UCLA
UCLA Electronic Theses and Dissertations

Title
Essays In Empirical Industrial Organization: The Determinants of the Firms' Investment Decisions

Permalink
https://escholarship.org/uc/item/5vv4j29r

Author
Cattivelli, Lorenzo

Publication Date
2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays In Empirical Industrial Organization: The Determinants of the Firms' Investment Decisions

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

Lorenzo Cattivelli

2022

© Copyright by

Lorenzo Cattivelli

2022

ABSTRACT OF THE DISSERTATION

Essays In Empirical Industrial Organization: The Determinants of the Firms' Investment Decisions

by

Lorenzo Cattivelli
Doctor of Philosophy in Economics
University of California, Los Angeles, 2022
Professor John W. Asker, Co-Chair
Professor Hugo A. Hopenhayn, Co-Chair

I study the determinants of the firms' investment decisions.

First, I quantify how building transport infrastructure affects investments. Whereas there is growing evidence that new infrastructure reduces trade costs, we lack a rigorous empirical understanding on how it impacts investments. To make progress, I leverage the U.S. crude oil industry. I specify a structural model of discrete investments centered around the producers' dynamic trade-off between current revenues and delaying investments to wait for additional transport infrastructure. I bring this model to a comprehensive dataset that allows me to match the drilling activity of oil producers with the construction of pipelines across space and over time. I estimate that pipelines reduce the producers' transportation costs by 15%,

increasing the amount of new oil wells by 28%. In sum, I find that firms substantially increase investments in response to the new transport infrastructure.

Then, I study how the firms' ownership structure affects investments. Vertical integration could increase investments for the integrating firm and reduce investments for its rivals. I measure the investment outcomes of eight vertical mergers in the US motion picture industry using 1997-2019 movie-level and companies' ownership data. Using a difference-in-difference research design, I estimate that vertical mergers increase investments for the integrating counterpart by 73.5%, while reducing investments for its rivals by 47%. Then I specify a within-firm model of resource allocation in order to separate the role of credit constraints and industry-specific technology from the change in the internalized return from investments. The results support the property rights theory.

The dissertation of Lorenzo Cattivelli is approved.

Michela Giorcelli

Martin B. Hackmann

John W. Asker Committee Co-Chair

Hugo A. Hopenhayn Committee Co-Chair

University of California, Los Angeles

2022

To my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my committee members John Asker, Hugo Hopenhayn, Martin Hackmann, and Michela Giorcelli, for their invaluable advice and continuing support.

I am indebted to my colleagues and friends, Victoria Barone, Alvaro Boitier, Benjamin Freyd, and Augusto Ospital, for their exceptional insights and encouragement.

I owe unique gratitude to Pauline Yang, Andrea Hamaui for their constant emotional support.

VITA

2014	B.A., Economics, Bocconi University
2017	M.S., Economics, Bocconi University
2018	M.A., Economics, UCLA

CHAPTER 1

The Investment Outcomes of Transport Infrastructure: The Crude Oil Industry

1.1 Introduction

Trade costs are a pervasive force in the production side of the economy, as they determine firms' profits (Firth, 2017). Given that transportation markets often rely on large infrastructure networks to connect sellers with buyers, a growing body of literature studies how building transport infrastructure affects trade costs, and how the impact propagates to prices, output, and aggregate productivity. (Donaldson, 2018, Allen and Arkolakis, 2019, Hornbeck and Rotemberg, 2021, Fajgelbaum and Schaal, 2020) Yet, we lack a rigorous empirical understanding of the firms' investment decisions causing the changes in output and productivity.

In this paper, I uncover how building new transport infrastructure affects the investments of upstream firms. I depart from the general equilibrium framework that characterizes the trade literature, and I model the firms' investment choices as dynamic optimal stopping point problems. Assuming firms respond to profits, firms' choices respond to the impact of infrastructure on costs. I exploit a revealed preference approach to recover how the firms' cost primitives depend on the transport infrastructure, while accounting for a rich set of time varying states. With these estimates at hand, I then simulate the investment response of firms to alternative levels of infrastructure.

This approach allows me to uncover how transport infrastructure projects affect both the timing and the level of investments. The dynamic interaction between transport infrastructure and upstream investments is ambiguous. On the one hand, once the infrastructure is built, firms have higher incentives to make sunk investments because of the lower shipping costs. On the other hand, firms have incentives to wait for the construction of new infrastructure, delaying sunk investments. This mechanism can be particularly acute when the infrastructure projects take long time to build, which is often the case with transport networks. The specification of a dynamic optimal stopping point has the benefit to incorporate both these factors in modeling the firms' investment decisions.

I apply this approach to the crude oil industry in Texas, where I estimate the effect of new pipelines on the drilling decisions of production companies.[1] My focus on the crude oil industry is driven by several factors that create empirical leverage to address this question. First, the transport sector plays a crucial role in this industry, since the crude's demand regions are distant from the supply basins.[2] Thus, changes in the cost of moving crude oil have large economic consequences for the industry. Oil producers drill wells to extract oil, then they ship it to the downstream demand centers, bearing the cost for transportation. Pipelines are the cheapest way to move crude oil. Since the construction of pipelines requires large fixed costs and a significant amount of time, the rapid increase of the oil supply between 2008 and 2019 wasn't immediately followed by new pipeline projects. Anecdotal evidence suggests that producers delayed the drilling of wells without a connection to a pipeline in order to save on trucking and rail costs.[3]

A second compelling reason to choose the crude oil industry is the unparalleled availability of micro-data on the firms' drilling activity and the presence of pipelines. Indeed, a crucial component to this paper is the construction of a comprehensive data set on the crude oil extraction and transportation in Texas, collected and synthesized from numerous sources.

[1] These companies are known in the industry's jargon as exploration and production (E&P) companies.

[2] In Texas the largest oil discoveries are located in the Permian Basin, which is distant from the coastal refineries and the main crude market hub located in Cushing Oklahoma

[3] "Because of the takeaway constraints, some operators with a geographically diverse portfolio of upstream assets plan to redirect capital expenditures to other onshore U.S. regions, while others may reduce well completions or look for alternative higher-cost transportation options." Energy Information Administration, September 2018

The data comprises two key components. First, I obtained the data on the universe of drilling prospects where a well has been drilled from Enverus, a private data provider, which I leverage to quantify the firms' discrete investments. Second, I was able to reconstruct the pipeline network across space and over time since 2008. Thanks to the geographic coordinates of wells and pipelines present in both datasets, I can match the pipelines with the individual wells. Compared to the past literature, this provides me a much richer variation in the access to transport infrastructure.

A descriptive look at the data indicates that pipelines affect the profitability of investments in drilling. Although I do not directly observe the producers' costs, prices at completion should contain information on the cost of shipping crude oil. Indeed, producers invest in drilling when the difference between the price and the variable cost of selling the crude is sufficiently high. I find that drilling prospects connected to the pipeline are completed at prices on average two dollars per-barrel lower than those without a connection. This evidence suggests that pipelines reduce the cost of shipping crude oil, due to the lower tariffs.

As a result, crude producers have incentives to wait for new pipeline projects in order to save in shipping costs. This creates a dynamic trade-off between the present revenues and the higher profit margins in the future. To provide evidence of this trade-off, I exploit the staggered timing of the new pipeline projects. During 2008 and 2019, I observe more than one thousands drilling prospects that get connected to a pipeline. I use the completion decisions of these prospects to assess the impact of the pipelines on the timing of drilling, modeling the time that a producer takes to drill a well as a function of the time that it takes to connect the prospect to a pipeline. I estimate that a one year delay in the pipeline connection delays the drilling decision by four months.

Although the previous evidence suggests that pipelines affect the producers' drilling decisions, it is not enough to understand how different levels of infrastructure affect investments. Furthermore, the estimated impact of pipelines on prices does not account for the effect of pipelines on the shadow cost of moving crude oil. New pipeline projects should relax the geographic constraints and the scarcity of trucking fleets in the region. As a consequence, the reduced form estimate of a two dollar price differential is a lower bound for the impact of pipelines on the producers' marginal costs.

To recover the producers' cost primitives, I specify and estimate a dynamic structural model of investments, where each drilling decision is an independent optimal stopping point problem. Every period, producers consider wether to complete the drilling prospects or to delay completion to the next period. A prospect is characterized by the availability of a pipeline, its expected production and additional profit shifters. Inspired by the descriptive facts outlined above, I model the connection to a pipeline as a stochastic state that affects the marginal cost of producers. Therefore, if a prospect is not connected to a pipeline, the producer holding the drilling option forms expectation over the prospect getting connected to a pipeline.

The model's solution centers around the dynamic trade-off between forgoing present sales for higher future profits if the well is connected to a pipeline. The dependence of the producers' payoffs on the cost of shipping crude oil creates option value from waiting. A producer decides to drill a well if the current profits exceeds the option value from waiting. Therefore, the probability that a prospect is completed depends on the impact of the pipelines on the producers' costs.

I exploit this relationship to estimate the cost primitives of the producers via maximum likelihood, following the dynamic discrete choice literature. That is, I use a nested fixed point algorithm in the spirit of Rust (1987) to recover the costs internalized by the producers. Thanks to the data on prices and quantities, I am able to separately identify the marginal cost of selling crude oil and the fixed costs of drilling. Given that I allow the marginal costs to depend on the pipeline connection, the difference between the per-barrel costs for prospects with and without a connection to the pipeline provides me an estimate for the impact of pipelines on the cost of moving crude oil.

I estimate the model using the observations coming from the drilling prospects in the

Permian Basin between 2008 and 2019, which include discovered and completed prospects, as well as discovered but uncompleted prospects. To establish the causal effect of the connection to a pipeline on the decision to drill a well, one would ideally compare the costs of the prospects with a randomly assigned connection to the pipeline to those without a connection. However, there are two main identification challenges.

First, the construction of infrastructure is not random. Pipeline builders are more likely to place a pipeline where there is a large amount of oil that allows them to recoup the large fixed costs. This creates a correlation between the local productivity of oil and the infrastructure, generating an endogeneity problem when producers react to this serial-correlation. A second challenge is that I observe a censored assignment process. By construction, I only observe drilling prospects getting connected to a pipeline conditional on not being completed.

The fine granularity of my data, combined with the crude's extraction process overcomes the endogeneity between infrastructure and local productivity. Assuming producers and builders form on average correct expectations on the amount of crude extracted by a well, the latter proxies for the expected productivity. Furthermore, the wells drilled in a region produce more oil when the underlying reservoir contains more oil. Hence, controlling for the prospect's production absorbs part of the unobserved local productivity shocks.

Since I observe the connection to a pipeline at the well-level, I can model the probability that a prospect is connected to a pipeline as a function of the prospect's cumulative production. Then I can feed this probability to my model as the producer's belief on the building of new infrastructure, explicitly accounting for the relation between infrastructure and local productivity. To further corroborate my identifying assumption, I model the dependence of the pipeline's constructions on the local supply of oil and past level of infrastructure. I show that my cost estimates are robust to the inclusion of these proxies for the local productivity in the producer's problem.

I am now left with the censorship plaguing the pipelines' stochastic process. I overcome this problem estimating the law of motion for pipelines using the observations coming from periods and fields where no prospect is completed. When this is true in the data, the conditional probability of receiving a connection to the pipeline coincides with the unconditional probability.

When I bring the model to the data, I find that on average a connection to the pipeline reduces the cost of moving crude oil by \$11.38 per-barrel. This effect is significantly higher than the two dollars difference that I estimated in the descriptive facts. This magnitude is consistent with the dynamic model capturing the additional effect of the pipeline on the shadow cost of moving crude. The access to a transport infrastructure relaxes local geography and market constraints that make moving crude disproportionately expensive. Given that in my sample, the average amount of crude extracted over a well's life cycle is 126,000 crude barrels, a pipeline connection generates savings equivalent to a \$1.5 million per-prospect.

With the support of the estimated dynamic model, I uncover the impact of the transport infrastructure on the industry's growth. Specifically, I simulate the investments in wells arising under two alternative pipelines' configurations. I then compare these investment patterns against the ones arising under the pipeline projects realized in the data.

First, I derive the counterfactual drilling if the pipelines that became active after 2008 would have not been built. Producers would substantially curtail investments, reducing the amount of completed wells by 28%. Thus, pipelines had a substantial impact on the industry's growth. Furthermore, the new pipelines contributed to the completion of larger and more productive prospects. Without building the additional pipeline projects, the average well would extract 15% less barrels of oil over its life cycle. This is because in absence of the transport infrastructure, producers have additional incentives to delay the completion of more productive wells in order to benefit from the lower transport costs.

Next, I simulate the drilling activity if all prospects got connected to the pipeline. Cumulative investments would increase by 9.5% compared to the baseline drilling patterns, although the drilling activity would slow down in the initial years of the sample compared to the baseline scenario. This is because the connection to a pipeline increases both the current profits and the future expected profits. Nonetheless, after the oil prices topped in 2012, producers boosted their investment and surpassed the investment patterns in the baseline scenario.

My results highlight how transport infrastructure affects upstream investments. The large impact of pipelines on the producers' drilling activity suggests that transport infrastructure projects generate sizable investment spillovers in the upstream industry. These results contribute to several strands of literature. First and foremost, I contribute to the growing empirical literature that studies the outcomes of infrastructure on trade costs. As recognized by Banerjee et al. (2012), transportation infrastructure is often mentioned as a key to promoting growth and development. This motivated the authors to provide reduced form evidence on the impact of access to transport networks to regional income growth in China. Motivated by the same pursuit, Donaldson (2018) exploits a general equilibrium trade model to quantify the impact of railroads in India on trade costs, income and regional productivity. The author quantifies a spatial equilibrium model to further measure the welfare's effect of the transport infrastructure. In a similar fashion, Allen and Arkolakis (2019), incorporates traffic congestion in a general equilibrium framework to measure the welfare impact of new highways.[4] These papers do not empirically address the investment patterns leading to these increase in production.

For this reason, I depart from the general equilibrium spatial framework, and instead I model investments as dynamic optimal stopping point problems that depend on the transport infrastructure. This approach is better suited to uncover the firms' investment decisions, and allows me to provide compelling evidence that building new infrastructure projects affects the investment decisions of upstream firms. The use of micro-level data sheds further light on the mechanisms behind the impact of infrastructure on investments, showing that building transport infrastructure reduces the producers' option value from waiting, thus boosting investments.

[4] Other relevant papers include: Chatterjee (2019), Fajgelbaum and Schaal (2020), Hornbeck and Rotemberg (2021)

My modeling choice connects this paper to the vast discrete choice literature that started with Rust (1987). The applications closer to my paper are Kellogg (2014), and Brancaccio et al. (2020).[5] The former studies the impact of price volatility on the optimal drilling decisions of oil producers. Methodologically, my paper has multiple points of contact with Kellogg (2014), since we both study sunk investments in oil drilling. Nonetheless, my research questions drastically differ, and I focus on how pipelines affect investments.

Instead, Brancaccio et al. (2020) study the interaction between the market for transportation services and the market for world trade in goods. The authors model the optimal ballast decisions of ships as dynamic optimal stopping point problems to recover the ships' sailing and port costs. At the same time, this paper does not analyze the micro-level producers' response to infrastructure.

My paper also tangentially contributes to the industrial organization literature of natural monopolies and regulation in energy markets (Borenstein et al., 2002,Timmins, 2002, Bushnell et al., 2008, Lim and Yurukoglu, 2018, Preonas, 2019). I provide empirical evidence on the investment outcomes of building new midstream infrastructure, which is the segment of the industry vulnerable to monopolization. Policy makers must carefully measure the investment spillovers in the upstream industry, when providing price incentives to the monopolist to build new infrastructure.

