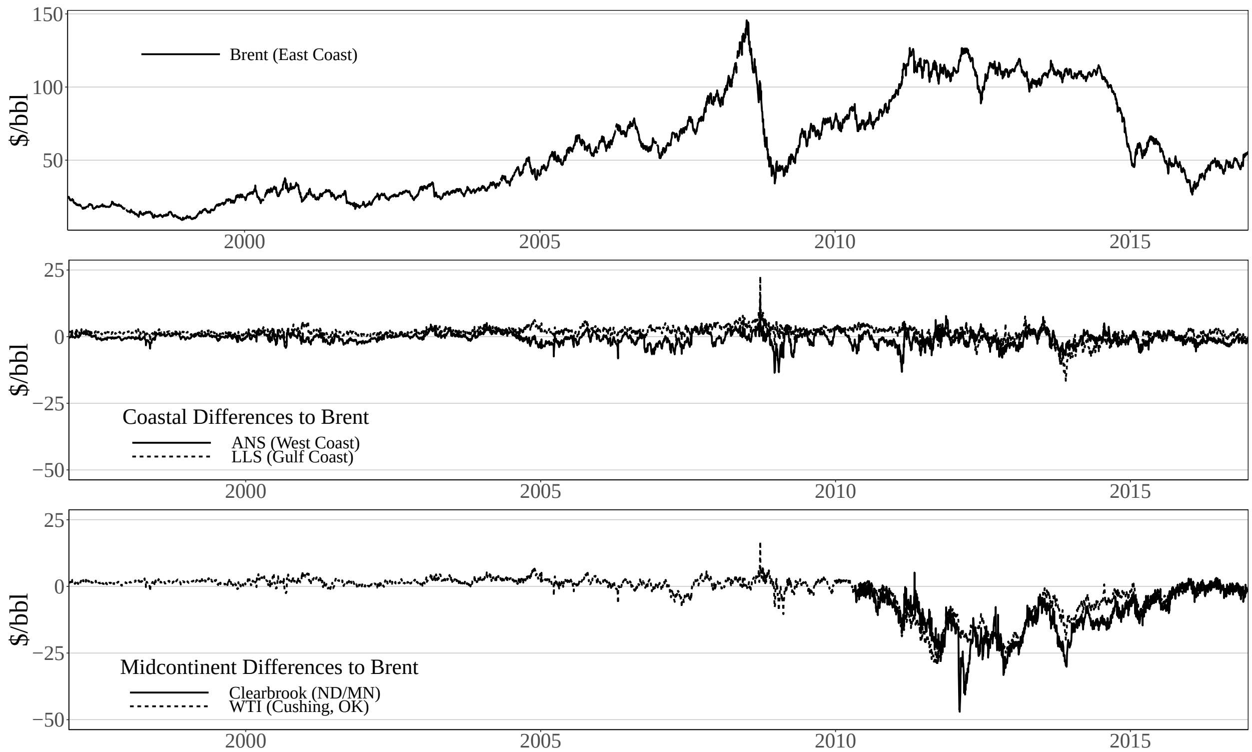
[17]Construction cost includes $1 billion for the Energy Transfer Crude Oil Pipeline (ETCO) from Patoka, IL to Nederland, TX on the Gulf Coast.

[18]The increase in DAPL capacity from 450 to 520 mbpd occurred following a supplemental open season held in early 2017 (Energy Transfer Partners LP, 2017).

[19]DAPL was built with the ability to expand to 570 mbpd, and it initiated an open season on the incremental 50 mbpd in March, 2018 (Darlymple August 7, 2016 and Energy Transfer Partners LP 2018).

[20]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 Platts starts on May 4, 2010, five months earlier than the series from Bloomberg.

Figure 3: Crude oil spot prices



Note: The top panel shows the spot price for Brent crude delivered to New York harbor. The middle panel shows the difference between prices at coastal markets (ANS, LLS) and Brent. The bottom panel shows the difference between prices at mid-continent markets (Clearbrook, WTI) and Brent. All three panels show daily data between Janary, 1997 and December, 2016.

for U.S. East Coast, Gulf Coast, and West Coast "downstream" destinations, respectively.[21] We also use the price of West Texas Intermediate (WTI) at the Cushing, OK pipeline and storage hub as another destination.

Figure 3 plots these spot price data, aggregated to the monthly level, over 1997–2016. The top panel shows the time series of the benchmark Brent price and illustrates the substantial oil price decrease that occurred during the second half of 2014. The middle panel of figure 3 shows that prices at the three coastal destinations are tightly correlated, typically differing by no more than a few $/bbl over the last 20 years. Moreover, no single destination has

[21] Because Bakken crude oil is quite light relative to ANS, the ANS benchmark may understate the value of Bakken crude on the West Coast (ANS crude is 32 API and 0.96% sulfur, and Bakken is 43.3 API and 0.07% sulfur (S&P Global Platts, 2017)). Because LLS and Heavy Louisiana Sweet (HLS) are comparable to Bakken and ANS, respectively (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)), we add the historic average price difference between LLS and HLS, equal to $0.53/bbl, to the ANS price series in all of our calculations. We calculate this price difference over January 6, 1988 (the first date at which both prices are available from Bloomberg) through December 31, 2016.

a consistent price advantage over another. In contrast, the bottom panel of figure 3 shows that the prices at Clearbrook and Cushing were substantially discounted relative to coastal destinations from 2011–2014.

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 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/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 quite close (figure 3).[22]

We use three approaches to estimate the variance of $P_{d,t}$ for each month t of shippers' commitment, where t ranges from 37 months (the start of pipeline service) to 156 months (13 years; i.e., contract expiration). Estimates from each of these approaches are presented in figure 7 in appendix B.2. In our baseline specification, we assume that shippers' expected price volatility over a t-month horizon is given by the historic Brent oil price volatility over a t-month horizon.[23] We estimate that historic volatility increases substantially over the 37 month to 13 year horizon, from 45% to 129%. Our second method for estimating future price volatility uses implied volatility from crude oil futures options markets, as in Kellogg (2014). Implied volatility in June, 2014 (when DAPL firm contracts were signed) was low relative to the historical average, so these volatility estimates are lower than in our baseline specification. We interpret them as plausible lower bounds, since they assume that volatility will not eventually revert to its long-run mean. Third, and finally, we take the historic one-month horizon Brent volatility from our baseline approach and extrapolate it to longer time horizons under the assumption that oil prices follow a random walk. The volatilities from this approach are larger than our baseline volatilities, and we interpret them as a plausible upper bound.

[22]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).

[23]Volatility at each horizon t is calculated by taking the standard deviation of t-month differences in the logged Brent front-month price, exponentiating, subtracting 1, and multiplying by 100.

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. Though the model from section 2.3 assumes that shipments to all rail destinations face a common cost r_0, our cost data, discussed in section 3.4, 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 $D = \widetilde{P} - (P_d - r_{0,d})$ as the amount by which the best rail netback exceeds a rail netback to LLS (the pipeline terminus).

Appendix B.2.2, figure 8 presents our estimates of the density $f(D \mid P_d)$, which includes a point mass at $D = 0$ (which occurs whenever the best netback is to LLS). Consistent with the raw price series presented in figure 3, our estimate of $f(D \mid P_d)$ implies that D is often zero and rarely takes on a value exceeding a few \$/bbl. For instance, at $P_d = \$99$/bbl, our estimate of $E[D \mid P_d = 99]$ is \$2.15/bbl.

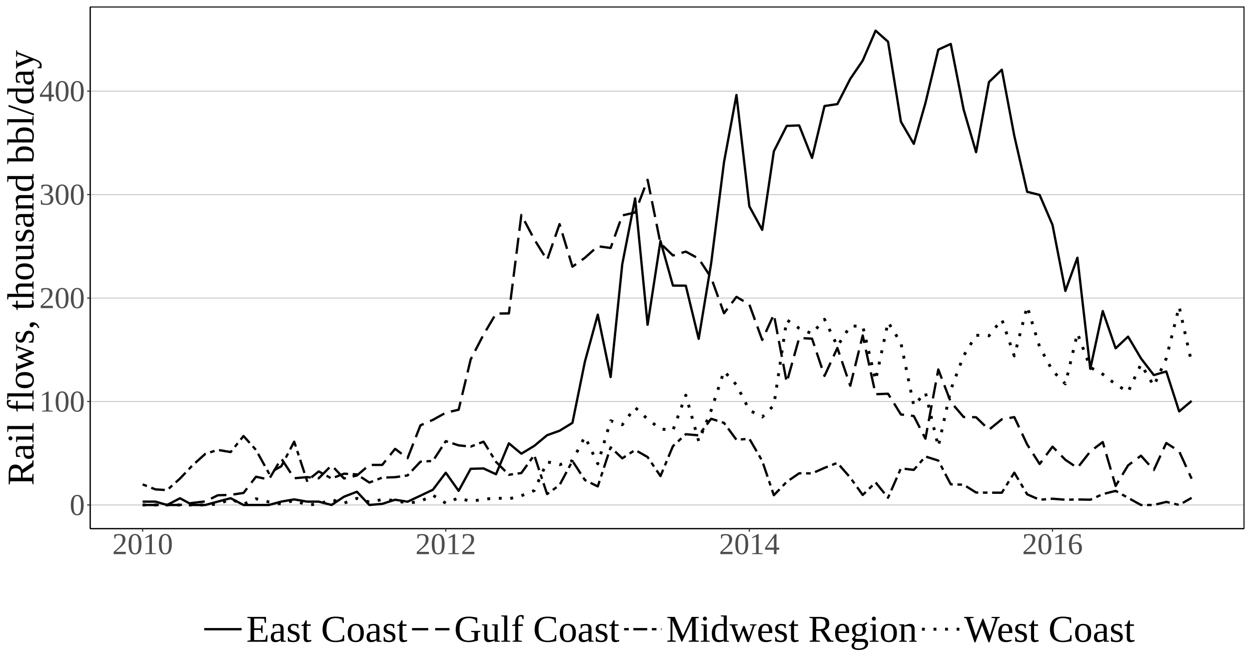
3.3 Crude-by-rail flows

We obtained data on monthly regional crude-by-rail flows from the EIA.[24] Figure 4 presents data on crude oil rail shipments from the Midwest region (which includes North Dakota and the Cushing, OK WTI pricing hub) to the East Coast, Gulf Coast, and West Coast, as well as within-Midwest shipments.[25] Volumes are dominated by shipments to the coasts rather than intra-Midwest shipments, in accordance 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 fall substantially. The rise and fall of crude-by-rail is consistent with the rise and fall in spatial price differentials shown in the third panel of figure 3, 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 rail contracts.