At last, I contribute to a recent and vibrant literature studying the onshore oil industry: Kellogg (2014), Covert and Kellog (2017), Anderson et al. (2018), Agerton and Upton (2019), and Herrnstadt et al. (2020). To the best of my knowledge, my paper is the first to study the investment outcomes of new pipeline projects.

[5] Other relevant applications include Collard-Wexler (2013), Kalouptsidi (2014), Hodgson (2021).

1.2 Institutional Setting and Data

1.2.1 The Organization of Crude Oil Production and Drilling Data

Oil and gas reserves lie beneath the earth's surface in geologic formations called fields. Oil producers secure the extraction rights signing lease contracts with private land owners. Subsequently, they extract crude oil from the ground drilling oil wells. Wells can be one of three types: exploratory, development, or infill. Exploratory wells are drilled into new prospective fields to discover the presence of oil. Development wells follow exploratory wells to drain the reserves. Finally, the infill wells are drilled later in the field to fully exploit the reservoir.

According to the Energy Information Administration, the average well's cost in Texas was $7.5 million in 2014, with producers reporting costs between $2.5 and $12.3 million, depending on the wells' depth and the location geological features. These drilling costs can be split among two main categories. First, the cost of actually drilling the well's hole through the rental of a drill bit. Second, once a well has been drilled, it must be completed before the well can extract the crude from the ground. Completion costs are associated with the perforation, fracking, and the disposal of water that are performed by fracking crews hired from third-parties service companies. In the end, a large portion of the well's costs are almost completely sunk, since the drilling rig's rental costs, the fracking crew's wages cannot be recovered.

Drilling wells can therefore be thought as a fully irreversible investment that is sensitive to the economic and operational conditions affecting the well's profitability. I will use this dependence to infer the cost of shipping crude oil from the producers' investment decisions. Specifically, I focus on the relationship between the decision of drilling and the presence of a pipeline to recover the impact of infrastructure on the investments' returns.

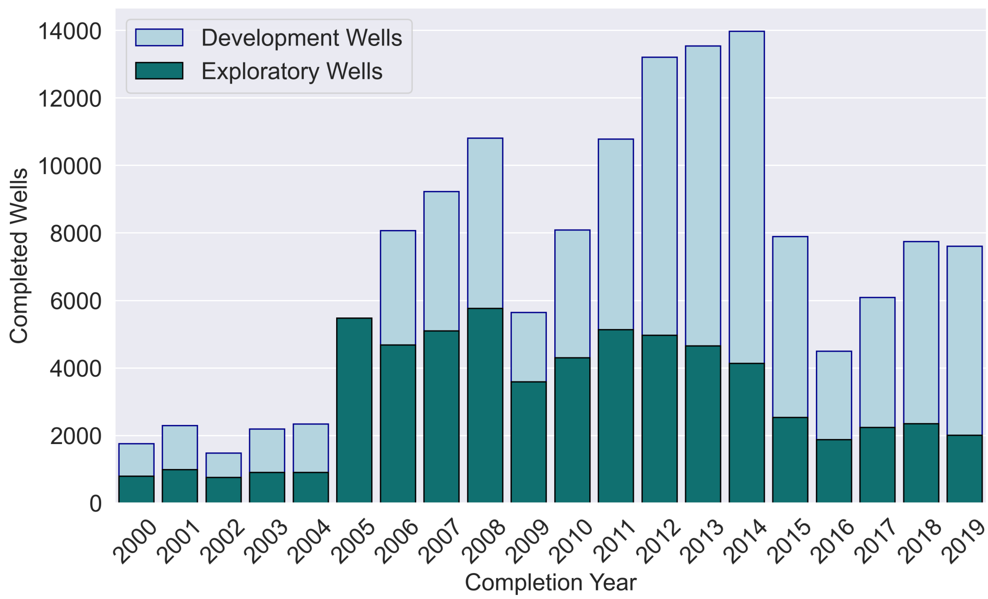
I identify the timing and location of the drilling activity using the data I obtained from Enverus, a private data provider. The dataset is at the well level, and covers all the wells that have been drilled between 2000 and 2021. For each well, the data reports its completion date and the geographic coordinates, which I use to identify the time and the location of investments in wells. Given that the focus of this paper is to understand the relationship between investment decisions and the access to the transport infrastructure, I decide to ignore investments in exploration in this paper. Exploratory wells are drilled with the precise purpose of obtaining information on the presence of oil in the underground reserves. Therefore, the amount of the crude oil obtained from the well is not central for exploration decisions, making those investments less sensitive to the presence of the transport network.

The Enverus dataset reports detailed well's characteristics. Among the more relevant ones, there are the oil producer holding the well and the crude oil extracted from the well, which combined with the crude prices allows me to measure the revenue obtained from each well[6]. Additionally, I observe the oil field where the well is located, and if this field is a wildcat - that is if there is no oil in the ground. I use this information to distinguish the exploratory wells from the development wells.[7] Figure 1.1 shows that the drilling activity in Texas drastically increased between 2005 and 2015, while exploration rates gradually declined after the new shale plays have been delimited.

[6] In the Appendix Section A.1 I describe the data that I collected on crude daily prices.

[7] In what follows I refer to development and infill wells simply as development wells. In the Appendix Section A.1 I describe the exact criteria that I use to define the development wells in the data.

Figure 1.1: Drilling Activity Over Time by Wells' Classification



Notes: The graph includes all the wells that have been completed in Texas between 2000 and 2019. The "Development" wells encompass development and infill wells.

In the next section I show that the increase in drilling activity was gradually followed by an expansion of the transport infrastructure. Then, I ask how this new transport infrastructure contributed to the development of new wells.

1.2.2 The Crude Oil Transport sector and the Pipeline Infrastructure Data

Crude oil is an intermediate good that must be refined to become gasoline. In Texas, the refining capacity is mainly located on the coastal region, since the crude's refining process requires large amount of water.[8] Consequently, oil producers must ship the crude barrels over long distances from the well-head to downstream refineries, bearing high transport costs, especially in the Permian Basin.

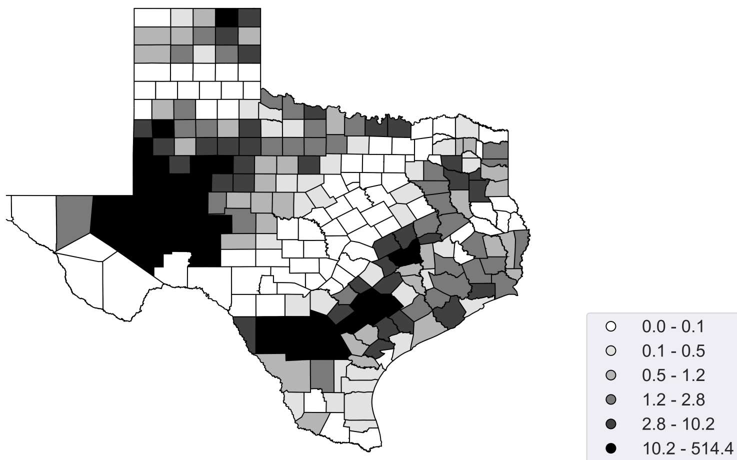
Figure 1.2 displays the allocation problem between the refining capacity and crude oil production in Texas. All the black and gray regions in panel 1.2a represent the daily crude oil supply, which must be shipped to the blue demand regions in panel 1.2b or outside the

[8]Depending on the refinery configuration, processing different types of crude can take 0.34, 0.44 and 0.47 barrel of water per barrel of crude.

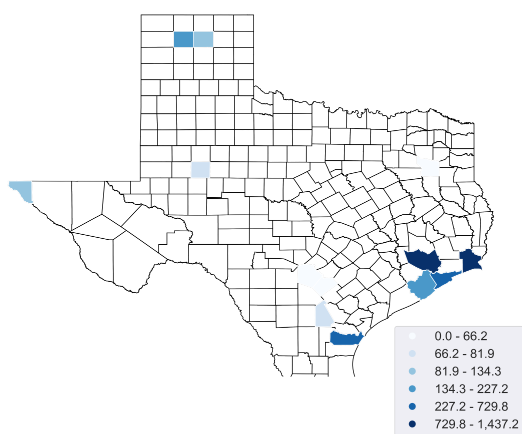
state Texas.

Figure 1.2: Spatial Mis-match of Supply and Demand

(a) Production



(b) Refinery Capacity



Notes: The supply of oil is measured in thousands of daily barrels of oil. The refinery capacity is expressed in thousands of daily crude barrel.

There are three main modes of transportation of on-shore crude oil: pipeline, rail and trucks. Pipelines are the predominant, and the cheapest shipping method for crude. The infrastructure network of pipelines can be divided in two types of pipes: gathering lines, and transmission lines. The former are short-haul lines that connect the wells to gathering stations. Transmission lines are long-haul lines with large diameters starting from the gathering stations and delivering the crude to the end points, such as refineries or market hubs.

To accurately reconstruct the pipeline infrastructure in Texas, I combine two datasets. These data are separately available from the Texas Railroad Commission of Texas. The first portion of the data includes the current stock of pipelines, which can be downloaded from their Public GIS Viewer Service (Maps) available the RRC website. The observation level of the dataset are segments of the pipeline lines. The data has detailed information about the geographic coordinates, the diameter and the length of the pipe of each segment, which I use to locate the infrastructure across space.

The construction permits for new pipelines constitute the second portion of the data.[9] I digitized the time series of those permits filed to The Railroad Commission from 2008 onward. These permits report the proposed start date, the pipeline's operator, the fluid carried by the pipe, the diameter and the miles of the pipeline. Crucially, the Texas' Maps data reports the T4 permit indicator for each pipeline segment, which I employ to combine the two separate data sources and reconstruct the dynamic evolution of the pipeline network since 2007.

Additionally, pipeline operators have to file a permit when pipelines are abandoned. The pipeline's map indicate which segments are currently abandoned, and their permit number. With this information at hand, I can identify the abandoned project in the RCC archives. Then I collected the abandonment date, and the reason for abandonment from each individual permit. Combining these multiple data sources allows me to observe the evolution of the pipeline network starting from 2008. Observing how the network evolves over time provides me a key input to quantify the impact of the transport infrastructure on the upstream's investment patterns.

I focus on the pipeline network active since 2008. Table 1.1 indicates that the data cover

[9] According to the Texas Administrative Code, each operator of a pipeline or gathering system shall obtain a pipeline permit, to be renewed annually, from the Railroad Commission of Texas. Upon the receipt of a complete application, the Commission has 30 calendar days to issue, amend, or deny the pipeline permit as filed.

a total of 48309 pipelines' segments, whose average length is 1.14 miles, with a maximum of 65.4 miles. Short segments are likely to be associated with gathering lines. The pipeline infrastructure underwent substantial changes during the period under consideration, as shown by the 19% of segments that became active after 2008, and the 30.6% of segments that got abandoned and stopped carrying crude oil between 2008 and 2020.[10] Figure 4.2 in the Appendix shows the spatial distribution of these pipelines.

TABLE 1.1. Infrastructure Dataset: Summary Statistics

	N	mean	std	min	50%	max
Abandoned segment	48039	0.306	0.461	0.000	0.000	1.000
Active segment	48039	0.694	0.461	0.000	1.000	1.000
New segment	48039	0.191	0.393	0.000	0.000	1.000
Gathering segment	48039	0.846	0.361	1.000	1.000	1.000
Miles	48039	1.147	3.788	0.000	0.142	65.429

Notes: The data covers pipeline segments that has been active since 2007. Miles indicate the length of the segments.

Pipelines require large capital investments and a long time to construct. According to the *Oil and Gas Journal* estimates, the construction cost of pipeline in 2013 was $6.57 million per-mile. After the pipeline is built, the maintenance and operation costs are around $135,00-$170,000 per-mile. Additionally, the average time to build a pipeline system in my dataset is about four years. Given the substantial expenses necessary to build pipelines, the midstream segment is much more concentrated than the upstream segment of the industry. Pipeline builders are predominantly large public companies, vertically separated from the upstream on-shore oil producers.[11]

The companies building the transport infrastructure recoup the large fixed costs charging a per-barrel shipping tariff to producers. Oil producers pay this tariff, and only bear the

[10] This percentages are not adjusted for the mileage of each segment.

[11] There are only few large global companies, such as Chevron, BP and Exxon Mobil that operate in the upstream, midstream and downstream segments of the crude industry. These companies own their own pipeline infrastructure. However, their activity in the Texas crude oil industry has been limited. The Texas's crude oil boom was led by independent E&P company only active in the upstream sector.

capital expenditures to build short pipe segments that hook-up the wells to the gathering lines.[12]

1.2.3 The Relationship between Investments in Drilling and Pipelines

Because the transport facilities are owned by third parties companies, the producers pay a fee proportional the volume of crude delivered. The shipping costs can have a large impact on the profitability of drilling, indeed, according to recent data from the Energy Information Administration, oil producers spend approximately between 2.25 million to transport crude oil over a well's life-cycle. Thus, the cost of moving crude oil is approximately 20% of the revenues obtained by the average well, assuming that the crude is sold at a price of 100$.

Crucially, there are large differences in tariffs across transport methods. In 2016 shipping the crude by pipelines over short distances costed between \$0.25 and \$1.50 per bbl. Trucks were more expensive, ranging between \$2.00 and \$3.50 per bbl. Although I did not find a tariff breakdown for Texas, a publication by the Oil and Gas Journal reports that the pipeline's tariffs for moving the crude from North-Dakota to Oklahoma are one half and one third of rail and truck's ones, respectively. Because of these large tariff differentials, the presence of a pipeline significantly affects the profitability of drilling and in turn producers' should respond to the presence of pipelines.

To understand how producers respond to pipelines, I need to identify which drilling prospects are connected to a pipeline. I limit the candidate points to those prospects where a well has actually been drilled between 2008 and 2022. For each well in the data, I observe its completion date, that is once the drilling and casing are completed and crude oil flows out of the ground. Given my focus on development wells, I assume that a drilling prospect is discovered after the producer completes the first successful producing well in a field.

Then, I compute the minimum distance of each drilling prospect from each pipeline segments. I impute a connection to a pipeline if there is gathering segment within a five miles

[12] These segments are generally less than a mile, and thus not subject to the Texas Railroad Commission pipelines' regulation. Hence, they won't appear in the map and in my dataset.

radius from the drilling prospect. To improve the precision of my measurement, I leverage the information on the midstream pipeline operator, which is reported in the infrastructure data. At the same time, oil producers must report the oil gatherer on every lease they hold, hence each well is associated with a pipeline or trucking company in charge to take the oil barrels away from the production site. When computing the minimum distance of the drilling prospect from the pipeline infrastructure, I restrict the sample to pipeline segments operated by the same firm listed as gatherer on the lease.[13]

Drilling prospects can be classified in three categories according to when they received a connection to a pipeline. A fraction of drilling prospects is never connected to a pipeline before completion, another portion has always enjoyed a connection to a pipeline, and the remaining portion obtained a connection between the date of discovery and the date of completion.

[13] I exclude from my sample those prospects that have been assigned multiple producing entities. These prospects have multiple wells extracting oil from the same location and potentially different gathering operators. The data do not report information about the names of the gathering companies for those prospects. Therefore, I cannot identify which pipeline is in charge of shipping the crude oil away from the producing entities. For this reason, I focus uniquely on drilling prospects with a single producing well.

TABLE 1.2. Descriptive Relationship between Investments and the Connection to a Pipeline

	N	mean	std	min	50%	max
Never had Connection						
Cumulative oil	23470	0.096	0.124	0.000	0.040	1.464
Months to completion	23487	44.672	30.654	1.000	38.000	144.000
Price at completion	23487	76.689	21.479	26.578	82.250	130.575
Always had Connection						
Cumulative oil	6667	0.132	0.148	0.000	0.067	2.438
Months to completion	6667	45.108	31.280	1.000	38.000	144.000
Price at completion	6667	72.865	20.958	26.578	72.589	130.575
Obtained Connection						
Cumulative oil	1567	0.193	0.129	0.000	0.184	0.813
Months to completion	1567	78.600	36.297	7.000	75.000	144.000
Price at completion	1567	60.919	17.324	26.578	54.056	102.589

Notes: The data covers pipeline segments that has been active since 2008. The cumulative production is measured in hundred of thousands of crude oil barrels. Prices at completion are lagged three months before completion and computed using the daily dollar delivery prices from the West Texas.

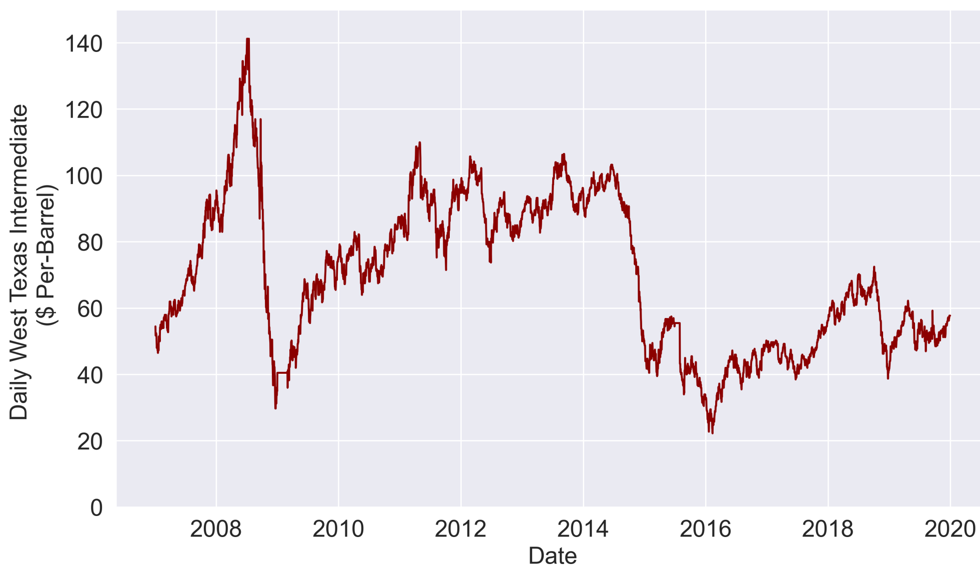
Table 1.2 shows that it takes on average 30 months longer to complete the prospects that obtained a connection to the pipeline relative to the other drilling prospects. This could indicate that producers delay completion, if they expect that a pipeline will be built in the future. The construction of the pipeline reduce the cost of moving crude oil, raising the option value from waiting.

Furthermore, on average, pipelines reach more productive prospects that generate higher shipping volumes. The average cumulative production of wells without access to the pipeline is 96,000 barrel of crude. This corresponds approximately to half of the average well's production of the prospects that obtained a connection to the pipeline, and two third of the average production of the prospects that were always connected. Larger prospects benefit more from a connection to the pipeline, because the transport costs increase with the volume of crude moved. Additionally, the pipeline builders who higher revenues from connecting more productive wells. This set the basis for an empirical correlation between prospect's productivity and the likelihood to obtain a connection to the pipeline.

At last, the bottom row of each prospect's group in Table 1.2 show that a connection to the pipe correlates with a lower price at completion. The average price at completion for wells without a connection to the pipeline is $4 per-barrel higher than wells that have always enjoyed a connection to the pipe, and $16 per-barrel higher than wells that obtained a connection to the pipeline. This is consistent with producers incurring lower transport costs to move oil from wells connected to a pipeline.

However, part of this price difference can depend on the the time series of crude prices, and the longer time of completion. Prospects that obtained a connection to the pipeline are more likely to be drilled after the 2014 oil crash, which is reported in Figure 1.3.

Figure 1.3: Time Series of Crude Oil Prices



Notes: This is the price paid for crude's delivery from the West Texas region. The crude price is measured in dollars per-barrel.

This calls for the use of an empirical strategy that controls for the time and productivity selection, in order to make some progress in the understanding of the relation between the drilling activity and the pipeline infrastructure. In what follows, I develop an empirical strategy to address these confounding factors.

1.3 Facts

This section provides robust evidence on how producers react to pipelines. First, I provide compelling evidence that the pipeline lowers the producers' variable costs. Second, I show that producers delay the completion of wells to wait for the pipeline. I use a regression approach to control for time trends and the selection of infrastructure on the prospects' expected productivity.

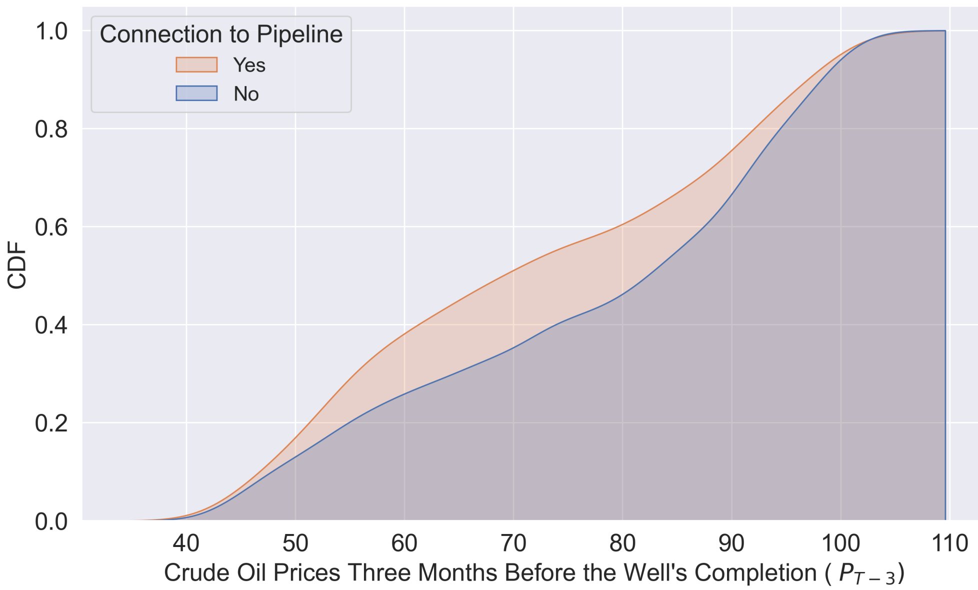
1.3.1 The Prices at Well's Completion Depend on the Presence of the Pipeline

The relationship between the crude prices at completion and the connection to a pipeline contains information on the impact of pipelines on the producer's transport costs. The range of prices when drilling becomes profitable expands if the transport costs drop. Hence, if pipelines reduce the producers' shipping costs, there should be a range of crude prices where it is profitable to complete only the drilling prospects connected to a pipeline .

Based on this intuition, the empirical distribution of prices at completion should differ by the prospects' connection state. Following Kellogg (2014), I assume that the crude price realized three months before the well's completion is the price that producers consider when deciding to drill.[14] Figure 1.4 shows that the cumulative distribution of prices for the wells connected to the pipeline is stochastically dominated by the one of wells without a connection to the pipeline. Wells connected the transport network are more likely to be completed when producers face completion prices that are below 80$ per barrel.

[14]This lag is consistent with the drilling and casing time of a well. This is also true in the data, where the median well's completion time since a prospect has been spudded is three months.

Figure 1.4: Cumulative Distribution of Crude Prices at Completion



Notes: Prices reflect the price of the West Texas Intermediate crude oil. I use wells completed for crude prices between the 5th and the 95th percentile of the empirical distribution. All prices have been lagged by three months. The sample includes single entities development or infill drilling prospects that have been completed between 2008 and 2020.

At the same time, the endogeneity of the infrastructure on the prospect's characteristics complicate the interpretation of these distributions. In order to control for this selection, I model the crude oil prices at the well's completion as a function of the access to the pipeline network, a set of wells specific covariates, and a linear time trend in the prospect's year of discovery. Specifically, for each well i, in a field l, I model the crude oil prices three months before completion, $t-3$ as:

$$P^c_{i,l,t-3} = \gamma_0 + \beta \mathbb{I}\{Pipeline\ Access\}_{i,t-3,l} + \gamma_1 X_{i,l} + \gamma_2\{T_{i,l} - 2008\} + \gamma_l + \epsilon_{i,l,t-3} \tag{1.1}$$

The main regressor of interest in equation 1.1 is the indicator variable for the connection to a pipeline, $\{Pipeline\ Access\}$. The vector X contains the well's productivity, the type of hole that is drilled, its depth and the gravity of oil extracted from the well. The variable T

is the prospect's year of discovery

Pipelines are associated with lower prices at completion, indicating that the prospects connected to a pipeline are profitable at a price per-barrel lower than those without a connection. This price difference ranges between \$1.32 and \$1.63, as reported in Table 1.3. The well's cumulative production is negatively correlated with the price at completion. The more productive wells are more likely to be profitable when crude prices are low.

TABLE 1.3. Relationship Between Prices at Completion and Transport Infrastructure

Dependent Variable:	Price at Completion				
Specification	(1)	(2)	(3)	(4)	(5)
Connection to a pipeline	-1.318***	-1.540**	-1.565**	-1.633**	-1.557**
	(0.220)	(0.661)	(0.696)	(0.688)	(0.791)
Well's cumulative oil	-0.025***	-0.020***	-0.020***	-0.020***	-0.021***
	(0.001)	(0.004)	(0.004)	(0.004)	(0.003)
Producer's size			0.003	-0.003	
			(0.006)	(0.006)	
Linear time trend	✗	✓	✗	✓	✓
Oil field FE	✗	✓	✓	✓	✗
Producer FE	✗	✗	✗	✗	✓
County FE	✗	✗	✗	✗	✓
Observations	30,402	30,402	30,402	30,402	30,402
R-squared	0.383	0.440	0.440	0.446	0.342

Notes: Standard errors in parentheses; *** p<0.01; ** p<0.05;* p<0.1. Standard errors are clustered at the field level in column (1) to (4). Standard errors are clustered at the producer level in column (5). The sample includes only single producing entities that have been completed in Texas between 2008 and 2019. The well's cumulative production is measures in thousands of crude oil barrel. Prices at completion are measured using the daily West Texas Intermediate crude oil price. Producer size in a given date is proxied adding the total number of wells that the producer has drilled across Texas.

The wedge in prices persists using different regression's specifications that account for the producer's size and producers fixed effects. In columns (3) and (4) of Table 1.3 I control for the producer's size, which is measured by the total number of wells drilled by the firm between 2008 and 2019. In column (5) I replaced the field-level fixed effects with county fixed effects in order to avoid multicollinearity issues due to single-producer operated fields.

Whereas these findings suggest that pipelines affect the producers' variable costs, one remark is due. Since I am only using the prices at completion, the reduced form estimates in Table 1.3 only capture the dollar difference between transport tariffs. However, they ignore the impact of pipelines on the shadow price of moving crude oil. The construction of new infrastructure projects relaxes market and geographical constraints, reducing the shadow price of moving crude oil where the alternative methods of transportation would be prohibitively expensive. This implies that the estimated dollar amount in Table 1.3 is a lower bound of the impact of infrastructure on the producers' costs.

1.3.2 The Timing of Investments Depends on the Presence of the Pipeline

In this section I investigate the relationship between the building of pipelines and the timing of drilling. The previous sections suggest that when a well is connected to a pipeline, the producer saves on shipping costs. Hence, producers have incentives to substitute away from trucks and rails. Consistently with this argument, Figure 4.3 in the Appendix section A.1 shows that the percentage of crude disposed by pipelines in Texas went from 33% in 2013, to 55% in 2019.

Nonetheless, the long time required to build pipelines creates a lag between the discoveries of new crude reservoirs and the construction of an infrastructure network that can absorb the additional oil supply. There is a dynamic trade-off between revenues today and higher profit margins in the future, because oil producers have incentives to wait for new pipeline projects to reduce the transport costs. Do producers actually delay the completion decisions of drilling options without a connection to the pipeline?

I answer this question accounting for the fact that pipeline operators build the pipes where there is enough crude to recover the large construction costs. Given this correlation, a cross section comparison of the completion time for drilling prospects connected to the pipeline to those without a connection produces a biased estimate. To minimize the concerns for the pipeline's endogeneity, I restrict the sample to only those prospects connected to a pipeline at the time of completion.

To identify the effect of the pipelines on the timing of investments, I leverage the time variation in the data generated by the different construction dates of the pipelines built between 2007 and 2020. The staggered time of new pipeline projects creates variation in the time that prospects get connected to the pipeline. As a result, I estimate the effect of the time incurred for connecting a prospect to a pipeline on the time for completing it. If producers delay completion, I expect that the time for obtaining a connection has a positive impact on the time for completion.

Both time variables are computed starting from the date of the prospect's discovery. Using the prospect's completion and discovery dates, I recover the share of prospects that were connected to a pipeline between their discovery and completion. I model the time for completing a drilling prospect as a linear function of the time that it takes to connect the drilling prospect to a pipeline according to

$$T^c_{i,d,f,p} = \alpha + \beta T^a_{i,d,f,p} + \gamma X_{i,d,f,p} + \lambda_d + \lambda_f + \lambda_p + \epsilon_{i,d,f,p} \tag{1.2}$$

I allow the completion time for the drilling prospects i to depend on the prospects' discovery year, d, the specific geological features of the crude oil field f through discovery year and field fixed effect. I also account for unobserved time-invariant heterogeneity, adding crude producer p fixed effect.

Identification is achieved as long as, conditional on the well's realized production, the field's geological features, and the producer's unobserved heterogeneity the timing for the pipeline connection is as good as random. Given the complex regulatory process combined with the long time for constructions, it seems reasonable to assume that producers do not have perfect forecasting over the pipeline's construction. I find that producers delay the completion of the drilling prospects without a connection to the pipeline.[15] Columns 3 of

[15] As a control group, I also include in the sample the prospects that have always had a connection to the pipeline. In the Appendix I report the estimates without the adding the prospects that have always had a connection to the pipeline.The estimates are larger, but similar. This strengthens the confidence in the results.