[24]The EIA's crude-by-rail data are available online at `https://www.eia.gov/dnav/pet/PET_MOVE_RAILNA_A_EPC0_RAIL_MBBL_M.htm`.

[25]Regions are defined as Petroleum Administration for Defense Districts (PADDs). A map of PADDs is presented in figure 9 in appendix B.3.

Figure 4: Crude-by-rail monthly volumes shipped from the Midwest region



Source: EIA. Regions are defined as Petroleum Administration for Defense Districts (PADDs); see text and appendix B.3.

3.4 Railroad transportation costs

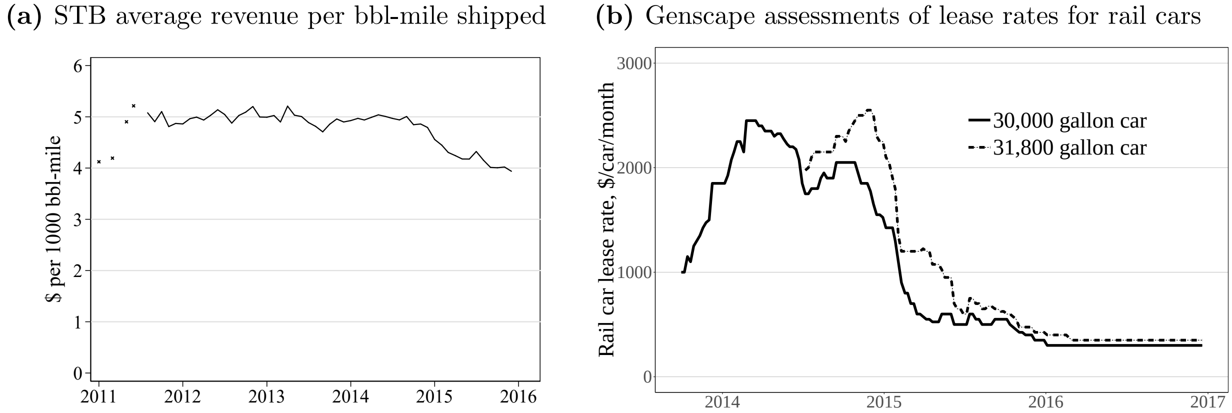
3.4.1 Railroad cost data

We obtained data on the rates charged by railroads to transport crude oil from the Surface Transportation Board (STB). The STB is the United States's regulator of interstate railroads, and we obtained STB's restricted-access waybill sample for 2009–2015. For each month of our sample, we calculate the total revenue (across all shipments originating in the EIA's Midwest region) earned by railroad carriers and divide by the total number of bbl-miles of crude oil shipped. Figure 5, panel (a), presents the resulting time series of average revenue per bbl-mile.[26] 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.[27] This decrease in transportation rates follows the sharp drop in

[26]To protect the confidentiality of individual carriers' rates, we are unable to present results that are disaggregated (either spatially or temporally) beyond the region-month level. We are also unable to present data prior to 2011.

[27]The 2015 rate decrease does not appear to merely reflect a compositional change in shipments: figure 10 in appendix B.4 shows that it persists after controlling for destination, distance travelled, the number of carrying railroads, and the number of cars.

Figure 5: Crude-by-rail freight and tank car lease costs



Note: STB data in panel (a) cover sampled waybills originating in the EIA's Midwest region and terminating in the EIA's East Coast, Gulf Coast, West Coast, and Midwest regions. 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): Genscape's rail car lease rates are not region-specific.

crude oil prices that began in late 2014 (figure 3) and coincides with a decrease in crude-by-rail volumes (figure 4). We calculate that roughly \$0.50 (i.e., half) of the rate decrease can be attributed to a direct reduction in locomotive fuel costs; the remainder is due to either reduced congestion or a decrease in markups.[28]

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% 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.

We also obtained data from Genscape, a private industry intelligence firm, on the cost of leasing rail cars (which are provided by third parties, not the railroads themselves) and the costs of 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

[28]We obtain this \$0.50 per thousand bbl-mile figure by combining the roughly \$50 per barrel oil price decrease during 2014–2015 (assuming complete pass-through to diesel prices) with information on locomotives' fuel use per bbl-mile. Specifically, we use: (1) the calculation from Clay et al. (2017) that crude-by-rail emits 4.578 short tons of CO_2 per million bbl-miles; and (2) Environmental Protection Agency (2009) stating that combustion of one gallon of diesel results in emission of 10,217 grams of CO_2.

cost of making a spot crude shipment to a particular destination. The Genscape data series begins in October, 2013.

Panel (b) of figure 5 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.[29]

3.4.2 Calibration of the crude-by-rail cost function

We estimate $r_{0,i}$ for each destination i using Genscape's minimum assessed "all-in" rail cost—including rail freight, tank car leasing, and terminal fees—for transportation between the Bakken and i. For most destinations, this minimum cost occurs in the spring of 2016, after: (a) millions of bbls/day of loading and unloading capacity had been constructed; and (b) both rail car lease prices and railroad freight rates had fallen substantially. We select the minimum reported costs 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. Our estimates of $r_{0,i}$ are then: \$13.00/bbl for shipments to the East Coast, \$10.94 for the Gulf Coast, \$9.23 for the West Coast, and \$8.54 for within-Midwest shipments to Cushing, OK.

We estimate r_1 by projecting the two sources of crude-by-rail costs for which we have credible time series, STB freight rates and Genscape tanker car lease rates, onto contemporaneous aggregate rail flows, measured with either the STB waybill sample or the EIA's rail flow estimates. The results, shown in table 10 in appendix B.4, are consistent with an upward-sloping rail services supply curve, though confidence intervals admit a wide range of total (freight services plus tanker car leasing) r_1 values.

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 exceed 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 bpd (mmbpd), consistent with the sum of the point estimates in the freight rate and tanker car lease rate regressions when the regressor is STB rail flows. These high costs, however, likely reflect short-run constraints in rail terminals and tank cars during 2013–2016 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.

[29] See Tita (May 29, 2014) and Arno (2015) for discussions of the boom and bust in railcar lease rates.

3.5 Upstream oil supply

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 estimates of a dynamic model of Bakken drilling and production do not exist in the literature and are beyond the scope of this paper. Instead, we calibrate our model's crude oil supply curve using a range of elasticities that we obtain from previous work on the supply of U.S. shale oil and gas.

Because changes in oil supply come from the drilling of new wells rather than from changes in production from existing wells (Anderson, Kellogg and Salant 2018), we use estimates of the drilling elasticity of U.S. shale wells. Newell and Prest (2017) finds an elasticity of 1.6 for all U.S. shale oil, and Hausman and Kellogg (2015) and Newell et al. (2016) estimate price elasticities of 0.8 and 0.7, respectively, for U.S. shale gas.[30] However, drilling elasticities likely over-estimate the impact of oil prices on upstream production over the 10-year pipeline contract 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 smaller than the percentage change in drilling. 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. Therefore, we roughly halve these drilling elasticity estimates and use a range of Bakken supply elasticities between 0.4 and 0.8.[31]

To calibrate the supply curve intercept, we use the North Dakota Pipeline Authority (NDPA) expected production forecast from April, 2014 (North Dakota Pipeline Authority, 2014). This forecast provides monthly expected production volumes throughout the ten-year pipeline contract; average expected production during the ten-year period is 1,616 thousand bpd (mbpd). In an alternative set of specifications, we allow for uncertainty in upstream supply by letting the supply intercept be stochastic, using a conservative production forecast from the NDPA that is, on average, about 200 mbpd lower than the expected production path. We construct an "optimistic" forecast that is symmetric to this conservative forecast and, in alternative specifications, run a version of our model in which prospective shippers face a stochastic supply intercept (in addition to the stochastic downstream price), with a probability of 1/3 assigned to each production path.[32]

[30] All three of these papers use instrumental variables regressions of drilling rates on lagged prices.

[31] Use of an elasticity higher than 0.8 would cause our model to estimate responses of pipeline capacity to rail costs that are even larger than those presented in section 6. In addition, we calculate that a gross minimum for the Bakken supply elasticity is about 0.3, based on 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, relative to an oil price that roughly halved (data accessed on March 18, 2018 from EIA at `https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=MCRFPND2&f=M`).

[32] These probabilities are arbitrary; the NDPA does not assign probabilities to its production cases.

3.6 Pipeline economies of scale

Finally, we consider the extent to which there are increasing returns to scale in pipeline construction; that is, we calibrate the $d(C(K)/K)/dK$ term in equation (10). As a conservative baseline, we assume that the construction of DAPL has locally 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 use an elasticity estimate of 0.59 derived from the Soligo and Jaffe (1998) study of Caspian Basin oil export pipelines.[33]

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 downstream price, net of transportation costs; and (b) fluidly adjust the magnitude of these shipments as downstream prices rise and fall. In practice, however, rail shipments violate both of these assumptions: figure 4 shows that every destination has positive rail flows in every month, starting in 2012, and figures 4 and 3 together suggest that rail volumes follow price movements with a lag.