TABLE 1.4. The Impact of Transport Infrastructure on the Timing of Investments

Dependent Variable:	Time for Completion				
Specification	(1)	(2)	(3)	(4)	(5)
Time for connection	0.630***	0.325***	0.296***	0.332***	0.235***
	(0.070)	(0.049)	(0.039)	(0.042)	(0.054)
Well's cumulative oil		0.005	-0.008	-0.009**	-0.009**
		(0.007)	(0.005)	(0.004)	(0.004)
Price at completion		-0.036***	-0.032***	-0.026**	-0.025**
		(0.011)	(0.010)	(0.011)	(0.011)
Discovery Year FE	✓	✓	✓	✓	✓
Oil Field FE	✗	✗	✓	✓	✓
Producer FE	✗	✗	✗	✓	✓
Pipeline Operator FE	✗	✗	✗	✗	✓
Observations	7,375	7,197	7,197	7,197	7,197
R-squared	0.353	0.602	0.666	0.580	0.594

Notes: Standard errors in parentheses; *** p<0.01; ** p<0.05;* p<0.1. Standard errors are clustered at the year of discovery level. The sample includes only single producing entities that have been completed after 2008 and before 2020. Well cumulative production is measures in thousands crude oil barrel. Prices at completion are lagged by three months and measured using the daily West Texas Intermediate crude oil price. The time to complete the drilling prospect and the time to build the pipeline connections are measured in months. I impute 1 month as the time to build the pipeline connections for those drilling prospects who have always enjoyed the connection to a pipeline

Table 1.4 shows that one month delay for the pipeline's construction, delays completion by 0.33 months. This is equivalent to say that an extra year in the connection to the pipeline delays the well's completion by approximately four months. All the specifications report a strictly positive and statistically significant correlation between the time for building the pipeline and the time for completing the drilling prospect.

This results provide evidence in favor of the dynamic trade-off between the current revenues and the future profit margins. In the next section I build a structural model of wells' completion centered around this trade-off. I use this model to recover the impact of pipelines on the trade costs of crude oil, that include the tariffs' differential between pipelines and alternative method of transportation together with the shadow cost of moving crude in absence of the pipeline infrastructure.

CHAPTER 2

The Investment Outcomes of Transport Infrastructure: A Structural Model of Investment

2.1 Model

In this section, I develop a dynamic model of investments that predicts the completion decisions of the producers' drilling options. Each producer holds a drilling option over a prospect after its discovery. In every period, the producer's problem is to decide wether to drill a well. I treat the drilling and the completion decisions jointly, and in what follows I will refer to them interchangeably.

If the producer does not drill the prospect in the current period, she keeps the option of drilling the next period. If the producer pays the well's sunk cost she completes the prospect. The producer simultaneously incurs the cost of shipping the crude to the delivery point and she sells the extracted oil for a profit. Oil producers are price takers. Completion decisions are irreversible, making it an absorbing state. One key output from the specification and estimation of the model is the impact of pipelines on the investment decisions. This informs the degree to which the producers value the transport infrastructure when making the strategic investment decisions.

2.1.1 Setup

The model generates data on a large set of drilling prospects indexed by $i = 1, ..., N$. A prospect enters the producer's decision problem when it is discovered. The passing of time is discrete and it takes value $t = 1, ..., T$.[1] At the beginning of period t, each drilling prospect is characterized by an exogenous state variable $x^t_{i,c}$, which represents its completion status.

The producer decides wether to complete an uncompleted drilling prospect $a_{i,t} = 1$, or to delay the completion to the next period $a_{i,t} = 0$. Therefore, if the prospect is uncompleted at the beginning of period t the producer's choice set coincides with $\mathcal{A}_i(x^t_{i,c} = 0) = \{0, 1\}$. Instead, completion is understood as an absorbing state such that $\mathcal{A}_i(x^t_{i,c} = 1) = \{\emptyset\}$.

I assume that a producer holding multiple completion options treats them independently of one another. This is motivated by my focus on development options, which are drilled following the exploration stage, and consequently they contain little new information on the presence of oil. This limits the concern for serial correlation across the drilling decisions.

In the period when the producer decides to complete the prospect she pays a sunk cost for completion, κ_i, which depends on some observed feature of the drilling prospect. I denote these cost shifters as w_i, which include geological and technological factors. For instance, the drill type (horizontal or vertical) dictates the technology to drill the well and therefore the cost of completion materials (e.g. fracking liquid).

Following the completion decision, the well extracts the crude oil from the ground. To realize the sales, the producer moves the crude from the wellhead to the delivery point, paying the shipping costs of moving the crude. These costs are proportional to the quantity of oil moved, and they depend on the method of transportation used. Thus, a drilling prospect is characterized by its connection to a pipeline.

I incorporate the prospect-specific connection to the pipeline, $x^t_{i,p}$, in the set of the observed payoff-relevant state variables.[2] I want to capture to which extent the presence of the pipeline affects the cost of shipping crude oil. For this reason, I allow the variable costs to depend on the connection to the pipeline and the quantity of oil recovered from the ground, q_i. I denote the producer's variable costs with $c(x^t_{i,p}, q_i)$. These costs capture also the other per-barrel

[1] This data generating process creates an unbalanced panel. Different prospects enter and exit the sample at different points in time.

[2] Whereas I observe if a drilling prospect has a connection to the pipeline, I do not observe the delivery point of the crude for each well. Therefore, I cannot specify different trade routes.

costs, such as the royalty rates paid to the landowner or the well's maintenance. Crucially, the difference in per-barrel cost between prospects with different connection status identifies the cost impact of the pipeline.

The producers' profits are proportional to the quantity of oil recovered from the ground, q_i. I include the quantity of crude oil in the set of exogenous profit shifters rather than in the producer's choice set. This is motivated by the crude oil extraction process, where drilling a well is irreversible, and oil flows on the surface following the pressure differentials generated by the drilling. This limits the ability of producers to actively control the quantity of oil extracted after the well's completion, except for shutting-in the well. However, producers rarely take this decision given the high risk of permanently making unrecoverable the crude left in the ground. For this reason, I augment the set of exogenous profit shifters with the well's production, denoting it as $\tilde{w}_i = \{q_i, w_i\}$.

Producers' decisions will also respond to the crude prices, p_t. The price at completion determines the expected revenues obtained from the well, which I denote as $r(p^t, q_i)$. In sum, the common knowledge payoff-relevant state space in period t comprises

$$\mathcal{X}^t = \left\{x^t_{i,c}, x^t_{i,p}, p^t, \tilde{w}_i\right\} \tag{2.1}$$

Thus, the one-and-for-all payoff from drilling an uncompleted prospect can be written as:

$$r(p^t, q_i) - c(x^t_{i,p}, q_i) - \kappa(w_i) + \epsilon^t_i \equiv \pi(x^t_{i,p}, p^t, \tilde{w}_i) + \epsilon^t_i \tag{2.2}$$

I assume that revenues are measured with random measurement error i.i.d., ϵ^t_i, which follows a mean zero type-I extreme value distribution governed by the scale parameter σ_e. If a producer does not drill the well, the revenues and the cost from delaying completion are zero. Oil producers don't suffer any capital depreciation since they do not purchase the drilling equipment, which is rented from the oil service companies, and they pay land owners a royalty rates upon oil extraction, avoiding the costs from holding the land before completion. Thus, if a producer decides not to complete the drilling prospect, the payoffs are zero.

2.1.2 Law of Motions

In the model there are two stochastic processes that evolve over time, the construction of pipelines connecting the drilling prospects to the point of delivery, and the evolution of crude oil prices.

I assume that pipelines are stochastic from the producers' standpoint, since third-party companies build the infrastructure and new pipelines are subject to substantial regulatory delay.[3] These factors create exogenous uncertainty for the producers about the exact date when a pipeline becomes active. For this reason, I need to model the producers' beliefs over the new infrastructure.

To model the construction of pipelines, I take into account that builders aim to recoup the fixed costs of building the infrastructure. To this end, they charge a per-barrel tariff on the amount of crude oil that flows through the pipe. Whereas this tariff is regulated under the common carrier regulation, the builders are "free" to choose where to place the pipeline, becoming more prone to place a pipeline where there is a large amount of oil. This creates two types of selection. Pipelines are more likely to connect high productive prospects and to be placed in fields where the local supply is booming.

Therefore, I model the probability of receiving a connection to the pipeline as:

$$Pr(x_{i,p}^{t+1} = 1 | x_{i,p}^{t} = 0) = h(q_i, Q_f^t, S_f^t) \tag{2.3}$$

I use this probability as the producers' belief over receiving a connection to a pipeline. In equation 2.3, q_i represents the well's cumulative production, and Q_f^t denotes the supply of crude from field f. Additionally, I take into account that the initial level of pipeline infrastructure might impact the cost of new pipeline projects. For instance, it is cheaper to extend an existing network, rather than building an entirely new pipeline system. Thus, I

[3] They need to meet a series of strict environmental and safety criteria

add the S_f^t component in equation 2.3, which denotes the share of prospects and wells in field f that are connected to a pipeline.[4]

Whereas a portion of the drilling prospects was connected to a pipeline between their discovery and completion, virtually no prospects enjoyed the access to the pipeline and lost it before completion. For this reason, the construction of a pipeline is an absorbing state.

The evolution of crude oil prices is the other time-varying stochastic process that impact the producer's problem. Indeed, Kellogg (2014) studies how the optimal drilling decisions of wells depends on the price volatility. However, this is not the focus of my paper and thus I will fit a simple random walk on the crude prices. I specify the distribution of crude oil prices following Hodgson (2021), where the log oil price follows:

$$p^{t+1} = exp(log(p^t) + \zeta_t) \tag{2.4}$$

In this formulation the unobserved component ζ_t follows a normal distribution, $N(\mu_\zeta, \sigma_\zeta)$.

2.1.3 Optimal Investment Decisions

The producer's problem at a given period t is to maximize the present value V_{it} of the drilling prospect by optimally choosing the period of completion. This optimal stopping problem is given by equation 2.5 below, in which Ω denotes a decision specifying wether the prospect should be completed in each period $d \geq t$ as a function of the $x_{i,p,d}$, p_d, $\epsilon_{i,d}$ and the set of exogenous profit shifters $\tilde{w}_i$. I denote I_d a binary variable indicating the outcome of this decision rule in each period, while β denotes the monthly producer's real discount factor.

$$V_{it} = \max_{\Omega} E\Big\{ \sum_{d=t} \beta^{d-t}[\pi(x_{i,p,d}, p_d, \tilde{w}_i, I_d) + \epsilon_{i,d}(I_d)] \Big\} \tag{2.5}$$

Because the completion of a prospect is irreversible, the producer has an option value in delaying completion when the future arrival of the pipeline is uncertain. If the prospect

[4] I do not explicitly model the builders' decision problem. However, the exogeneity of the pipelines holds assuming that Q_f and S_f are sufficient statistics for the state variable in the builders' dynamic problem.

obtains a connection to the pipeline, postponing the completion decision when the shipping cost of crude are lower increases the well's profitability. The producer must trade-off the crude oil revenues from completing the prospect immediately, against the expected profits in a later period.[5]

This trade-off is captured restating the optimal stopping problem as the Bellman equation 2.6 below, in which V_i represents the current maximized value of the completion option as a function of the state variables $x_{i,p}$, p, ϵ_i, $M = \{Q, S\}$, $\tilde{w}_i$:

$$V(x_{i,p}, p, \tilde{w}_i, M, \epsilon) = max \left\{ \pi(x_{i,p}, p, \tilde{w}_i, 1) + \epsilon(1), \beta EV[(x'_{i,p}, p', \tilde{w}_i, M', \epsilon', 0)] + \epsilon(0) \right\} \quad (2.6)$$

An oil producer decides to complete a drilling prospect if and only if the profits from drilling today exceeds the option value from delaying drilling to the next period. Formally, the optimal control function $I(x_{i,p}, p, \tilde{w}_i, M, \epsilon)$ is defined by

$$\underset{\{0,1\}}{argmax}\{ \ \pi(x_{i,p}, p, \tilde{w}_i) + \epsilon_t(1), \beta E[V(x'_{i,p}, p', \tilde{w}_i, M', \epsilon')] + \epsilon_t(0)\}$$

Under the identical and independent distributed logit assumption the model provides me a closed form solutions for the probability that a prospect is drilled.[6] Given the realized state variables, the choice probability that a prospect is drilled is given by

$$Pr(I = 1|x_{i,p}, p, \tilde{w}_i, M) = \frac{exp(\pi(x_{i,p}, p, \tilde{w}_i)/\sigma_e)}{exp(\pi(x_{i,p}, p, \tilde{w}_i)/\sigma_e) + exp(\beta * E[V(x'_{i,p}, p', \tilde{w}_i, M', \epsilon')]/\sigma_e)} \quad (2.7)$$

This probability increases in the per-period profits from drilling today, while it decreases in the option value from waiting. Vice versa, the choice probability that a prospect is not drilled increases in the option value from waiting, while it decreases in the per-period profit

[5] Uncertainty in prices creates additional option value from waiting, however this is constant across drilling prospects

[6] The proof is in the Appendix section A.3

from completion and it is given by

$$Pr(I=0|x_{i,p},p,\tilde{w_i},M)=\frac{exp(\beta * E[V(x'_{i,p},p',\tilde{w}_i,M',\epsilon')]/\sigma_e)}{exp(\pi(x_{i,p},p,\tilde{w_i})/\sigma_e)+exp(\beta * E[V(x'_{i,p},p',\tilde{w}_i,M',\epsilon')]/\sigma_e)} \tag{2.8}$$

The option value from waiting is increasing in the connection to a pipeline if this generates savings in shipping cost. Given equation 2.8 this increases the probability that a prospect is not completed prior to the construction of the pipeline. Therefore, producers holding newly discovered drilling prospects without a connection to the pipeline have incentives to delay completion until a gathering line is built.

It is important to note that in this formulation, the option value from waiting is increasing in Q, S because of their effect on the producers' beliefs. An higher oil supply increases the likelihood that a pipeline is built, pushing the producer to wait for the construction of new pipeline projects.

2.2 Baseline Model

The primary goal of the model is to measure the producer's savings in transport costs generated by the pipeline infrastructure.

2.2.1 Parametrization

The data legwork described above, allows me to observe if a drilling prospect is connected to a pipeline. Given the volumetric tariff that producers pay to pipelines' operators, I parametrize the producer's variable costs as a function of the prospect's connection state:

$$c(x^t_{i,p},q_i)=q_i\Big(\sum_{x,\in\{1,0\}}\tau_x[\mathbb{1}\{x^t_{i,p}=x\}]\Big) \tag{2.9}$$

In this specification, τ_0 captures the transport costs without a connection to the pipeline together with the additional well's operating costs, which include the artificial lift, maintenance, royalty rates and income taxes. Similarly, τ_1 captures the sum of the transport costs having a direct connection to the pipeline and the additional operating costs. The key estimate of interest for this project is the difference between τ_1 and τ_0, which provides the impact of the pipeline on the producer's shipping costs.

A large portion of the costs associated with drilling and completing a well are fixed. In this specification, I allow them to depend on the technique used to drill the prospect, that is if a well is vertical or horizontal. The choice of the drilling technique is dictated by the geological characteristic of the oil reservoir, therefore I model vertical and horizontal wells as an exogenous cost shifter:

$$\kappa(w_i) = \kappa_0 + \kappa_1 \mathbb{1}\{Horizontal\} \tag{2.10}$$

Additionally, I allow the fixed cost of completion to depend on a linear time-trend in the estimation routine. However, I that assume producers' do not anticipate this variation when making the completion choice. In other words, the passing of time does not enter the set of relevant state variables in the bellman equation .

Producers do not delay completion in the anticipation of a future decline in fixed costs. In a recent contribution, Agerton (2020) shows that firms extracting natural resources obtain large productivity gains learning the resource quality and the most profitable locations for drilling. The acquisition of this new information is costly, and not driven by the exogenous arrival of new information.
For this reason, I do not incorporate the passing of time as a state variable, but I add it to the estimation routine as follows:

$$\kappa_t(w_i) = \kappa(w_i) + \kappa_2(t - t_0) \tag{2.11}$$

In equation 2.11 t is the calendar year and t_0 is the beginning year of the sample, 2008. To complete the parametrization of the per-period producers' payoffs, I need to model the present value of the expected revenues from selling the crude. In principle, this is equivalent to the sum of the product of the well's monthly production and the expected crude oil price each month. Rather then modeling this discounted sum explicitly, I follow Kellogg (2014) and model it as

$$r(p^t, q_i) = p_t q_i \tag{2.12}$$

In this formulation, q_i represents the sum of the expected monthly crude oil production. I use the crude oil price realized three months before the actual completion date as p_t. This approximation subsumes that firms value the future production using the crude oil price three months before completion. [7]

I am left to characterize the producer's expected value from delaying completion. This depends on the expectations of oil producers over the likelihood that a pipeline is built in the future, connecting the drilling prospect to the transport infrastructure. I assume a simplified stochastic process for the constructions of pipelines. In the baseline model, I suppress the dependence of infrastructure on time varying location characteristics, Q_f, S_f, conditional on the prospect-specific expected productivity. That is, I model the probability of being connected to the pipeline only as a function of the project specific production. Formally, I specify the probability that a prospect obtains a connection to the pipeline using the specification

$$h(q_i) = \frac{exp(\delta_0 + \rho q_i)}{1 + exp(\delta_0 + \rho q_i)} \tag{2.13}$$

In sum, the baseline set of parameters recovered from the data and the model comprises:

$$\tilde{\Theta} = \{\{\kappa_n\}_{n=0}^{2}, \{\tau_n\}_{n=0}^{1}, \delta_0, \rho, \mu_\zeta, \sigma_\zeta, \sigma_e, \beta\} \tag{2.14}$$

[7] Kellogg (2014) uses the 18-months crude oil futures prices. This should be the average oil price that prevails during the extraction period.

2.2.2 Estimation and Identification

I recover the cost primitives following the dynamic discrete choice literature (Rust, 1987). I build a nested fixed point algorithm, which solves for the producer's value function at every guess of the parameters, and then it estimates the cost parameters from the optimal investment choice probabilities via maximum likelihood.

Before solving the model, I estimate the parameters governing the stochastic law of motions of pipelines outside the model, using the partial likelihood implied by the evolution of the state variables. Rust (1987) proves the validity of this approach under the independence of the unobserved error term. The estimation of δ_0 and ρ is carried out by maximum likelihood of the logit model governing the pipeline data generating process specified in equation 2.13. Then, I match σ_ζ and μ_ζ to the variance and the mean of the monthly changes in the log oil price. Recovering these parameters allows me to use the transition laws to compute the producer's option value from waiting.

Each month, the producers decide wether to complete the prospects that they hold or to delay completion to the next period. Given the the exogenous profit shifters, the realized time series of pipeline constructions and oil prices, the model's solution yields the probability that any given prospect will be drilled in each month t as well as the probability that the prospect will not be drilled by the end of the sample.[8] These choice probabilities form the basis for the likelihood function, which is obtained multiplying the choice probabilities derived in equation2.7 and 2.8.

Let I_{it} denote an indicator variable that takes on a value of one if prospect i is completed in month t and zero otherwise, let T denote the final month of the sample. Let N_t denote the number of wells actually drilled at t, and denote N_0 the number of prospects not drilled.

[8]Some of these prospects won't be completed by 2019. Therefore, the model also generates the probability that a prospect won't be drilled by the end of the same conditional on being completed in 2021. This is unbiased if the prospects completed by July 2021 approximates the universe of discovered prospects by the end of 2019. This can be justified by the lower exploration rates.

The log-likelihood function can be written as

$$l((N_1, N_2, ..., N_T), N_0|\mathbf{x}, \mathbf{p}, \tilde{w}; \tilde{\Theta}) =$$
$$\sum_{t=1}^{T} N_t logPr(I_{it} = 1|\mathbf{x}, \mathbf{p}, \tilde{w}; \tilde{\Theta}) + N_0 logPr(I_{it} = 0|\mathbf{x}, \mathbf{p}, \tilde{w}; \tilde{\Theta}) \quad (2.15)$$

The estimation of $\tilde{\Theta} = \{\{\kappa_n\}_{n=0}^{1}, \tau_0, \tau_1, \sigma_e, \beta\}$ is carried out by maximizing the likelihood function using a nested fixed point routine. An inner loop computes the unknown function $EV_{\tilde{\Theta}}$ for each value of $\tilde{\Theta}$ and an outer hill climbing algorithm searches for the value of $\tilde{\Theta}$ which maximizes the likelihood function. I provide the details of the nested fixed point algorithm in the section A.4 of the Appendix.

The marginal cost and the fixed costs are separately identified up to a normalization for the discount factor and the variance of the unobserved logit term. I formally prove the identification of the different cost primitives in the Appendix section A.4, where I also show that the discount factor and the unobserved logit variance are not identified. Thus, I will calibrate those two parameters. In the rest of this section, I discuss the intuition behind some additional identification hurdles present in my setting.

I use the construction of new pipelines across space and over time to recover the prospect-level probability of obtaining a connection to the pipeline. The logit specification would allow me to parametrically identify the distribution of infrastructure shocks assuming that conditioning on the prospect production a pipeline connection was as good as random. Ideally, to identify δ_0 and ρ I would use data on the unconditional realizations of the pipelines segments connecting the universe of the drilling prospects. Observing these unconditional realizations would allow me to estimate equation 2.13. Then, I would feed these estimates to the dynamic optimization problem as the producers' beliefs.

In my setting, the data on the pipeline connections are generated by a censored process. The completion decision of a prospect is an absorbing state, which prevents me from recovering the unconditional probability of the connection to a pipeline from the data. By construction, I only observe a connection to the infrastructure conditional on not having completed a drilling prospect in the current period. This posits a threat to the conditional independence assumption. However, by the Bayes theorem:

$$Pr(x_{i,p}^{t+1} = 1 | x_{i,c}^{t+1} = 0) = Pr(x_{i,p}^{t+1} = 1) \iff Pr(x_{i,c}^{t+1} = 0) = 1 \tag{2.16}$$

That is, in periods and locations where the probability of completion decision is zero, the conditional pipeline's constructions process is equivalent to the unconditional one. Empirically, this condition is satisfied for those periods when the fields' completion rate is approximately zero. When this is true in the data, I assume that receiving a connection to a pipeline is independent of the completion's decision. Hence, I conduct the estimation of the law of transition h, only on the portion of the data coming from the months and fields when no prospect is completed.

I have already discussed how aggregate market and geographical characteristics could affect the presence of infrastructure. I evaluate the relevance of these concerns in section 2.2.5, and momentarily assume that these factors don't affect the producers' expectations over pipelines' constructions.

2.2.3 Results

I now present the estimates for the producers' cost primitives. I use the completion decisions generated by the prospects that were never connected to a pipeline and the ones that got a connection between their discovery and completion dates. That is, I exclude the prospects that have always had the connection to a pipeline. [9]

To estimate the model, I only use the prospects in the Permian basin. I choose to impose this data restriction for two reasons. First, the Permian Basin is the largest production

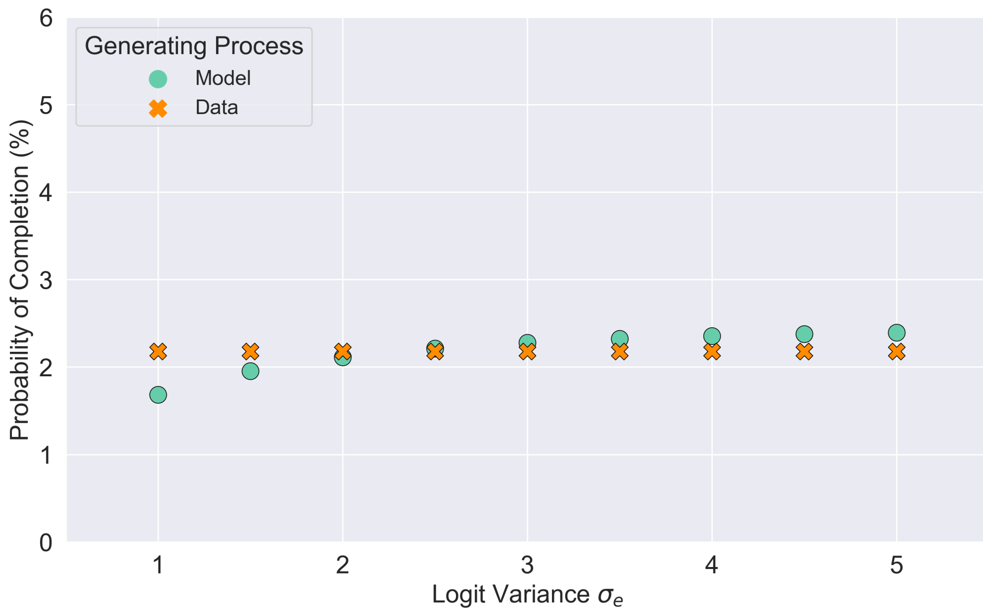
[9] The prospects that have enjoyed a connection to the pipeline could be part of a pad-drilling project. These involves multiple wells drilled in sequence. Therefore, their access to the pipeline is driven by serially correlated unobservables in the model, e.g. past discoveries and firms' private information. The economic incentives behind the investment decisions might largely differ from the completion decisions of single producing entities, and at the same time they become difficult to incorporate as a state of the dynamic model.

basin in Texas, where I observe 19896 unique development drilling prospects and a total of 1123691 decisions. This strikes a balance between the computationally tractability of the estimation and having meaningful variation in the data. Second, focusing on a single basin limits the unobserved heterogeneity in the shipping routes.

As mentioned before, both the discount factor and the logit scale parameter are not identified from the data. Thus, I pin down the discount factor using data on the average inflation and the discount rate. Between the 2008 and 2019 the average inflation rate was 1.57%. At the same time, according to a 2016 Mercer Capital report, the market applied approximately an 8.5% median discount rate to the future cash flows of wells drilled located Permian Basin. The inflation data combined with the median discount rate implies a monthly discount factor about $\frac{1.0157}{1.085} = 0.935$. Then, I estimate the cost parameters for different values of the variance of the unobserved the logit component. I evaluate the pair of discount factor and variance that optimize the model's ability to fit the data, given the estimation result. Afterward, I discuss the estimates of the cost primitives.

I assess the ability of the model to match the data comparing the choice probabilities generated from the model to their empirical counterparts. The latter are computed as the sum of completion decisions, over the total number of observations. I estimate the model for different values of the variance governing the prospects' unobserved productivity shock. Given a monthly discount factor of .935, the variance parameter that better fits the data is 2.5. This is shown in Figure 2.1. The graph benchmarks the model's probability of completion with those recovered from the data. Given that I introduced a limited number of sources of heterogeneity in the baseline model, the model demands high values for the variance of the unobserved component to rationalize the patterns in the data. The addition of an unanticipated linear time trend in the producers' fixed costs greatly improves the model's goodness of fit, as it is apparent juxtaposing Figure 2.1 and Figure 4.5 in the Appendix. This corroborates the choice of excluding the time trend from the vector of state variables, supporting the assumption that producers do not wait for an exogenous decrease in the sunk

Figure 2.1: Goodness of fit



Notes: The discount factor used across all the specifications is $\beta = .935$

costs.

Table 4.2 in the Appendix A.5 reports the cost estimates for the alternative values of the variance of the unobserved logit component. These estimates imply that having a connection to the pipeline reduces the variable cost for producers, which is consistent with pipelines having lower tariffs than trucks or rail. Whereas the presence of savings in the cost of shipping oil persists across all the specifications, the magnitude of these savings depends on the specific variance assigned to the unobserved logit component. I discuss the magnitude of the cost parameters relative to the parameter taking a value of 2.5, which provides the best fit for the data with a .935 discount factor.

This combination of parameters provides one of the most conservative estimate for the impact of the pipeline on costs. These estimates are reported in Table 2.1. The average cost saving per-barrel generated by having a connection to the pipeline corresponds to \$11.38 per-barrel. This is given by the difference between τ_0 and τ_1. Given that the average amount of crude extracted from a well is 126,000 barrels, this estimate implies an average saving of \$1.5 million per-well.

TABLE 2.1. Cost Estimates for the Baseline Specification

	Marginal Costs		Fixed Costs		
Baseline Specification	τ_0	τ_1	κ_0	κ_1	κ_2
	77.954	66.676	13.470	2.065	-1.134
	(1.521)	(1.159)	(0.084)	(0.214)	(0.014)

Notes: Standard errors in parenthesis. Standard errors are obtained from bootstrapping the standard errors with the random sampling of wells. They are computed using 50 repetitions. The discount factor across all specification is $\beta = .935$. Revenues are expressed in million of dollars. Fixed costs parameters are expressed in million of dollars. Transportation costs parameters are expressed in dollars per-barrel. The mean of the unobserved term logit distribution is calibrated at zero.

To better interpret this quantity, I benchmark these estimates against the transport tariffs and the drilling costs reported in 2016 by the Energy Information Administration. Initially, the crude must be moved from the well to the gathering stations. In 2016, the cost of moving crude oil by pipeline over the short distances in the Permian basin was about \$0.75 per-bbl. The cost of moving the crude by truck was about \$2.50. This generates an approximate cost saving of \$1.75 per barrel. Crude oil must be moved over the long distances from gathering stations to refineries or market hubs. The cost of moving crude over long-distances from the Permian ranged between \$4.00 and \$13 per barrel. Assuming that the former are the shipping cost using pipelines, and the latter the costs of using rail, the implied savings generated by the pipeline would be about \$11.75 per bbl. This number is remarkably close to the one predicted by my benchmark model.

Nonetheless, this back of the envelope calculation provides an upper limit to the dollar savings. Using \$4.00 as the pipeline tariff and \$13 as the pipeline tariffs, I assigned the largest tariffs gap for moving crude from the Delaware and Midland basins.

The large difference in the primitives recovered from the model can be explained by the shadow cost imposed by physical and market constraints from moving the crude oil in absence of infrastructure. For instance, without a pipeline the crude has to move through a longer route to reach the nearest available rail carrier. Alternatively, the scarcity of trucking freights would increase the cost of shipping of crude oil. The connection to a pipeline relaxes these constraints, decreasing the marginal cost of moving crude oil and generating cost savings that exceed the pure differential between tariffs.

The estimation routine provide an average $13 million expenditures in the fixed cost of drilling and completion, with an additional cost of two millions for horizontal wells. These large estimates reflect that during many months no drilling occurs. As expected, horizontal wells are more expensive than vertical ones. Fracking is a new technology, and it demands pumping more materials to recover the crude.

At last, the estimates at the bottom of Table 4.2 in the Appendix section A.5 show that the prospect's cumulative production has a sizable impact on the probability of that the well receives a connection to the pipeline. In turn, pipelines are more likely to reach large wells, suggesting that the scarcity of infrastructure disproportionately delays the development of the most profitable wells. This correlation magnifies the impact of the pipelines on the industry's growth.

2.2.4 Sources of Heterogeneity

In this section, I assess the ability of the model to capture the dimensions of heterogeneity present in the data. First, I show that the model can mimic the drilling activity over time and across space. I use years and sub-basins as the time and spatial dimensions of interest, leveraging the fact that the territory covered by the Permian Basin is classified in three sub-basins with underlying geological differences between reservoirs.