A likely driver of these violations of 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. Though we treat crude-by-rail costs as entirely marginal in the baseline model, these contracts were presumably necessary to finance investments in loading facilities in North Dakota and unloading facilities in refining regions.[34] 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 test the hypothesis that contracts are important, by empirically examining the response of rail flows to present and past price differentials. We then develop and alternative version of our pipeline investment model that requires rail shipments to use contracts rather than operate on a spot basis.

[33] We were not able to find a study of scale economies in U.S. oil pipelines. We obtain the estimated elasticity of 0.59 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.

[34] These investments are reported to be substantially cheaper, per unit of capacity, than pipeline investments. For example, at the low end, RBN Energy (2018) documents that the Plains All American loading facility in Manitou, ND cost \$40 million, with 65,000 bpd of capacity, or roughly \$600 per bpd of capacity. At the higher end, Area Development News Desk (2018) documents that Enbridge spent \$145 million on an 80,000 bpd facility, or \$1,8125 per bpd of capacity. Both of these figures lie substantially below the \$9,200 per bpd of capacity that ETP spent on DAPL.

4.1 Empirical relationship between rail flows and crude prices

We correlate 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 the Midwest region. 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;[35] and transported by rail to the East, West and Gulf coasts, as well as rail transportation within the Midwest region.[36]

We assume shippers to the East coast earn the Brent price for their cargoes, shippers within the Midwest region earn WTI, shippers to the Gulf coast earn LLS, and shippers to the West coast 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. We assume that infinitesimal shippers to destination j in month t get indirect utility from a linear combination of current and lagged prices $p_{j,t-l}$, a fixed effect δ_j, a time-varying unobserved mean utility shock $\xi_{j,t}$ specific to destination j, and an *i.i.d.* type-1 extreme value "taste" shock $\epsilon_{i,j,t}$ specific to the shipper i, destination j, and month t. 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 $s_{j,t}$ with current and lagged 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 3 to 24 months. In general, the coefficient on the oldest price realization in

[35] 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. 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.

[36] Because refined consumption is empirically at or just below the reported capacity for refineries in North Dakota during the entire time period, we subtract it from total production and focus on the destinations that appear to be unconstrained.

Table 1: Multinomial logit shipment destination share regressions

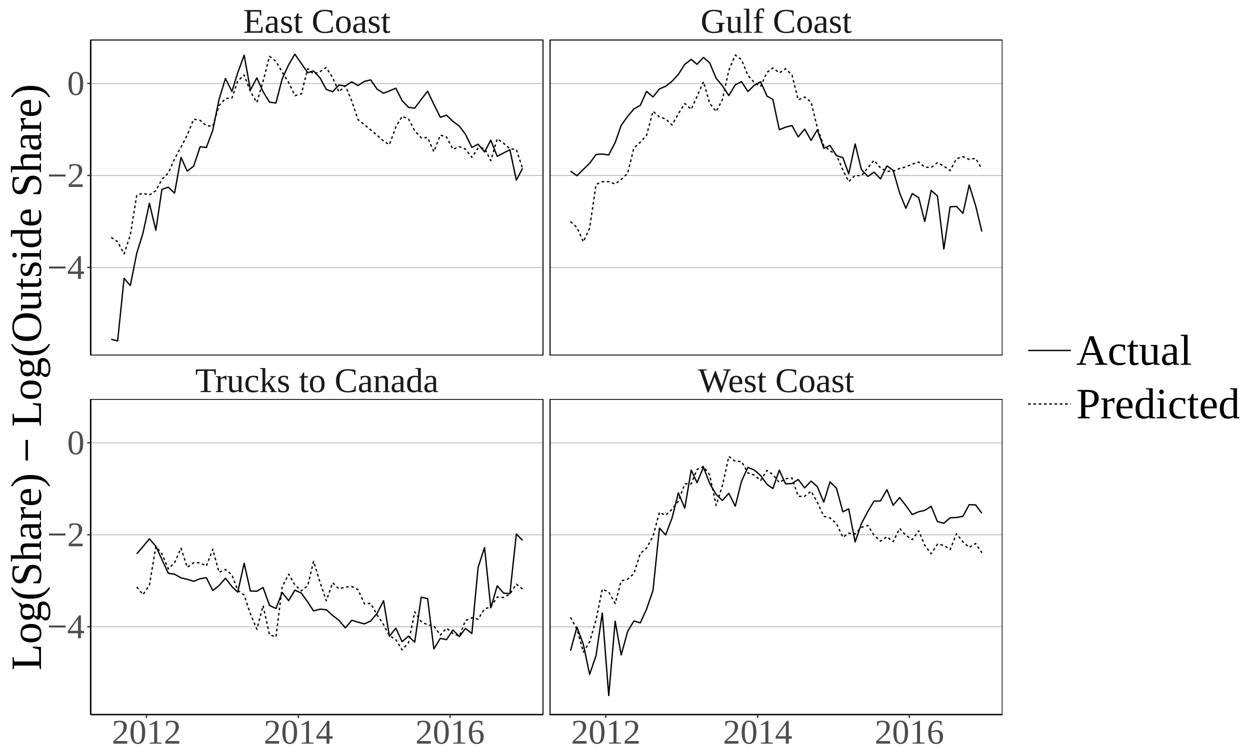
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
P_t	−0.00	−0.07	−0.08	−0.07	−0.09	−0.08	−0.06	−0.03	−0.01
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
P_{t-3}		0.09	0.02	0.00	0.02	−0.00	−0.01	−0.01	−0.01
		(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
P_{t-6}			0.09	0.02	0.01	0.03	0.01	−0.00	−0.01
			(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
P_{t-9}				0.09	0.01	0.00	0.02	0.01	0.00
				(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
P_{t-12}					0.10	0.03	0.02	0.03	0.02
					(0.01)	(0.01)	(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)
N	400	397	386	371	356	341	326	311	296
R^2	0.15	0.21	0.30	0.40	0.52	0.60	0.65	0.68	0.68
Within R^2	0.00	0.07	0.14	0.23	0.34	0.41	0.46	0.43	0.37

Note: Multinomial logit regressions of monthly shipment destination shares onto current and lagged monthly destination prices. Pipeline shipments to Cushing, OK, 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, with a maximum lag of 4 months. The within R^2 values are the squared correlation of the destination-demeaned outcomes and their predicted values.

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.

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.

Figure 6: Actual vs predicted rail shares under 18 month lagged model



Note: Pipeline transportation to Cushing, OK is the outside option. Because predicted rail shipments from the Midwest region to other destinations within the Midwest are constant at the estimated fixed effect, their actual and predicted values are excluded from this figure.

Despite the simplicity of this destination-share model, it fits the data reasonably well. Figure 6 plots actual and fitted time series of the log odds ratios for each destination. There are 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 respond to oil price shocks, even those that occur two years in the past.

To verify that these correlations are indeed indicative of a lagged relationship between flows and prices that can be explained by contracting phenomena, we estimate a structural model of rail flows in the presence of contracts. In the model, which we describe in detail in appendix C, rail flows to each destination are the sum of previously contracted flows and new flows, which respond to price differentials. The key model parameter is the average contract length, which is identified off of the correlations shown in table 1, as well as Nevo

(2001) style instruments which posit a relationship between lagged flows across destinations. We estimate an average length of 24 months, which is consistent with both the trade press on crude-by-rail contracts and the correlations in table 1.

4.2 Embedding rail contracting into the pipeline investment model

To address the importance of crude-by-rail contracting for pipeline investment, 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 (see appendix E.3 for computational details). Because we do not observe individual rail contracts, we do not know whether 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). Thus, as a bounding exercise on the importance of rail contracts, we assume that rail shipments require firm ship-or-pay commitments.

Informed by our estimated average contract length from section C, we impose that rail shippers must sign 24-month ship-or-pay commitments in order to access the railroad. 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 24 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 24-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 five rail contracting "cycles". This modeling approach therefore diminishes the flexibility of crude-by-rail. Instead of being able to freely adjust volumes every month in response to downstream price shocks, rail shippers may only adjust their capacity limit every 24 months. In addition, we assume that the 24-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 large realizations of P_d, relative to the baseline model.[37] Thus, pipeline shippers are willing

[37] 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

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 crude-by-rail and an upper bound on pipeline shippers' willingness to pay for capacity.

Finally, we 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.[38]

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. 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).

Column (3) of table 2 presents the implied average per-bbl cost of DAPL from our baseline (single destination, spot crude-by-rail) model, covering a range of assumptions on the upstream supply elasticity and on r_1, the responsiveness of rail costs to rail flow (note that the implied cost is not affected by assumptions about pipeline returns to scale). 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.[39] The implied cost increases with r_1 because higher values of r_1 increase the margin earned by pipeline shippers when rail volumes are large.

When r_1 takes values of \$0 or \$3/bbl per million bpd (mmbpd), the average cost of DAPL implied by our baseline model is generally in the neighborhood of the actual \$5.50–\$6.25/bbl tariff. The implied cost exceeds this range when we let r_1 = \$6/bbl per mmbpd, consistent with this large value of r_1 reflecting short-run capacity constraints during 2013–2014 rather than the long-run rail capacity responses that would be of interest to shippers

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.

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

[39] 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 supply elasticity because large rail 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.

Table 2: Model validation

Input parameters		Implied average cost per bbl		
Supply elasticity	r_1 $/bbl per mmbpd	Baseline model	Multiple dest. model	Rail contract model
0.4	0	$6.62	$5.85	$7.68
0.6	0	$5.71	$5.03	$7.06
0.8	0	$5.24	$4.62	$6.71
0.4	3	$7.19	$6.43	$8.45
0.6	3	$6.41	$5.75	$8.03
0.8	3	$6.11	$5.51	$7.93
0.4	6	$7.73	$6.99	$9.15
0.6	6	$7.07	$6.43	$8.88
0.8	6	$6.92	$6.34	$8.93

Note: The actual DAPL tariff for ten-year committed shippers is between \$5.50/bbl and \$6.25/bbl. All rows assume $r_0 = \$10.94/\text{bbl}$. 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 rail contract model allows rail to reach a single destination and constrains rail to use 24-month ship-or-pay contracts.

signing long-term contracts (see the discussion in section 3.4.2).

In the model that allows rail shippers to enjoy spatial option value, shown in the fourth column of table 2, the implied cost is roughly 10% lower than in the baseline model. This decrease is driven by the fall 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 (figure 3).

When rail shipments require 24-month ship-or-pay contracts, shown in the right-most column of table 2, the value of crude-by-rail diminishes and the expected margin to pipeline shippers increases. The expected margin, and thus the implied average cost of DAPL, exceeds \$6.25/bbl in all specifications, and substantially so when the upstream supply elasticity is 0.4 or when $r_1 > 0$. We believe that the 24-month ship-or-pay contracts we model are overly restrictive relative to actual crude-by-rail contracts, so that the "true" model for expected shipper returns lies between the fully flexible spatial model and the rail contract model.

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 2 (which uses the actual installed DAPL capacity of 520 mbpd), but costs are substantially higher if one assumes that only 320 mbpd of new Bakken export capacity were committed to in June, 2014. Table 8 presents implied costs using alternative values for future oil price volatility (discussed in section 3.2.2) and allowing for supply uncertainty (discussed in section 3.5). Different assumptions about price volatility, which change the model's implied average cost by \$0.57–\$1.27/bbl, depending on specification, are more important than allowing for supply uncertainty, which decreases DAPL's implied cost by only a few cents per barrel.

6 Results: substitution between pipeline investment and crude-by-rail

6.1 The sensitivity of pipeline investment to rail costs

Table 3, column (4) presents our calculations of dK/dr_0—the sensitivity of pipeline capacity to per-bbl rail costs—from the baseline model, with no spatial option value or rail 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, crude-by-rail exhibits constant returns to scale. Under this assumption, our estimates of dK/dr_0 range from 41.1 thousand bpd (mbpd) per \$/bbl (for an upstream supply elasticity of 0.4) to 61.1 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 between 0.35 and 0.52. 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 crude-by-rail's externalities are addressed through regulation.

The middle and bottom sections of table 3 display results when we set r_1 equal to \$3 and \$6/bbl per mmbpd, respectively. dK/dr_0 decreases modestly with r_1, and our smallest estimate is $dK/dr_0 = 32.0$ mbpd per \$/bbl (with an upstream elasticity of 0.4 and $r_1 =$ \$6/bbl per mmbpd). Because this estimate neglects railroad spatial option value, railroad contracting, and pipeline economies of scale, and because letting $r_1 = \$6$/bbl per mmbpd yields an implied average cost of DAPL that substantially exceeds the actual tariff, 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 DAPL pipeline capacity commitment K_d. The results do not substantially deviate from baseline results in table 3 for values of K_d between 450 mbpd and 570 mbpd. For $K_d = 320$ mbpd and an upstream supply elasticity of 0.4, however, the estimate of dK/dr_0 is nearly

Table 3: 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	41.1	0.35	-26.1	-1.49	-0.19
0.6	0	50.4	0.43	-28.1	-1.29	-0.20
0.8	0	61.1	0.52	-31.8	-1.17	-0.20
0.4	3	36.0	0.31	-22.8	-1.33	-0.19
0.6	3	42.7	0.36	-23.9	-1.12	-0.20
0.8	3	50.1	0.43	-26.1	-0.99	-0.20
0.4	6	32.0	0.27	-20.3	-1.20	-0.19
0.6	6	37.1	0.32	-20.7	-1.00	-0.20
0.8	6	42.4	0.36	-22.1	-0.87	-0.20

Note: All rows assume constant returns to scale in pipeline construction and that $r_0 = \$10.94$/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.

double its value in the baseline model.[40] Appendix table 9 shows that our results are not highly sensitive to alternative assumptions about price volatility or supply uncertainty: these alternatives yield estimates of dK/dr_0 that vary by at most 25% from our baseline estimates in table 3, and often by much less than that.

The fifth column of table 3 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.[41] Expressed as an elasticity, however, the estimated impact of rail costs on expected rail flows is large, ranging from -0.87 to -1.49. The right-most column of table 3 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, rail volumes are quite insensitive to changes in rail transportation costs: the elasticity of $E[Q_r]$ to r_0 is only -0.2. Thus, our results indicate that the main channel by which environmental regulation will reduce crude-by-rail volumes is substitution to investment in pipeline capacity.

[40] Under these assumptions, the implied cost of DAPL is \$9.98/bbl (appendix table 6), so that the cost of pipeline shipment is similar to the cost of rail shipment ($r_0 = \$10.94$/bbl). Pipeline and rail transport are therefore close substitutes, inflating the estimate of dK/dr_0.

[41] 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 $dE[Q_r]/dr_0$.

Table 4: 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	36.0	0.31	-22.8	-1.33	-0.19
0.6	3	42.7	0.36	-23.9	-1.12	-0.20
0.8	3	50.1	0.43	-26.1	-0.99	-0.20
Multiple rail destination model						
0.4	3	41.0	0.35	-26.1	-1.47	-0.18
0.6	3	48.6	0.41	-27.1	-1.24	-0.19
0.8	3	57.6	0.49	-29.9	-1.10	-0.19
Rail contracting model						
0.4	3	61.6	0.52	-37.7	-2.20	-0.22
0.6	3	67.7	0.58	-36.4	-1.73	-0.23
0.8	3	75.2	0.64	-37.8	-1.46	-0.23
Baseline model, increasing returns to scale						
0.4	3	57.7	0.49	-34.6	-2.01	-0.19
0.6	3	80.5	0.69	-41.3	-1.95	-0.20
0.8	3	118.6	1.01	-54.7	-2.08	-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. Rail contract model allows rail to reach a single destination and constrains rail to use 24-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 = \$10.94/bbl.

Table 4 presents results from models allowing for multiple rail destinations, rail contracting, and pipeline economies of scale. We focus on the case with $r_1 = \$3$/bbl per mmbpd (rather than $r_1 = 0$) to obtain conservative estimates of dK/dr_0. The fourth through sixth rows of table 4 show that when crude-by-rail can flow to multiple destinations, the sensitivity of both pipeline capacity and expected rail flows to changes in r_0 is greater (by roughly 15%) than in the single-destination baseline model, reflecting the intuition discussed in section 2.3.

When we require crude-by-rail to use 24-month ship-or-pay contracts, our estimates are substantially larger than in the baseline model: the estimated dK/dr_0 ranges from 61.6 to 75.2 mbpd per \$/bbl, depending on the upstream supply elasticity. These increased magnitudes reflect the fact that requiring crude-by-rail to use ship-or-pay contracts makes rail transportation 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

Table 5: Changes in expected crude oil flows if Dakota Access were not constructed

	Baseline model			Multiple rail dests.			Rail contracting		
Supply elasticity	0.4	0.6	0.8	0.4	0.6	0.8	0.4	0.6	0.8
Change in pipe flow (mbpd)	-459	-408	-369	-459	-408	-369	-459	-408	-369
Change in rail flow (mbpd)	408	346	303	417	355	311	416	353	309
Change in total flow (mbpd)	-51	-62	-66	-42	-53	-58	-43	-55	-60
Δ rail flow as % of Δ pipe flow	-89%	-85%	-82%	-91%	-87%	-84%	-91%	-87%	-84%

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. Rail contract model allows rail to reach a single destination and constrains rail to use 24-month ship-or-pay contracts. Changes in pipeline, rail, and total flows are all expectations over future downstream price distributions, and are in units of thousands of barrels per day (mbpd). All columns assume r_0 = \$10.94/bbl, r_1 = \$3/bbl per mmbpd, and constant returns to scale in pipeline construction.

model, at least locally to $r_0 = \$10.94$/bbl. Recall that the ship-or-pay contracting framework for crude-by-rail 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 rail 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 construction.

Finally, the bottom-most section of table 4 presents results that allow for increasing returns to scale in pipeline construction, using the baseline model for crude-by-rail flows. Here, the values of dK/dr_0 and $dE[Q_r]/dr_0$ are substantially amplified relative to the constant returns case, especially when the upstream supply elasticity for crude oil is 0.8, where we estimate an elasticity of pipeline capacity with respect to r_0 that exceeds one.

6.2 Crude-by-rail in the absence of Dakota Access

Substitution between pipeline and rail transport is reflected in not only the response of pipeline investment to changes in railroads' costs, but also the response of rail flows to changes in pipeline capacity. This sub-section addresses the question of by how much rail volumes would increase were DAPL not completed.

Table 5 presents the model's counterfactual changes in expected pipeline and rail flows of crude oil when we reduce Bakken pipeline export capacity by DAPL's 520 mbpd. Across models and assumed upstream supply elasticities, expected pipeline flows decrease by 369 mbpd to 459 mbpd. These magnitudes are smaller than the 520 mbpd capacity decrease because at very low downstream price realizations, DAPL (if built) is not fully utilized. The magnitude of the decrease in pipeline flows is decreasing in the upstream supply elasticity because at high elasticities, there is a greater probability that DAPL is under-utilized. The effect of foreclosing DAPL on pipeline flows (or rail flows) is essentially the same across the baseline, multiple destination, and rail contracting models.