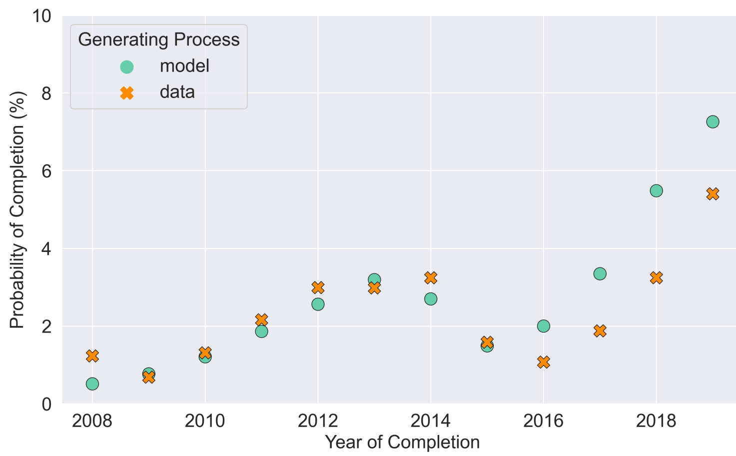
I compare the choice probabilities predicted by the model with their empirical counterpart. The model is able to replicate the increase in drilling activity after the 2008 financial crisis, and the subsequent drop in investments triggered by the crash of oil prices in 2014.[10] The model also replicates the bounce back in drilling after the crash.

Figure 2.2a displays a divergence between the model and the data at the end of the sample

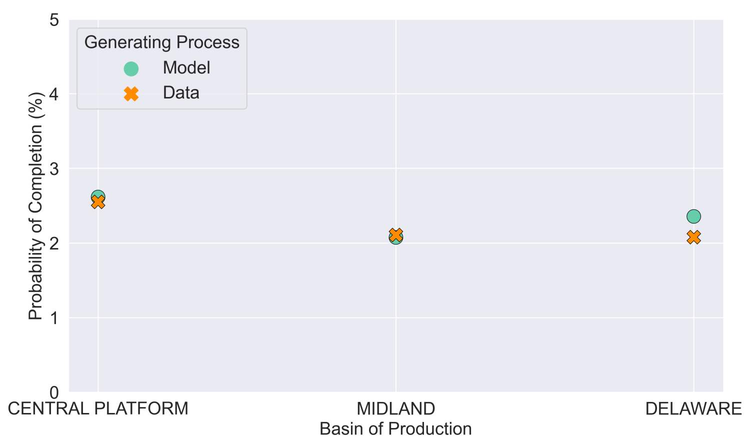
[10] The price crash can be visually seen in Figure1.3

period. The estimated cost of extraction declines too fast to fully rationalize the data. This sharp bounce back is likely to be driven by the high coefficient imposed on the linear time trend, which makes the fixed costs disproportionately low over time. Nonetheless, the model captures the main time-trend underlying the data. This speaks to the predominance of the price dynamics, the construction of infrastructure, and the decline in drilling costs into explaining the relevant time heterogeneity in the data.

Figure 2.2: Time and Location heterogeneity



(a) Time



(b) Location

Notes: The discount factor used across all the specifications is $\beta = .935$

Now, I shift my focus to the spatial heterogeneity. Across the Permian's sub-basins, there is a small discrepancy between the model predictions and the data. This is displayed by the overlapping points and crosses in Figure 2.2b. The accurateness of the baseline model can be rationalized by the fact that the distribution of drill types together with the cumulative production capture the spatial dimensions correlating with the investment's decisions.

The baseline model does account for differences in the producers' cost structure. Nonetheless, the firm's size might explain part of the variation in the costs of drilling wells and extracting crude oil. Large producers can leverage economies of scales splitting the drilling rental costs between multiple prospects. They also enjoy a better bargaining position with landowners and oil service companies.

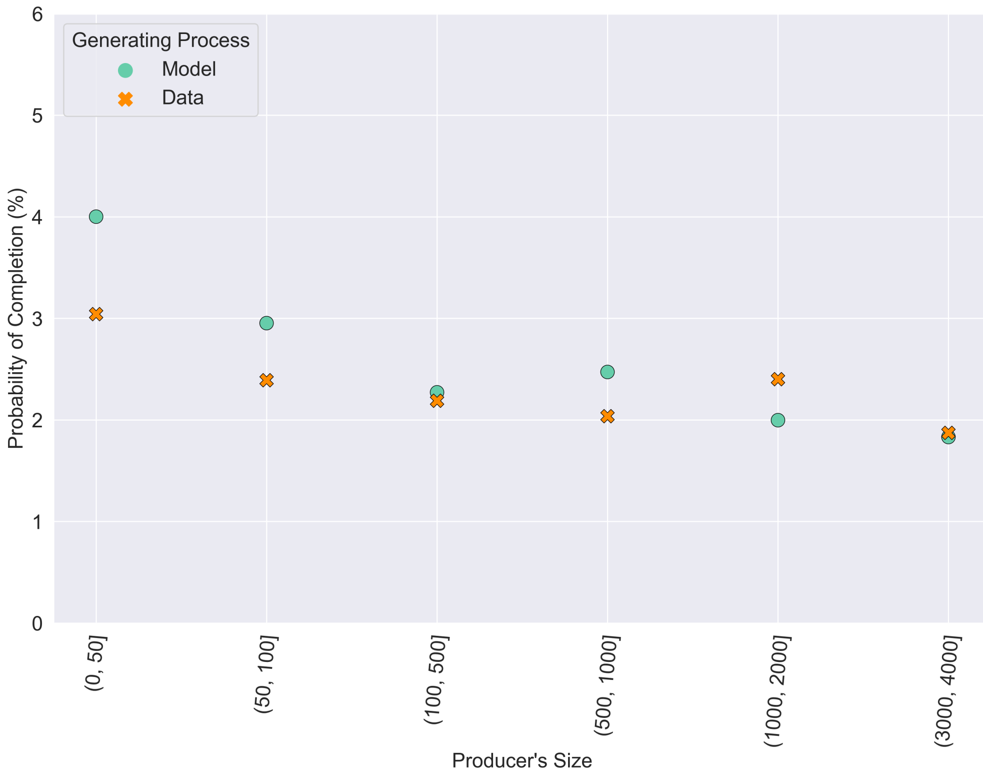
These factors reduce the cost of investing and completing a well relative to a small producer. To assess the empirical validity of this claim, I classify the upstream producers based on the total number of drilling prospects that they hold during the 2008-2019 sample years. On average, the model comes close to predict the probability of completion by size. However, it over-estimates the probability of completion for those producers that held one hundred or less drilling prospects. Figure 2.3 displays a one percentage point difference between the model and the data predictions for producers with less than fifty prospects.

This pattern is consistent with small producers bearing higher costs than the estimated ones. Large producers drive down the average cost of completion in the estimation routine, making the average estimated costs lower than the true cost primitives of small producers. Lower costs put upward pressure on the choice probability predicted by the model. This gap quickly closes as the number of prospects held by producers exceeds one hundred.

Another interesting pattern emerging from the data is that the absolute probability of completion declines with size. This can be rationalized by larger producers holding a broadly differentiated set of prospects and taking on more risky projects in new fields. By contrast, small producers might enter where reserves have already been proved increasing their drilling rates.

I further investigate the fixed heterogeneity across producers. In Figure 2.4, I plot the choice probability predicted by the model as a function of the choice probability in the data.

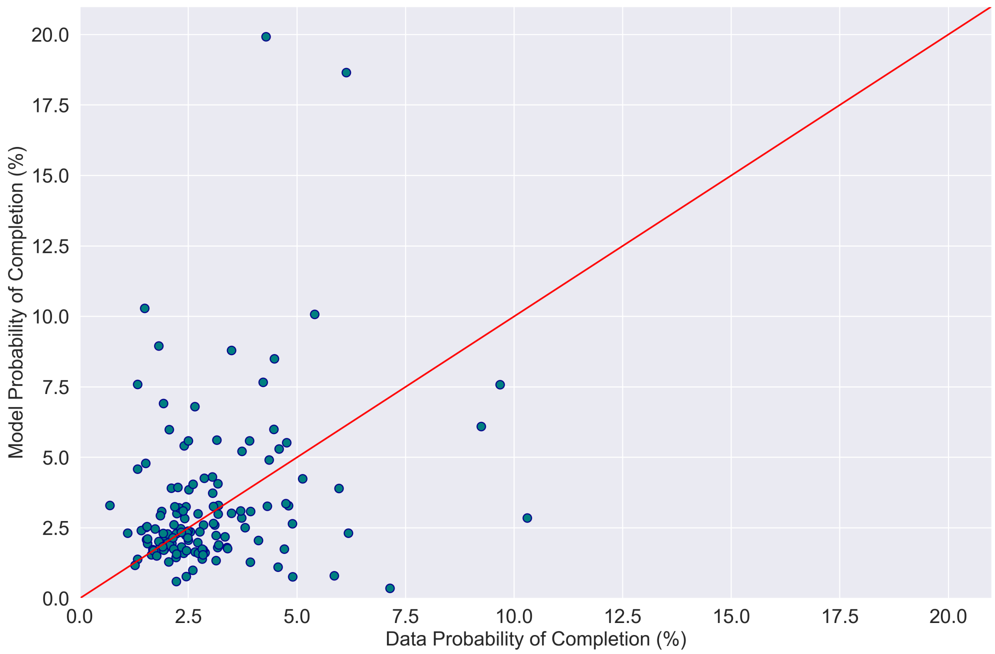
Figure 2.3: Size heterogeneity



Notes: The discount factor used across all the specifications is $\beta = .935$. The sample includes all the producers holding a drilling prospect during the years 2008-2021. The producers' size is computed using the total number of development prospects held by the firm in Texas.

Any divergence from the forty-five degree line reflects a discrepancy between the model's prediction and the data. There is dispersion across producers. However, the dispersion in the data seems uncorrelated with the drivers of the investments' decisions, since the model does not consistently over-predict or under-predict the choice probabilities. The dots in the graph are evenly distributed below and above the forty-five degree line.

Figure 2.4: Producer heterogeneity



Notes: The sample includes the producers that held more than one-hundred development projects during the years 2008-2021.

2.2.5 Robustness checks

In this section I extend the model to account for additional heterogeneity. The baseline specification does not accurately capture the differences in completion decisions between producers with different sizes, posing a threat to identification if the producers' size correlates with getting a connection to the pipeline.

In order to address this concern, I allow the fixed cost of drilling and the variable costs of extraction to depend on the producer's size. I classify each producer as small or large based on the total number of wells the firm drills during the 2008-20219 sample period, splitting the sample at the 25th percentile number of wells per producer. Then, I specify the producer's cost structure as a function of the producer's size, which enters the payoffs as an additional fixed profit shifter. Formally, I model variable and fixed costs as follows:

$$c(x_{i,p}^t, q_i) = q_i \Big(\sum_{x, \in \{0,1\}} \tau_x \mathbb{1}\{x_{i,p}^t = x\} + \tau_2 \mathbb{1}\{Small\} \Big) \tag{2.17}$$

$$\kappa(w_i) = \kappa_0 + \kappa_1 \mathbb{1}\{Horizontal\} + \kappa_3 \mathbb{1}\{Small\} \tag{2.18}$$

In this specification, τ_2 captures the impact of the producer's size on the variable costs. Instead, κ_3 captures the effect of size on the sunk costs of drilling. [11]

A natural concern when estimating the impact of infrastructure projects is that of bias due to a potential correlation between project placement and unobserved changes in local environment (Donaldson, 2018). These concerns are likely to be less important in my setting, given the granularity of the data. My model allows for selection at the prospect level proportional to the cumulative oil extracted by each well.[12] The quantity of crude extracted from a well depends on the past exploration activity and the geological features of the prospect's location. Therefore, the realized production correlates with the spatial characteristics, partially mitigating the concerns for unobserved location characteristics.

Nevertheless, to better control for the correlation between the building of new infrastructure and changes in the local economic environment, I expand the set of factors driving the construction of new pipelines. I specify a stochastic process for the pipeline' connections that allows the infrastructure to depend on time-varying field-level characteristics. I assume that the likelihood that a prospect gets connected to a pipeline depends on the local supply of crude oil, which I measure with the cumulative oil extracted from the field. Additionally, I incorporate the correlation between the connection to a pipeline and the amount of infras-

[11] I do not incorporate producers' fixed effects as an additional state variable. This would greatly expand the state space. At the same time, the unobserved heterogeneity across producers appeared uncorrelated to the drilling rates, reducing the threat of endogeneity along this dimension.

[12] Oil producers conduct extensive analysis before drilling to forecast the crude present in the ground. Additionally, focusing on development wells reduced the uncertainty on the presence of oil. Therefore, producers are likely to be able to accurately predict the crude extracted from a development prospect. Hence, the realized production is a credible proxy for the production expected by producers when making the completion decision.

tructure built on the field. This is measured with the field-level share of prospects connected to pipelines.

Incorporating the local supply of oil is motivated by the fact that builders have more incentives to construct infrastructure where the supply will grow. Given a volumetric tariff structure, builders have more chances to recover the large infrastructure costs when large volumes of crude are shipped through the pipelines. Separately controlling for the past level of pipeline infrastructure is motivated by the fact that midstream companies are likely incur lower construction costs where there is a well developed infrastructure network. The law of motions for pipelines follows a logit probability distribution according to:

$$h(Q_f^t, S_f^t, q_i) = \frac{exp(\delta_0 + \delta_1 Q_f^t + \delta_2 S_f^t + \rho q_i)}{exp(\delta_0 + \delta_1 Q_f^t + \delta_2 S_f^t + \rho q_i) + 1} \tag{2.19}$$

Where Q_f, S_f and q_i represent the field supply of crude, the field level infrastructure and the prospect's production of crude, respectively. The field's oil supply and the infrastructure level enter the producer's drilling choice problem only through her beliefs on the pipeline's arrival. Therefore, I do not incorporate them as separate state variables.

I first estimate the parameters governing the logit law of transition for new pipeline connections in equation 2.19. Then, I predict the probability that a prospect receives a connection to the pipeline given its cumulative expected production, the local supply of oil and the level of infrastructure. I incorporate this predicted probability as a new state variable for the producer's problem. This approach has the advantage of reducing the state space, keeping the estimation routine computationally tractable.

I first estimate the model adding only the size's indicator, without changing the pipeline's law of motion. Small firms have fixed costs \$1.7 million higher than the producers beyond the 25th percentile distribution, according to the first column of Table 4.3 in the Appendix section A.5. This result is consistent with large producers having substantial economies of scales.

However, small producers have \$11 per-barrel lower variable costs from extracting oil. This result can be rationalized by the differences in the areas of activity. Anecdotal evidence suggests that small producers enter an oil plays as followers, which might confine them in regions with less crude. Thus, they drill smaller and shallower wells. Those prospects demand less maintenance and fracking materials, which reduces the variable costs of drilling.

As a second robustness check, I augment the pipeline's law of motion. Both the field crude supply and the amount of pipeline's infrastructure present in the field have strong predictive power for receiving a connection to a pipeline. The positive coefficients in the second column of Table 4.3 supports this evidence. Adding these variables to the law of motion decreases by half the coefficients on the well's cumulative production. This also implies that controlling for the well's productivity largely captured the impact of location-specific characteristic.

Interestingly, in column (2) the size's estimates shrinks compared to column (1). Controlling for the local characteristics capture part of the size's effect on costs. This strengthen the intuition that the producer's size affect the costs' estimates because it correlates with different drilling areas. Additionally, the complete model reduces the producer's specific dispersion. This can be seen comparing Figure 2.4 above and Figure 4.7 in the Appendix section A.5. In the latter graph, the probability of completion is more concentrated along the diagonal.