The second through fourth rows of table 5 show that the foreclosure of DAPL leads to increases in rail flows that offset 82% to 91% of the decrease in pipeline flows, depending on the specification. These increases are large, in that they reflect between a doubling and a tripling of rail flows, relative to the DAPL-inclusive baseline. The net decrease in crude oil flows out of the Bakken caused by removing DAPL's 520 mbpd of capacity is then between 42 mbpd and 66 mbpd. Thus, railroads' ability to effectively, even if not completely, serve as a substitute for pipeline transportation implies that policies curtailing pipeline construction will cause most of the precluded pipeline oil flow to divert to the railroads instead.

7 Conclusions

The development of the Bakken shale was associated with substantial investment in new pipeline capacity and an unprecedented boom in railroad transportation of crude oil. Following the late-2014 collapse of world oil prices, however, crude-by-rail volumes fell substantially. One interpretation of this shift 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. As a result, we show that rail and pipelines can effectively substitute for one another over the long run, even though rail volumes may ebb and flow over time. Indeed, crude-by-rail has recently staged a comeback as oil prices have risen: volumes shipped out of the Bakken have nearly tripled from June, 2017 through June, 2018 (North Dakota Pipeline Authority 2018).

While we focus on the question of pipeline versus railroad transportation of crude oil, we believe that, more broadly, our model is the first to distill intuition for how costly but flexible technologies substitute for investments in infrastructure that require large up-front commitments but have a low amortized cost. For instance, our modeling framework can be used to evaluate and intuitively understand the tradeoffs between urban light rail, which requires up-front investment in dedicated tracks and passenger loading stations, and re-routable passenger buses. The same comparative statics that we derive here for how pipeline and rail transportation are affected by factors such as relative costs of service, scale economies, and demand uncertainty are readily applicable to other settings involving tradeoffs between technologies that differ in the extent to which their costs are sunk versus variable.

Our results also speak directly to policy questions stemming from crude oil transportation's environmental externalities. Clay et al. (2017) finds that the externalities associated with crude-by-rail are particularly severe, exceeding $2 per barrel shipped. Our results imply that, had policies caused rail transporters to internalize a $2/bbl externality at the time of the Dakota Access Pipeline's investment decision, its capacity would have been at least 64,000 bpd larger than its actual 520,000 bpd capacity, and counterfactual rail flows would have been lower by at least 41,000 bpd. Under more aggressive but plausible assumptions on input parameters, the reduction in rail flows could have been greater than 75,000 bpd. We also demonstrate that policy analyses that ignore the pipeline investment mechanism would under-estimate the policy's effect on rail flows by at least a factor of five.

Finally, we show that regulatory foreclosure of pipeline investment will primarily lead to increases in rail flows rather than a decrease in overall oil production and transportation. If those opposed to Dakota Access had succeeded in blocking its construction, we estimate that roughly 80–90% of the displaced pipeline oil flow would have simply moved onto the rails. Overall then, our results illustrate that pipeline and rail transportation of crude oil are sufficiently substitutable that policies targeting one mode result primarily in a shift in crude oil flows to the other mode, with only small impacts on total crude oil production and transportation.

References

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

Amihud, Yakov and Haim Mendelson, "Asset Pricing and the Bid-Ask Spread," *Journal of Financial Economics*, 1986, *17*, 223–249.

Anderson, Soren T., Ryan Kellogg, and Stephen W. Salant, "Hotelling Under Pressure," *Journal of Political Economy*, 2018, *126*, 984–1026.

Area Development News Desk, "Enbridge Pipelines Completes Berthold, North Dakota Rail Terminal, Continues Pipeline Expansion Project," 2018. [Online; `http://www.areadevelopment.com/newsItems/8-19-2013/enbridge-pipelines-sandpaper-pipeline-project-berthold-north-dakota289123.shtml`; accessed 6 September, 2018].

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," *American Economic Review*, 1962, *52* (5), 1052–1069.

Berry, Steven T, "Estimating discrete-choice models of product differentiation," *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*, October 18, 2016.

Blas, Javier, "US oil glut shifts to Gulf of Mexico," *The Financial Times*, February 26, 2013.

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

Burton, Mark L., "Railroad Deregulation, Carrier Behavior, and Shipper Response: A Disaggregated Analysis," *Journal of Regulatory Economics*, 1993, *5*, 417–434.

Busse, Meghan R. and Nathaniel O. Keohane, "Market effects of environmental regulation: coal, railroads, and the 1990 Clean Air Act," *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.

Chamberlain, Gary, "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 1987, *34* (3), 305–334.

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," NBER Working Paper #23852, 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*, August 7, 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].

_ , "The United States is now the largest global crude oil producer," 2018. [Online; `https://www.eia.gov/todayinenergy/detail.php?id=37053`; accessed 8 December, 2018].

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].

Environmental Protection Agency, "Emission Factors for Locomotives," 2009. [Online; `https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100500B.pdf`; accessed 13 February, 2018].

Federal Register, "Control of Emissions of Air Pollution From Locomotive Engines and Marine Compression-Ignition Engines Less Than 30 Liters per Cylinder; Republication; Final Rule," June 30, 2008. Vol 73, No. 126, p. 37096.

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

_ , "LLS discount to Brent could be permanent on light sweet crude invasion," *Platts Oilgram News*, November 20, 2013.

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*, May 23, 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].

Nevo, Aviv, "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 2001, *69* (2), 307–342.

Newell, Richard G. and Brian C. Prest, "The Unconventional Supply Boom: Aggregate Price Response from Microdata," NBER Working Paper #23973, 2017.

— , — , **and Ashley Vissing**, "Trophy Hunting vs. Manufacturing Energy: The Price-Responsiveness of Shale Gas," NBER Working Paper #22532, 2016.

North Dakota Pipeline Authority, "Website Slide Deck, North Dakota Pipeline Authority, April 1, 2014," 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].

— , "Website Slide Deck, North Dakota Pipeline Authority, August 16, 2018," 2018. [Online; `https://ndpipelines.files.wordpress.com/2018/08/ndpa-monthly-update-aug-16-2018.pdf`; accessed 30 August, 2018].

Pástor, Ľuboš and Robert F. Stambaugh, "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy*, 2003, *111*, 642–685.

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].

Preonas, Louis, "Market Power in Coal Shipping and Implications for U.S. Climate Policy," Energy Institute at Haas Working Paper #285, 2017.

RBN Energy, "Crude Loves Rocking Rail Plains, Enbridge And Global Terminals In The Bakken," 2018. [Online; `https://rbnenergy.com/plains-enbridge-and-global-terminals-in-the-bakken`; accessed 6 September, 2018].

Shaffer, David, "Minnesota Regulators Order a Broader Look at Pipeline Route," *Minneapolis Star Tribune*, September 12, 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*, May 29, 2014.

Winston, Clifford, "The Success of the Staggers Rail Act of 1980," Related Publication 05-24, AEI-Brookings Joint Center for Regulatory Studies 2005.

A Online appendix: additional results 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	\$9.98	\$8.23	\$7.19	\$6.44
0.6	3	\$8.64	\$7.18	\$6.41	\$5.88
0.8	3	\$7.90	\$6.71	\$6.11	\$5.70

Note: The actual DAPL tariff for ten-year committed shippers is between \$5.50/bbl and \$6.25/bbl. All rows assume $r_0 = \$10.94$/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.

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	63.5	0.64	-53.4	-1.83	-0.18
0.6	3	56.0	0.57	-40.9	-1.31	-0.19
0.8	3	57.8	0.58	-38.2	-1.09	-0.20
Baseline model, DAPL capacity = 450 mbbl/d						
0.4	3	42.5	0.38	-30.5	-1.46	-0.19
0.6	3	46.3	0.42	-28.7	-1.18	-0.20
0.8	3	52.2	0.47	-29.7	-1.02	-0.20
Baseline model, DAPL capacity = 520 mbbl/d (main results in paper)						
0.4	3	36.0	0.31	-22.8	-1.33	-0.19
0.6	3	42.7	0.36	-23.9	-1.12	-0.20
0.8	3	50.1	0.43	-26.1	-0.99	-0.20
Baseline model, DAPL capacity = 570 mbbl/d						
0.4	3	32.6	0.27	-18.7	-1.26	-0.19
0.6	3	40.8	0.33	-21.0	-1.09	-0.20
0.8	3	48.8	0.40	-23.8	-0.97	-0.20

Note: All rows assume constant returns to scale in pipeline construction and that $r_0 = \$10.94$/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.

Table 8: Expected margin for pipeline shippers: baseline model with alternative values for price volatility and supply uncertainty

Input parameters		Implied average cost per bbl			
Supply elasticity	r_1 \$/bbl / mmbbl/d	Baseline	High price vol	Low price vol	Supply uncertainty
0.4	3	\$7.19	\$5.92	\$8.35	\$7.06
0.6	3	\$6.41	\$5.52	\$7.24	\$6.36
0.8	3	\$6.11	\$5.50	\$6.68	\$6.08

Note: The actual DAPL tariff for ten-year committed shippers is between \$5.50/bbl and \$6.25/bbl. All rows assume $r_0 = \$10.94$/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. Baseline volatility uses Brent history of long differences. High volatility case uses Brent history of one-month differences, extrapolated to longer time horizons under a random walk assumption. Low volatility case uses 18-month implied volatility in June, 2014, extrapolated to longer time horizons using a random walk assumption. Supply uncertainty case uses baseline price volatilities and an uncertain intercept for the upstream oil supply function.

Table 9: Sensitivity of pipeline capacity and expected rail flow to the cost of crude-by-rail: baseline model with alternative values for price volatility and supply uncertainty

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, baseline volatility (main results in paper)						
0.4	3	36.0	0.31	-22.8	-1.33	-0.19
0.6	3	42.7	0.36	-23.9	-1.12	-0.20
0.8	3	50.1	0.43	-26.1	-0.99	-0.20
Baseline model, high volatility case						
0.4	3	41.1	0.35	-20.6	-1.09	-0.13
0.6	3	52.2	0.44	-23.4	-0.91	-0.13
0.8	3	62.6	0.53	-26.6	-0.77	-0.12
Baseline model, low volatility case						
0.4	3	34.4	0.29	-26.0	-1.65	-0.26
0.6	3	36.9	0.31	-24.4	-1.38	-0.28
0.8	3	41.4	0.35	-25.2	-1.24	-0.29
Baseline model, baseline price volatility with supply uncertainty						
0.4	3	37.2	0.32	-23.2	-1.29	-0.19
0.6	3	43.6	0.37	-24.1	-1.11	-0.19
0.8	3	50.6	0.43	-26.2	-0.98	-0.20

Note: All rows assume constant returns to scale in pipeline construction and that $r_0 = \$10.94$/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. Baseline volatility uses Brent history of long differences. High volatility case uses Brent history of one-month differences, extrapolated to longer time horizons under a random walk assumption. Low volatility case uses 18-month implied volatility in June, 2014, extrapolated to longer time horizons using a random walk assumption. Supply uncertainty case uses baseline price volatilities and an uncertain intercept for the upstream oil supply function.

B Details on data and calibration

This appendix provides additional detail on the data we collect and use to calibrate the model. The organization of this appendix parallels the structure of section 3 in the main text.

B.1 Dakota Access Pipeline facts

No additional detail beyond that presented in section 3.1.

B.2 Oil prices and estimation of downstream oil price distributions

B.2.1 Downstream price distribution for pipeline shippers

In our baseline specification, we assume that shippers' expected price volatility over a t-month horizon is given by the historic Brent oil price volatility over a t-month horizon. We 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. We calculate historic volatility at each horizon t by taking the standard deviation of t-month differences in historic logged Brent prices for $t \in [37, 156]$. 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.

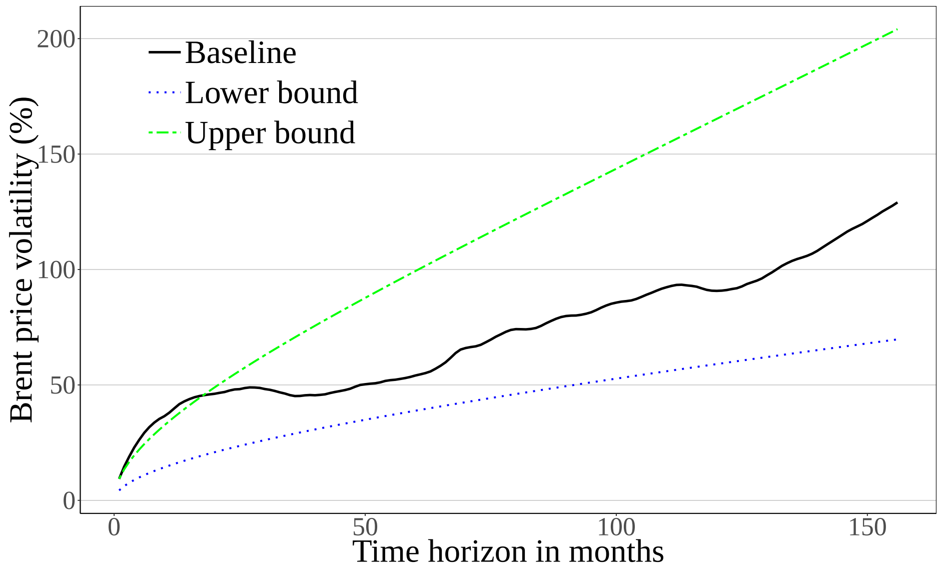
We calculate price volatilities for all horizons using Brent price data from May, 1996 (the first observation for which a 13-year lag is available in the data) through May, 2014 (the last full month before shippers committed to DAPL). We find that uncertainty over the future price of Brent increases substantially over the 3-year to 13-year horizon, from a volatility of 45% at 37 months to 129% at 13 years. Figure 7 presents the full sequence of these baseline volatilities from 37-month to 13-year horizons.[42]

Our second approach to estimating the price volatility faced by shippers uses implied volatilities from crude oil futures options, as in Kellogg (2014). This approach allows shippers' volatility expectations in June, 2014 to differ from the long-run historical average. Implied volatility data for time horizons of up to 18 months are available from the EIA via its monthly Short Term Energy Outlook; futures options are generally not liquidly traded at longer time horizons.[43] The 18-month future price implied volatility in June, 2014 was 19.7%; we extrapolate this volatility through the 13-year pipeline horizon under the assumption that oil prices follow a random walk, so that volatility increases with the square root of the time horizon. Because the June, 2014 18-month implied volatility is much smaller

[42] Volatility in figure 7 does not monotonically increase with the time horizon 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 model calculates the expected return to pipeline shippers over the full duration of their 10-year commitments.

[43] See `https://www.eia.gov/outlooks/steo/special/pdf/2009_sp_05.pdf` for a description of the EIA implied volatility data.

Figure 7: Brent price volatility estimates at horizons up to 13 years



Note: Baseline volatilities are calculated from standard deviations of long differences in historic logged Brent crude prices. Lower bounds are calculated by extrapolating the June, 2014 18-month implied volatility to longer time horizons assuming a random walk. Upper bounds are calculated by extrapolating the historic Brent one-month volatility to longer time horizons assuming a random walk. Volatilities in percent are calculated for each horizon by exponentiating the standard deviation, subtracting one, and multiplying by 100.

than the historical average 18-month volatility from our baseline forecast (45.6%), this approach produces lower price volatilities throughout the ten-year pipeline contract than does our baseline method, as shown in figure 7. We interpret the volatilities from this method as plausible lower bounds, since they assume that volatility (at any horizon) will not eventually revert back up to its long-run mean after June, 2014. This assumption is strong: when we estimate a GARCH(1,1) model on the monthly return on Brent spot prices, we find evidence of substantial historic mean reversion of volatility to its long-run average.[44] The estimated GARCH equation (combining the ARCH and GARCH terms) is $V_t = 0.0071 + 0.0733V_{t-1}$, where V_t is variance in month t. The coefficients imply a long-run one-month volatility of 9.1% (naturally close to the historic one-month average volatility of 9.3% in our baseline forecast), and the small coefficient on V_{t-1} implies rapid reversion back to this long-run mean after a volatility shock.[45] For instance, if volatility at $t = 0$ were 4.0%, volatility would reach 9.1% by only $t = 2$.

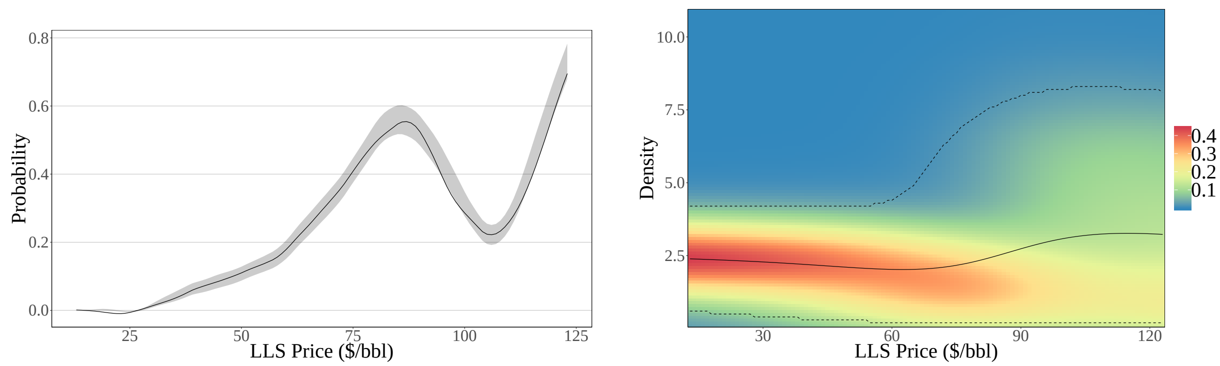
Finally, to generate an upper bound on downstream price volatility, we take the long-run

[44] Our GARCH model assumes an AR1 process for Brent monthly returns, which is sufficient to eliminate serial correlation in the residuals (at up to six lags using Breusch-Godfrey tests). A GARCH(1,1) is then sufficient to eliminate one lag of serial correlation in the unexplained variance of the residuals.

[45] We obtain 9.1% as $100 * (\exp(\sqrt{0.0071/(1 - 0.0733)}) - 1)$.

Figure 8: 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.

historic one-month horizon Brent volatility (9.3%) and extrapolate it via the square root of the time horizon (i.e., using a random walk assumption). Because the empirical relationship between historic volatility and the time horizon increases at a rate less than that from the square root assumption (as is apparent in figure 7), the volatilities from this final approach are larger than our baseline volatilities.

B.2.2 Downstream price distribution for rail shippers

This sub-section discusses our non-parametric procedure for estimating the density $f(D \mid P_d)$, which includes a point mass at $D = 0$ (which occurs whenever the best netback is to LLS). Recall that our model from section 2.3 assumes that $D \leq r_{0,d}$, so 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 we estimate $f(D \mid P_d)$ only for $D \in [0, r_{0,d}]$ and exclude those 8 days from the estimation procedure.

We non-parametrically estimate the conditional density $f(D \mid P_d)$ in two steps. First, we estimate $Z(P_d) = \Pr(D = 0 \mid P_d)$, the probability that P_d is the best downstream rail destination. 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.[46] Second, we estimate $f(D \mid P_d, D > 0)$, using local linear conditional density estimation methods in Hyndman and Yao (2002).[47] 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 $P_d - r_{0,d}$. 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.

[46] We use the **nprobust** R package, provided by Calonico, Cattaneo and Farrell (2017).