Despite these improvements in precision, the estimated effect of pipelines on the marginal costs is consistent across the model's specifications. This estimate ranges between \$10.8 per-barrel and \$ 10.6 per-barrel in the augment model . The magnitude of this estimate is only \$0.5 lower than the one generated by the baseline model. The proximity of all these estimates suggests that the heterogeneity missing in the baseline model is uncorrelated with the building of pipeline.

This strengthens the credibility of the parsimonious specification in the baseline model. The main goal of this project is to capture the effect of the midstream infrastructure on the upstream investments, which is almost unaffected by incorporating additional states in the model. Given the purpose of this project, to properly decompose the factors driving the producers' beliefs and producer's fixed effect is second order. For this reason, I favor the baseline specification to conduct the counterfactual analysis in the next section.

2.3 Counterfactual Analysis

To illustrate how the transport infrastructure and investments interact, I simulate the drilling dynamics implied by alternative pipeline's configurations. In the first configuration, I assume that none of the prospects gets connected to a pipeline. In the second configuration, I assume that each prospect gets connected to a pipeline. Then, I compare the investments given these two configurations against the ones under the pipelines that were actually built in the Permian Basin. I define the latter as the baseline investments produced by the model.[13]

The difference between the first counterfactual and the baseline informs us how investments in wells would have changed if the pipelines built after 2008 were not built. New pipelines significantly contributed to the growth of the extraction of crude oil. The impact of new pipelines on the drilling activity is quantified by the difference between the dotted line and the continuous line in Figure 2.5a.

The absence of new infrastructure would have increased the cost of shipping crude, and thus depressed the yearly investments. The additional infrastructure increased the total number of completed wells by 1672 in the sample period 2008-2019. This is equivalent to approximately a 8.5 percentage point difference in the cumulative number of investments, that translate into a 28% increase in additional wells.

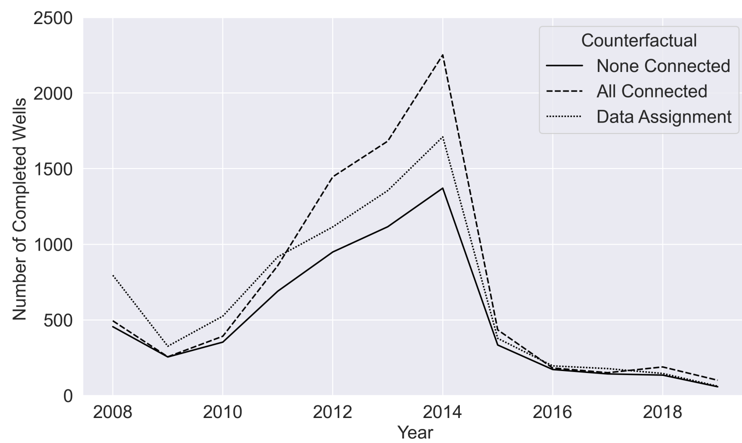
Instead, the difference between the second counterfactual and the baseline investments, informs us how the industry dynamics would have changed if additional pipelines reached the prospects that did not obtained a connection to the pipeline. I assume that the counterfactual connections would have been realized at the date of discovery. Connecting the prospects without a connection to a pipeline further raises investments in drilling. This is in line with

[13]In both counterfactuals, I hold unchanged the producers' beliefs on the probability of receiving a connection to the pipeline.

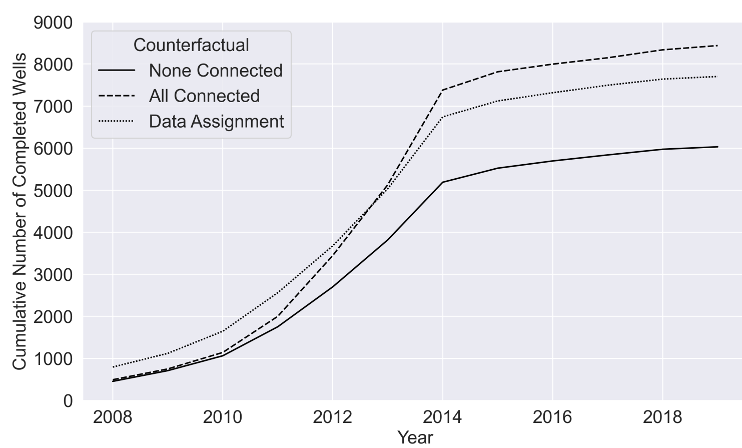
the significant impact of pipelines on the marginal cost of moving crude oil. Oil producers would complete 8438 wells if each prospect received a connection to a pipeline.

Nonetheless, the producers would delay the completion of some wells between 2008 and 2010 to wait for higher prices. Figure 1.3 shows that crude prices were healing from the 2008 financial crisis during these two years. Once the prices peak after 2011, drilling quickly ramps up and exceeds the baseline levels.

Figure 2.5: Counterfactual Investments



(a) Yearly Drilling Activity



(b) Cumulative Drilling Activity

Notes: The discount factor used across all the specifications is $\beta = .935$. The sample includes drilling prospects located in the Permian Basin discovered between 2008-2019.

Additionally, the construction of pipelines increases the prospects' average revenue and the extracted quantity of crude. In Table 2.2 the average well extracts 138,000 barrels of crude oil over her lifecycle when all the prospects received a connection to the pipeline. By contrast, the average well extracts 126,000 and 107,000 barrels of crude oil over her lifecycle in the baseline scenario and when none of the prospects received a connection to the pipeline, respectively.

This implies that producers delay the completion of more productive prospects to avoid large profit cuts due to high shipping costs. Indeed, oil producers have a larger option value from waiting to drill the more productive prospects. Therefore, the impact of pipelines on the industry's output and revenues is even larger when adjusted for the prospects' quantities. The transport infrastructure raises the drilling of more productive wells.

TABLE 2.2. Counterfactual Descriptive Statistics

			Data Baseline			
	N	Mean	Min	25%	75%	Max
Months for Completion	7703	44.879	0.000	25.000	64.000	84.000
Revenues	7703	12.252	0.000	1.402	17.360	114.375
Quantity	7703	0.126	0.000	0.018	0.183	1.028
			None Connected			
	N	Mean	Min	25%	75%	Max
Months for Completion	6031	42.024	0.000	20.000	62.000	84.000
Revenues	6031	10.111	0.000	1.345	13.099	105.443
Quantity	6031	0.107	0.000	0.018	0.145	1.028
			All Connected			
	N	Mean	Min	25%	75%	Max
Months for Completion	8438	41.642	0.000	20.000	60.000	84.000
Revenues	8438	12.934	0.000	1.762	21.408	140.396
Quantity	8438	0.138	0.000	0.022	0.226	1.419

Notes: The data covers drilling prospects in the Permian Basin. Quantities are measured in millions of crude oil barrels. Revenues are measured in millions of dollars.

However, the second counterfactual is somewhat a limit case. Connecting some of these prospects might be prohibitively expensive due to geographical constraints. A more realistic counterfactual would be to predict how the introduction of an infrastructure subsidy affects the constructions of pipelines. Then, based on the resulting pipeline connections I would simulate the upstream investment decisions.

To perform this exercise, I need to model the builders construction choices of connecting different oil prospects. The current model only accounts for a statistical dependence between the prospects' characteristics and the connection to the pipeline, without explicitly modeling the builders' decisions. In my future research agenda I plan to endogenize the builders' decisions to recover the cost of building pipelines. This allows me to simulate how a subsidy to infrastructure would affect the construction of new pipelines, and in turn drilling investments.

2.4 Conclusions

This papers advances the empirical understanding of the impact of the transport infrastructure on the firms' investment decisions. I show that building new infrastructure reduces the shipping costs for firms, raising the profitability of investments. Therefore, the upstream firms increase their investment levels in response to the new infrastructure. By contrast, the scarcity of transport infrastructure increases the firms' option value from waiting, causing substantial delays in investments.

I use the crude oil industry as my empirical laboratory. Due to the distance between supply and demand regions, the transport sector plays a crucial role in this industry.

I estimated a discrete model of investment and used it to evaluate the industry's response to new pipeline infrastructure. The model centers around the crude producers' incentives to delay drilling without a connection to a pipeline. I build a novel dataset that allows me to control for selection of the infrastructure on the well's productivity.

I find that pipelines have a substantial impact on the cost of shipping crude oil. In response, crude producers substantially curtail their investment activity if they can not access pipelines to move the crude oil. Specifically, I considered how investments in new wells would have changed if the new infrastructure wasn't built. The pipelines built between 2008 and 2019 were responsible for a 28% increase in the number of wells.

CHAPTER 3

The Investment Outcomes of Vertical Integration

3.1 Introduction

This article measures the impact of vertical mergers on investments. Vertical integration may solve hold-up problems for the integrating firm while exacerbating them for its downstream rivals (Grossman and Hart 1986, Bolton and Whinston 1991). Though integration could significantly affect the incentives to investment, empirical evidence on the investment outcomes of vertical mergers is still limited (Crawford et al., 2018). From an antitrust standpoint, quantifying these effects is important. How vertical integration affects investments, and its ultimate impact on welfare, has fueled the debate over recent and past mergers in the entertainment industry (e.g., Live Nation and Ticketmaster in 2010, AT&T and Time Warner in 2016).

Quantifying the investment effects of integration posits two major empirical challenges. First, it is difficult to isolate the input patterns between upstream suppliers and downstream distributors. Second, it is difficult to observe the measures of asset-specific investments. To overcome these challenges in this article I conduct a retrospective analysis of a series of vertical mergers and their investment effects both for the integrating firm and its rivals. The analysis is based on a comprehensive data set on the US motion picture industry, compiled from numerous sources.

The data set comprises project-level financial and technical information, the universe of companies involved in financing, producing and distributing each project, and firm-level ownership structure during the period 1997-2019. I use this unique data set to distinguish movies distributed by the production company's downstream counterpart from those distributed by the distribution divisions of unintegrated studios. It is crucial for the analysis that the data include the production budget of each project, which measures quality-enhancing investments that increase the movies' ability of capturing demand.

My focus on the US motion picture industry is motivated by several factors. First, the contracts involved in the production of movies offer a unique opportunity to conduct this analysis. The industry features upstream producers investing in the production budgets of multiple projects while securing long-term contracts with downstream distributors that implement marketing campaigns and secure exhibition slots. This structure resembles the multilateral settings studied in Bolton and Whinston (1991). In their framework, multiple downstream distributors compete to trade with an upstream supplier. Second, there is significant ownership variation across firms and time. The research design benefits from this variation, allowing me to implement a staggered difference-in-difference. Third, the detailed data on project-level production expenditures allows me to directly measure the changes in investments.

Moreover, I can distinguish the movies distributed by the integrating distributor from those distributed by its rivals following a vertical merger. This provides me enough power to split the data in two samples and separately identify the causal impact of mergers on investments both for the integrating counterpart and its rivals. In this article I focus on the upstream investment decision problem that is, the choice of how many resources to allocate to movies' production budget. I find that on average vertical mergers increase investments in movies distributed by the integrating firm by \$75 million, a 72% increase compared to the pre-merger scenario. By contrast, vertical integration reduces investments for the integrating firm's rivals by \$21 million, a 37% decrease compared to the pre-merger scenario.

Several factors can explain the causal effect of vertical mergers on investment. The most relevant from an antitrust standpoint is the change in marginal returns from investment.

This would support the property right theory and inform the analysis of future mergers. Indeed, as long as contractual incompleteness holds, industries experiencing vertical integration should feature investment outcomes analogous to the ones estimated in this article. However, vertical mergers could also relax the credit constraints for the upstream production company. Distributors are generally part of large entities, called studios, which have easy access to equity and debt markets. Therefore, vertical integration could reduce the cost of capital for a stand-alone company. This might conflate with the hold-up explanation. Lastly, industry-specific technological factors could magnify the effects of integration in movie investments.

An important contribution of this article is to speak to the mechanisms behind these outcomes. My goal is to separate the financial frictions and industry-specific technology from the changes in the marginal return on investment after integration. To this end, I build a within-firm resource allocation model in the spirit of Giroud and Mueller (2019) who illustrated how financially constrained firms operating in multiple regions allocate internal resources in response to regional shock. In my article, the purpose of the model is threefold. First, it provides testable empirical implications to rule out the relevance of financial constraints. Second, it specifies an empirical production function of movie tickets, whose out-of-sample estimation measures the role of the technological parameter. After estimating the movie tickets technology, I back out the implied change in the revenue internalization rates. This allows me to separate the role of industry-specific technology from the change in the internalized marginal returns from investments.

The main endogeneity threat in the production estimation comes from the unobserved productivity of upstream production companies. I leverage the movie-level covariates to sidestep the restrictive assumption of time-invariant productivity. Instead of recovering the unobservable productivity by inversion, such as in Olley and Pakes (1996), I model it as a function of companies' fixed effects and time-varying movie characteristics. Formally, the key identifying assumption is that the unobserved productivity term is separable and additive in the production company's fixed effect and movie-specific covariates. This assumption identifies the production function that transforms the production budget into movie tickets. Lastly, I replace the production function estimates and the difference-in-difference reduced-form estimates into investment first-order conditions. Inverting this equation I recover the implied change in revenue internalization rates when production companies are making investment decisions.

The change in revenue internalization rates is the key mechanism responsible for the effects on investments according to the property right theory. I estimate that vertical integration increases the ex ante internalized returns by 37% when the upstream production company invests in projects distributed by its downstream counterpart. By contrast, vertical integration reduces the internalized returns by 14% for movies distributed by its downstream rivals. This change in internalized marginal returns from investments explains the previously quantified effects of vertical integration on production budget.

These results suggest that the integrating firm solves hold-up concerns while reducing investments that would benefit its downstream rivals. My results highlight the importance of measuring the investment outcomes of integration when conducting merger evaluations. The impact of a vertical merger on investments is large and, most importantly, driven by a change in internalized marginal returns from investments. This implies that long-term contracts fall short in substituting firm boundaries. As long-term distribution contracts are pervasive of many industries, the patterns of the motion picture industry are likely to arise in mergers consummated in other industries.

Related Literature. Despite the voluminous literature examining the reasons of integration, few articles have explored its outcomes (Crawford et al., 2018). Most of these studies (e.g., Chipty (2001), Hastings (2004), Ciliberto (2006), Gil (2007), Villas-Boas (2007), HortaÃ§su and Syverson (2007), Mortimer (2008), Lee (2013), Atalay et al. (2014), Crawford et al. (2018), Luco and Marshall (2018), and Yang (2020)) focus on the effects of vertical integration on output prices, holding investments constant. A notable exception is Ciliberto (2006) who focused on investment decisions of hospitals that vertically consolidated the provision of healthcare services. Lastly, Yang (2020) built a structural industry model to evaluate the impact of integration on innovation via simulations in the system-on-chip industry.

My contribution to the literature is twofold. First, I measure the effects of vertical integration on investments in multilateral settings. Second, I provide direct empirical evidence on the mechanisms through which integration operates. Specifically, ownership affects investment, changing the ex ante internalized marginal return. This mechanism is not industry specific, but potentially applies to many industries and markets, and lies in the contractual primitives of the economy. Therefore, these results directly inform the antitrust debate on whether the ownership of content by distributors harms welfare in recent mergers in the cable TV industry. Moreover, I highlight the importance of accounting for the whole network of firms trading with a specific supplier.