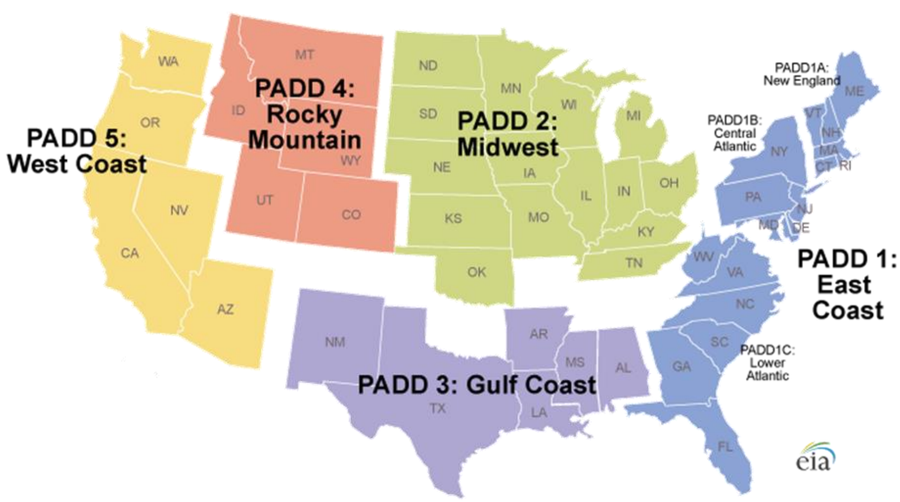
[47] We use the **hdrcde** R package.

Figure 8 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.[48] 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 ceased with the relaxation of the U.S. crude oil export ban in January, 2015.

Panel (b) of figure 8 shows the density of D, conditional on $D > 0$ and P_d. For most values of P_d, this density is concentrated around just a few dollars per barrel, with a downward trend in the mode except at values of P_d in excess of roughly \$80/bbl.

B.3 Crude-by-rail flows

Figure 9: Map of EIA Petroleum Administration for Defense Districts (PADDs)



Source: EIA

B.4 Railroad transportation costs and calibration of crude-by-rail cost function

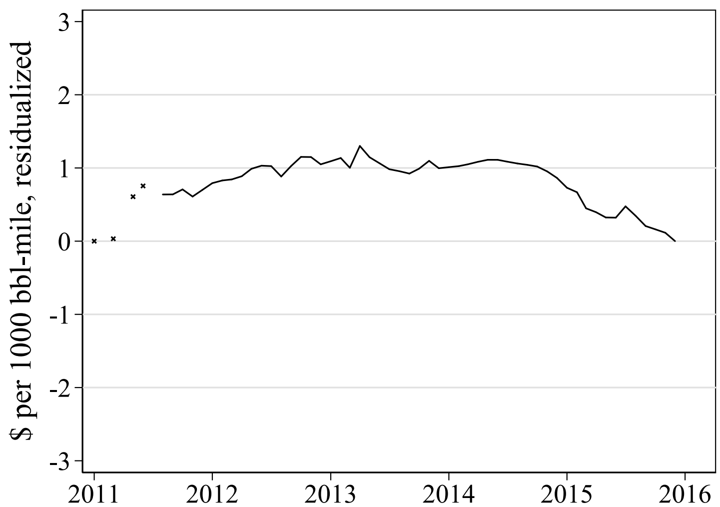
When using the STB waybill sample, we isolate the data to crude oil shipments by keeping only shipments with a Standard Transportation Commodity Code (STCC) of 1311110. To assign a movement date (and therefore a month) to each shipment, we follow the procedure described in Energy Information Administration (2017) to convert waybill dates to movement dates. The revenue measure we use is the sum of total freight line-haul revenue

[48]See, for example Blas February 26, 2013 and Hunsucker November 20, 2013.

with fuel surcharges. Our understanding from industry experts is that this revenue includes compensation for back-haul of empty rail cars; i.e., there are no separate, additional back-haul charges (personal communications with Kenneth Boyer (Michigan State University) and Justin Kringstad (NDPA)). 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.

Figure 10 presents a version of figure 5, panel (a), from the paper that is residualized. We perform a bbl-mile weighted regression of each waybill's revenue per 1000 bbl-mile on date fixed effects, region fixed effects, distance shipped, the number of carrying railroads, and the number of cars. Figure 10 plots the resulting date fixed effects.

Figure 10: STB average revenue per bbl-mile shipped, residualized



Note: Figure plots date fixed effects (relative to the January, 2011 observation) from a waybill-level regression of revenue per 1000 bbl-mile on date FE, region FE, distance shipped, the number of carrying railroads, and the number of cars, weighted by bbl-miles. Data cover sampled waybills originating in the EIA's Midwest region and terminating in the EIA's East Coast, Gulf Coast, West Coast, and Midwest regions. 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.

To calculate the percentage of crude oil moving on contract (rather than tariff) rates (87%) and the average contract discount ($0.52 per 1000 bbl-mile), we use the same sample of waybills as was used to generate figure 5, panel (a). To obtain the contract discount, we regress revenue per 1000 bbl-mile on a tariff vs contract dummy variable and on month-of-sample fixed effects, 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 control variables, we obtain a

\$0.71 contract discount.

To estimate r_1, we first normalize the freight rate and tanker car leasing rate cost series so that the units on the coefficients in the regressions are dollars per bbl per million bbl per day (mmbpd). We do so assuming that rail cars have a volume of 30,000 gallons, that rail cars complete 1.75 round trips per month (per communication from Genscape), and that the average trip distance to the Gulf Coast is 1,900 miles (Clay et al., 2017). Table 10 presents the results from regressing the resulting freight rates and tanker car lease rates on rail flows from the STB data (columns (1) and (2)) and flows from the EIA (columns (3) and (4)).

Table 10: Components of r_1

	Freight	Tanker	Freight	Tanker
STB Flows	1.36	4.50		
	(0.40)	(0.71)		
EIA Flows			0.56	2.78
			(0.30)	(0.82)
R^2	0.17	0.61	0.06	0.31
Num. obs.	57	28	57	28

Regressions of freight rates and tanker lease rates onto total rail flows, as measured using STB and EIA data. Freight rate data are monthly from January, 2011 to December, 2015, while tanker lease data are monthly from September, 2013 to December, 2015. Freight and tanker rates are scaled so that the units on the flow coefficients are dollars per bbl per million bbl per day.

B.5 Calibration of upstream oil supply function intercept

The expected Bakken production forecast we use is the NDPA's April, 2014 "case 1" forecast, which represents expected production under the prevailing EIA oil price forecast (personal communication from Justin J. Kringstad at NDPA (June, 2018)). The production forecast, supply elasticity, and expected oil price together pin down a supply intercept for each month of the contract.

In addition to its expected production forecast, the NDPA also provides a conservative ("case 2") forecast that is, on average, about 200 mbpd lower than the expected production path. We construct an "optimistic" forecast that is symmetric to this conservative forecast (i.e., in each month our optimistic forecast exceeds the baseline forecast by the same amount that the conservative forecast falls short) and solve for the series of supply intercepts that are consistent with each of these two scenarios. We then use these pessimistic and optimistic supply intercepts in our alternative specifications that allow for stochastic upstream oil supply.

C A destination share model with explicit rail contracts

To better infer average rail contract durations, we modify the empirical destination share model from section 4 to explicitly account for rail contracts. Our goal is to decompose current rail flows to destination j during month t, which we denote by $R_{j,t}$, 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. In the notation below, M_t is the total output of North Dakota crude during month t and $q_{j,t}$ is the share of "uncontracted" crude oil at month t that goes under contract to location j. The parameter s represents the share of previously contracted flows that are still under contract in the next month. Equation 5 formalizes this idea:

$$R_{j,t+1} = \underbrace{sR_{j,t}}_{\text{old flows still contracted}} + q_{j,t+1} \underbrace{\left(M_{t+1} - s\sum_{k} R_{k,t}\right)}_{\text{new flows available to be contracted}} \tag{5}$$

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 $q_{j,t+1}$, the share of available production that goes under contract to destination j in period $t+1$. We adopt a random utility choice model. Atomistic shippers, indexed by i, observe a vector of crude price differentials $X_{j,t+1}$ and destination fixed effects δ_j that account for shipment cost differences or other fixed factors. They then choose the location with the highest value of indirect utility $X_{j,t+1}\beta + \delta_j + \epsilon_{i,j,t+1}$, where $\epsilon_{i,j,t+1}$ is an *i.i.d.* logit shock.

To get an expression that relates current flows to a a destination with its lagged flows and the determinants of new flows, first define an accounting identity for $R_{0,t+1}$, the quantity of "outside" flows (via pipeline to Cushing) at time $t+1$:

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

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, which we assume is rail transport to Cushing.

Next, recall that new flow shares are given by a logit random utility model: $q_{jt} = \frac{\exp(X_{jt}\beta+\delta_j)}{1+\sum_k \exp(X_{kt}\beta+\delta_k)}$. By the standard properties of these models, the outside share in the expression for outside flows above is $(1 - \sum_k q_{k,t+1}) = \frac{q_{j,t+1}}{\exp(X_{j,t+1}\beta+\delta_j)}$ for any choice j.

Finally, we can use these two relationships to write an estimable version of the contracting model:

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

Our data on $R_{j,t}$ 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%.[49] To account

[49]This measurement error precludes using a standard log transformation to estimate the equation $\log\left(R_{j,t+1} - sR_{j,t}\right) = \log R_{0,t+1} + X_{j,t+1}\beta + \delta_j + \xi_{j,t+1}$. 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

for this issue, let $\mu_{j,t}$ be an iid measurement error that is orthogonal to $R_{j,t}$, so that observed flows are $R^*_{j,t} = R_{j,t} + \mu_{j,t}$. A feasible version of our estimating equation is then:

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

Because the measurement error $\mu_{j,t}$ is mechanically correlated with $R^*_{j,t}$, we instrument for $R^*_{j,t}$ using the second and third lags of destination price differences, in levels, squares and cubes, as well as a Nevo (2001) style measure of the average value of $R^*_{k,t}$ across destinations $k \neq j$. Lagged prices should influence $R^*_{j,t}$ through contracting forces but will be uncorrelated with sampling-based measurement error in the EIA data, and average lagged flows at other destinations will capture aggregate demand shocks but not location specific demand shocks. The instrument set also includes $R_{0,t+1}$ and destination dummy variables.[50] We estimate this model off of rail flows to the Midwest Region and the East, West, and Gulf coasts using GMM and approximately-optimal instruments in the spirit of Chamberlain (1987).[51]

The results of estimating equation (6) are in presented in table 11. The point estimate on s is 0.96, which implies an average contract length of 24 months. The coefficient on downstream prices is positive, though imprecisely estimated, consistent with the latent shipping choices being driven by variation in netbacks. Finally, the point estimates on the location dummies are all negative (and also imprecisely estimated), consistent with rail transportation to far away destinations being more expensive than rail transportation to other destinations within the Midwest region, which includes North Dakota.

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

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 $P_p(K)$, and shippers earn $P_d - P_p(K)$. This situation occurs when $P_d \geq P_p(K)$ and $\tilde{P} \leq P_r(K)$.

error in the $R_{j,t}$, 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.

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

[51] Specifically, we construct our instruments in four steps. First, we use nonlinear least squares to get a (biased) estimate of the model's parameters. Second, we regress our three endogenous variables onto our instrument set and recover their predicted values. Third, we compute a version of the conditional moment function implied by equation 6 using the predicted values, other exogenous variables in the model, and the NLLS parameter estimates. Finally, we use as instruments the gradients of this function with respect to s, β, and the δ's.

Table 11: GMM estimates of the rail contract model

	Estimates
s	0.96 (0.03)
β	0.06 (0.08)
$\delta_{\text{East Coast}}$	−2.07 (1.87)
$\delta_{\text{Gulf Coast}}$	−2.66 (2.09)
$\delta_{\text{West Coast}}$	−3.28 (1.87)
Num. obs.	231

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 j's oil price differential relative to Cushing. The δ's are fixed effects for each rail destination. The outside good is shipment within the Midwest region by rail. Two-step GMM HAC standard errors in parentheses. See text for details.

Thus, the expected value for committed shippers that accrues when the pipe is full but rail is not used is given by:

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

When rail is used (and the pipeline is therefore full), the upstream price is given by $S(K + Q_r(\tilde{P}))$, and pipeline shippers earn $P_d - S(K + Q_r(\tilde{P}))$. 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:

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

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.

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

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:[52]

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

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.[53]

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

[53]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.

E Numerical implementation of pipeline investment model

This appendix provides details on how the pipeline investment model presented in section 2 is implemented numerically. We begin with the baseline model and then discuss the models involving multiple railroad destinations and railroad contracting.

E.1 Baseline model

The primary output of the model is a calculation of the expected return to pipeline shippers over the duration of a ten-year shipping commitment, given input parameters. This calculation, which then feeds into a calculation of the derivative dK/dr_0, proceeds using four broad steps:

1. For each month t of the commitment period, derive the upstream oil supply curve and compute $P_p(K)$ and $P_r(K)$, the minimum downstream prices P_d at which the pipeline is full and rail is used, respectively.

2. For each t and for each possible realization of P_d, calculate pipeline flows, rail flows, and the pipeline shipping margin $P_d - P_u$.

3. For each t, calculate the expected return at t using the shipping margins calculated in step 2 and the distribution $f_t(P_d)$.

4. Compute the overall expected return to committed shippers across all t.

At each t, we compute the upstream supply curve intercept using the assumed value for the supply elasticity, the June, 2014 expectation of a long-run \$99/bbl oil price, and the NDPA oil production forecast for month t. $P_p(K)$ at t is then given by the upstream inverse net supply curve for oil at an output level equal to the pipeline capacity K.[54] $P_r(K)$ is then given by $P_p(K) + r_0$.

Given a value P_d, pipeline flows Q_p are either K if $P_d \geq P_p(K)$ or given by the upstream net supply curve at P_d for $P_d < P_p(K)$. Rail flows Q_r are zero if $P_d \leq P_r(K)$. Otherwise, Q_r is the solution to $S_t(K + Q_r) = P_d - r_0 - r_1 Q_r$, which is analytic for $r_1 = 0$ or can otherwise be easily solved numerically. The margin for pipeline shippers is then $P_d - S_t(Q_p + Q_r)$.

In commitment month t (where t ranges from 37 months to 156 months), the standard deviation of the lognormal distribution $f_t(P_d)$ is calculated as discussed in section 3.2. $E[P_d]$ is \$99/bbl for all t. The expected margin for pipeline shippers in month t is then the integral of $P_d - S(Q_p + Q_r)$ over $f_t(P_d)$. We compute the integral using Simpson's Rule with 1000 nodes.[55]

[54] Net supply is supply minus local refining capacity of 88 mbpd.

[55] We distribute the nodes evenly over the unit interval and then map the nodes to $\mathbb{R}^+$ using an inverse lognormal distribution that has a standard deviation given by the standard deviation of $f_t(P_d)$ at the end of the shipping commitment period (month 156). We drop the last node, which has $P_d \to \infty$. We use Simpson's Rule rather than quadrature so that the nodes on P_d are the same for all t even as the standard deviation of $f_t(P_d)$ varies, saving computation time.

Finally, to obtain the expected return over all t we multiply each expected return by the discount factor δ^t, sum over these products, and then divide by the sum of the δ^t. δ is based off an annual discount rate of 0.1 (Kellogg, 2014), so that $\delta = 1/1.1^{(1/12)}$.

To calculate the derivative dK/dr_0, we increment r_0 by \$0.10/bbl and solve for the new equilibrium pipeline capacity K such that the pipeline's average cost equals the expected shipping margin. Let M_0 denote the expected margin we calculate for the true DAPL and total export capacities, denoted by K_{d0} and K_0, respectively (i.e., M_0 denotes the values presented in table 2). Assuming that pipeline construction has no scale economies, the average cost is constant at M_0 regardless of the new capacity K. Otherwise, given a guess of K, the average cost is given by $((K - K_0 + K_{d0})/K_{d0})^{\varepsilon_c - 1}$, where ε_c is the elasticity of the pipeline's total cost with respect to capacity. The calculation of the expected pipeline shipping margin given K and the incremented r_0 proceeds as discussed above. For the parameters used in our calibration, the expected shipping margin decreases more quickly with K than does average cost, so that it is straightforward to solve numerically for the unique K that satisfies equilibrium at the incremented r_0. We repeat this process for a decrement of r_0 by \$0.10/bbl and then obtain dK/dr_0 using a two-sided derivative.

E.2 Multiple destination model

When rail can flow to multiple destinations, the only change from appendix E.1 is to the calculation of Q_r. Rail earns the downstream price $\tilde{P}$, which is distributed on $f(\tilde{P}|P_d)$. Whenever $\tilde{P} > P_r(K)$, Q_r is then the solution to $S(K + Q_r) = \tilde{P} - r_0 - r_1 Q_r$. Otherwise, $Q_r = 0$. For each P_d, we then compute the expected return to pipeline shippers by integrating the pipeline shipping margin $P_d - S(Q_p + Q_r)$ over the distribution $f(\tilde{P}|P_d)$.

E.3 Railroad contracting model

When rail flows on 24-month contracts, as discussed in section 4.2, several changes are required to our computations. We now proceed as follows:

1. Compute $P_p(K)$ as in appendix E.1.
2. Within each 24-month rail contract cycle, and given a rail contract volume K_r, compute rail flow Q_r and the expected shipping margin for each value of P_d.
3. At the start of a 24-month cycle, and at each initial downstream price P_d, compute the equilibrium volume of rail contracts K_r and the expected margin for shippers over the life of the rail contract.
4. Compute the overall expected return to pipeline shippers across all five rail contract cycles.

Within each month t of a 24-month cycle, and given capacities K and K_r, there is a critical price $P_c(K, K_r) = S_t(K + K_r)$ such that the pipe is full and all contracted rail capacity is used. Whenever $P_d > P_c(K, K_r)$, the margin earned by both pipeline and rail shippers is simply $P_d - P_c(K, K_r)$, pipeline flows Q_p equal K, and rail flows Q_r equal K_r. If $P_d \in [P_p(K) + r_0, P_c(K, K_r)]$, then the shipping margin is zero, $Q_p = K$, and Q_r is the solution to $S_t(K + Q_r) = P_d$. And if $P_d \leq P_p(K)$, then the shipping margin is zero, Q_p is given by $S_t(Q_p) = P_d$, and $Q_r = 0$.

Just before a 24-month cycle, rail shippers will commit to an equilibrium capacity K_r such that the marginal capacity cost $r_0 + r_1 K_r$ equals the expected shipping margin over the 24 months, given an initial price P_d that is observed when the contracts are signed. Given P_d and a guess of K_r, calculation of this expected margin follows steps 3 and 4 discussed in appendix E.1, but using $f_t(P_d)$ for $t \in [1, 24]$, the expected supply intercepts corresponding to the 24 months of the cycle, and margins obtained via the calculations discussed in the above paragraph. Because the expected margin over the 24-month contract strictly decreases in K_r, we can solve for the unique K_r that satisfies the equilibrium condition at each initial price P_d.

There are five rail contracting cycles over the life of pipeline shippers' 10-year commitment. The present value (at the pipeline commitment date) of the expected margin for a cycle beginning at date t is given by the integral of the expected margins obtained in step 3 over the distribution of initial prices $f_t(P_d)$, multiplied by the discount factor δ^t. We then sum these margins across the five cycles and divide by the sum of the δ^t.