Finally, I contribute to the strand of literature that studies the motion picture industry (e.g., Corts (2001), Einav and Orbach (2007), Gil (2007), Palia et al. (2008), Gil (2008), Gil (2015), Waldfogel (2017), Caoui (2019)). In terms of vertical arrangements, this literature mainly focuses on the relationships between theatrical exhibitors and distributors. Gil (2008) studies how firm boundaries affect prices and efficiency focusing on the relation between movie theaters and distributors. His main metric of interest is the length of movie run. One exception is Corts (2001), who studies how vertical structures between studios and production companies affect movie release dates, specifically, how vertically integrated distributors strategically determine the release dates to minimize profit cannibalization. The author interprets his results in light of the common agency theory elucidated in Bernheim and Whinston (1985). I contribute to this literature by analyzing how production budgets change as a consequence of the integration between studios and production companies. My focus on the production and distribution side, together with revenue and budget data, allows me to quantify the mechanisms explaining the effects of integration.

3.2 Institutional Detail and Data

This section describes the structure of the US motion picture industry and the contractual features that open the way for vertical mergers to shape investment patterns. It then summarizes the data at hand.

3.2.1 Industry Structure

The industry is composed of two types of firms: stand-alone *production companies* involved in the development and production of movies, which represent the *upstream* side of the industry, and *studios* responsible for distribution and marketing campaigns. However, studios are large corporations that generally own several production units together with their distribution and marketing divisions. In contrast, stand-alone production companies do not have a distribution division, and sign long-term distribution deals with studios to distribute their projects. This article focuses on the investment outcomes of vertical mergers between studios and stand-alone production companies.

Figure 3.1: Industry structure: Pre-Merger

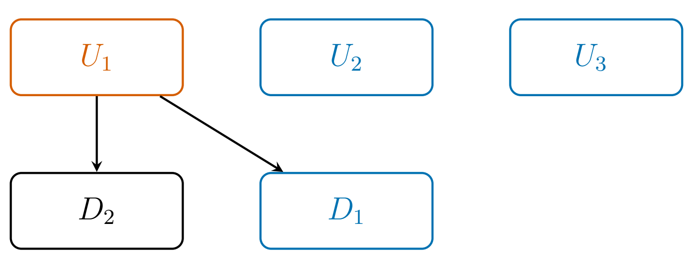
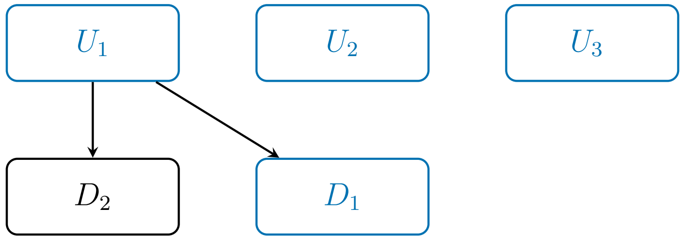


Figure 3.2: Industry structure: Post-Merger



To fix ideas, Figures 3.1 and 3.2 describes how these vertical mergers affect the industry structure. Specifically, I focus on scenarios where a stand-alone company, U_1 in the example, had signed long-term distribution deals with one or more studios before the merger. These studios each operate a downstream distribution unit; in the example the blue studio owns D_1, and the green studio, D_2. This situation is described in 3.1, where the arrows indicate U_1 is distributing movies through D_1 and D_2. The blue studio is already controlling two upstream production divisions, U_2 and U_3. This article studies how investments in projects produced by a stand-alone company, U_1, change after the integration with a studio, D_1. It quantifies the impact of the vertical merger on investments for movies distributed by D_1, which is part of the integrating firm, and by D_2, which directly competes with the downstream division D_1. The new industry structure is represented in 3.2.

To clarify the timing and the type of investments I describe the value chain of movies. Before theatrical exhibition, each movie goes through a sequence of distinct stages. The following description holds with generality for movies with a production budget above \$5 million. [1] During the development phase a production company acquires the intellectual property rights over a "source". The latter might be a comic character, a book or a video game. The company hires a writer, or a team of writers, to create a screenplay based on the source. Once the screenplay takes shape, it is matched to a creative team that selects the primary cast members, the line producer and the director. In the meanwhile, the projected budget is accurately compiled, net of the marketing expenditures. The screenplay, projected

[1] Low-budget movies rely on peculiar funding sources and go through heterogeneous bargaining protocols. For this reason I exclude them from my analysis.

budget and creative team form the "package". This is the essential input to produce the final good. The production company finances the project up to this point. Development expenditures range from 5% to 20% of the total production budget.[2]

After the development stage, substantial funding is needed to bring the movie into production. Corts (2001) and Palia et al. (2008) analyze contractual efficiency in the US movie industry. In their analysis, they provide an exhaustive description of the basic financing arrangements. Aggregating these forms to a relevant level for this article, two main scenarios arise. On the one hand, production funding is uniquely provided by the production company. On the other hand, the movie is co-financed by the production company and the studio. In the first scenario, the quality enhancing-investments are incurred only by the upstream company, whereas in the second scenario both the downstream and upstream firms invest in content quality. Once financing has been secured, the project enters the production phase, which encompasses both filming the movie and post-production activities such as editing and creating visual effects. Each title's production budget encompasses expenditures incurred up to this point.

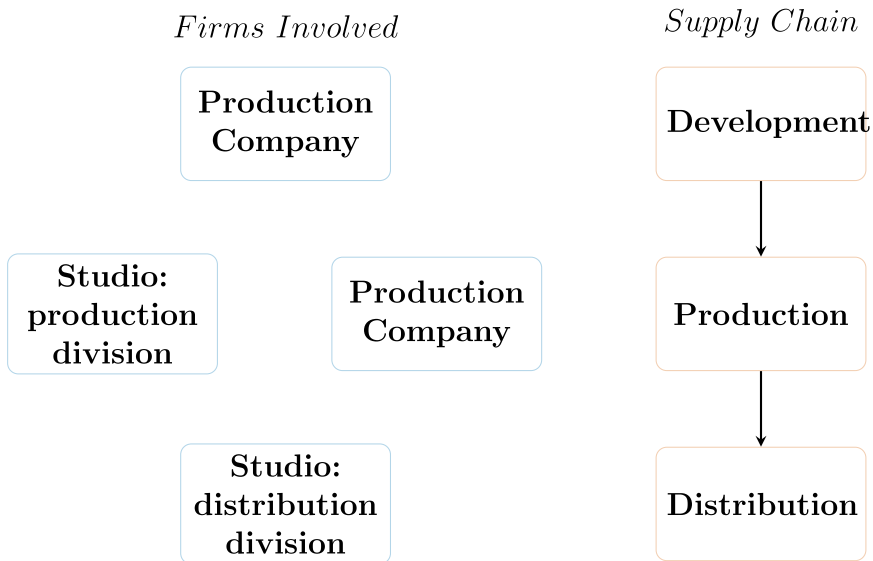
The very nature of movies as products clarifies why these expenditures are quality-enhancing. Adding more features, and therefore spending more on production, can make a movie more appealing to an audience[3]. Importantly, marketing and advertising costs do not enter the gross production budget. These expenditures are concentrated early in the release cycle, after production, and they are sustained only by the studio's distribution arm, which at the same time bargains with exhibitors to secure screens and favorable showtime. These stages and the firms involved are described in Figure 3.3.

Two contractual features create the basis for vertical mergers to affect investments. First, production companies sign long-armed *distribution agreements* with studios. The contracts

[2] *The Guardian: Anatomy of a blockbuster, 2004*; I corroborated this statistic by examining data from several movies whose budget breakdowns are publicly available due to the Sony data breach lawsuit, such as *Sahara* and *Pixels*

[3] In Appendix A I extensively discuss the extent to which the production budget measures quality-enhancing investments.

Figure 3.3: Supply Chain



cover a slate of movies during a well-defined time window[4]. Potentially, both parties benefit from the long-term nature of the contract. Production companies secure a distribution outlet, which in turn helps them to raise funding from external investors. Studios secure a steady flow of inputs, which avoids disruption or delays in production. However, due to the contracts' length, the specific projects that fall under the contracts' umbrella are not specified. Usually, only the quantity of movies to be supplied downstream is clearly defined. Thus, specific investments are ex ante unknown. Concretely, it is impossible to determine the best leading actor for a role without a finished screenplay. This prevents firms to write complete contracts that include ex ante the amount of resources a studio invests in exchange for quality-enhancing actions. Out of necessity, the production budget is not ex ante contractible but determined ex post in the project-specific development stage.

The other contractual feature relevant for investment outcomes is profit sharing. The production company and the distributor jointly hold rights over the gross profits of the movies within the distribution deal. Under the so-called net deal the distributor collects a percentage of the gross rental, to recoup its advertising expenditures from the remaining revenues before

[4]DreamWorks Animation and 20th Century Fox sealed a distribution deal on August 18, 2012, with an "output term of 5 years".

distributing the net proceeds to the production company (Corts (2001)). Sellout contracts allocating the full residual profit claim to a single party are virtually nonexistent. Two main reasons undermine the two-part tariffs in this industry. First, these contracts are ineffective to alleviate double-sided moral hazard. The latter arises when investments take place on both sides of the market, such as in the motion picture industry, where the distributor incurs marketing expenditures whereas the production company invests in the production budget. Second, franchise fees have significant drawbacks when firms are risk adverse and the final demand is uncertain. In a "nobody knows world" [5], such as the motion picture industry, this concern has a solid foundation. Therefore, the marginal returns from quality-enhancing investments are never allocated to a single company. By contrast, vertical integration makes the studio the only residual claimant over its investment. Consequently, the incentives to invest may drastically increase for the integrating studio.

These features of the contractual environment posit the theoretical basis for two outcomes and justify the analysis that follows. Specifically, integration could solve hold-up problems for the integrating firm, raising investment levels. However, during the unwinding of long-term distribution deals between a studio and a production company, a rival studio might acquire the upstream production company. After the merger, the production company directly competes against the unintegrated distributor in the output market. Thus, the production company might degrade the input quality, benefiting downstream rivals by reducing investments.

3.2.2 Data

I divide the core of my data into two broad categories: movie-financials, which measure expenditures and revenues associated with each project, and firm-financials, which include each company's ownership-organizational structure, entry year, and the activity carried out within the production and distribution of each movie. Additionally, the data set contains

[5] Waldfogel (2017)

detailed project-level technical details.

Opus Movie Database. The *Opus Movie Database* is the backbone of the data set. It spans 21,000 movies released or re-released in theaters from 1997 to the beginning of 2019. Among these titles 13,254 were released in the US. Movies' financials are the first key information contained in the *Opus Database*. For each movie the database includes box office revenues (domestic and international[6]), domestic DVD sales with associated revenues and the project's production budget. This information is complemented by detailed technical characteristics of each film. The most important for this article are the MPAA ratings and a sequel indicator, because they proxy project-specific risk. For the sake of exposition, I describe other technical attributes only when they are directly included in the following analysis. Table 3.1 provides summary statistics for movies with a production budget greater

TABLE 3.1. Opus Movie Database: Descriptive Statistics

	N	Mean	SD	Min	Max
Production budget ($ million)	2692	50.99	50.75	5.00	425.00
Domestic box office ($ million)	2692	66.13	85.36	0.00	936.66
International box office ($ million)	2692	88.91	153.30	0.00	2015.84
DVD & Blu-ray units (million)	1810	1.75	2.57	0.00	24.11
DVD & Blu-ray revenues ($ million)	1446	29.04	47.68	0.02	548.89
Opening weekend theaters	2692	2216.13	1349.89	1.00	4662.00
Maximum theaters	2692	2417.98	1182.04	1,00	4662.00
Running time (minutes)	2586	110.07	18.12	42.00	201.00
IMAX	2692	0.08	0.28	0.00	1.00
Sequel	2690	0.13	0.33	0.00	1.00
Wide	2692	0.77	0.42	0.00	1.00
G-rated	2688	0.02	0.14	0.00	1.00
PG-rated	2688	0.15	0.36	0.00	1.00
PG13-rated	2688	0.41	0.49	0.00	1.00
R-rated	2688	0.40	0.49	0.00	1.00

Notes: The sample includes data on movies released in theaters from 1997 to 2019 and having a production budget above $5 million. Production budget, box office data and DVD revenues are in millions of dollars. Production budget does not include marketing costs and advertising expenditures incurred to promote the movie.

than or equal to $5 million. This subset adds up to 2,692 movies. The average production

[6] In what follows the attribute "domestic" refers to US data, and "international" refers to other territories.

budget and domestic box office are \$56 and \$71.62 million, respectively.

The second key piece of information included in the *Opus Database* is the US distributor and the companies involved in financing, developing and producing the movie. Three sets of dummies allow me to disentangle the exact role of each company. The "financing" indicator identifies the companies that funded the movie. Additionally, the "production" dummy isolates those firms that carried out the physical shooting of the movie. Finally, co-financiers are recorded with the "in association with" flag set. This set of information falls within firm-financials, and allows me to match the studio's distribution branch with the upstream companies investing in the project.

Ownership Data. In this article, the research design leverages the time variation of ownership configurations to identify the investment outcomes of vertical mergers. I focus on firms that financed at least two projects with a budget greater than or equal to \$5 million dollars. The set of companies in my data spans 106 distribution arms and 151 production companies, for a total of 257 entities. However, the previous count conflates different subsidiaries under the same company with independent firms. Indeed, the *Opus Database* displays only the names of companies involved in each project. In order to fill this gap, I manually reconstructed the ownership of each company mainly from the *Orbis Database*. I was able to match each firm to its parent company. The latter is the corporate that the *Orbis Database* credits as directly holding the majority of shares. Then, I matched the parent companies with their ultimate global corporate owner. When possible, I corroborated this information with publicly available information such as companies' SEC filings. With this data at hand I observe the production companies that underwent a vertical merger between 1997 and 2019. Figure 3.4 displays all the vertical acquisitions consummated between 1997 and 2019 in the US motion picture industry, with the acquired upstream companies and their respective buyers. For instance, Pixar operated as a stand-alone company until 2005 and became vertically integrated with The Walt Disney Company in 2006 in a \$7.4 billion deal. This article ignores partial vertical integration and vertical disintegration, and focuses on

Figure 3.4: Vertical Mergers Timeline

Buyer/ *Target*	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Lionsgate																							
Mandate Picture	.	.	.	.	x	x	x	x	x	x	x	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
Summit Entertainment	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	100	100	100	100	100	100	100	100
Roadside Attraction	.	.	.	.	.	.	x	x	x	x	45	45	45	45	45	45	45	45	45	45	45	45	45
Paramount Pictures Corp				**CBS**												**Viacom**							
MTV Films	x	x	x	x	x	x	x	x	x	x	100	100	100	100	100	100	100	100	100	100	100	100	100
NBCUniversal		**Vivendi**						**General Electric**								**Comcast Corporation**							
Amblin Partners	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	x	x	<25	<25	<25
Dreamworks Animation	.	.	.	.	.	.	.	.	.	x	x	x	x	x	x	x	x	x	x	x	>50	>50	>50
Rogue Pictures	100	100	100	100	100	100	100	100	100	100	100	100	x	x	x	x	x	x	x	x	x	x	x
Relativity Media																							
Focus (distribution)	x	x	x	x	x	x	x	x	x	x	x	x	x	100	100	100	100	100	100	100	100	100	100
Twenty-First Century Fox																							
New Regency	x	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
The Walt Disney Company																							
ImageMovers	x	x	x	x	x	x	x	x	x	x	x	>50	>50	>50	>50	x	x	x	x	x	x	x	x
Lucasfilm ltd.	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	>50	>50	>50	>50	>50	>50	>50
Marvel Studios	x	x	x	x	x	x	x	x	x	x	x	x	x	x	>50	>50	>50	>50	>50	>50	>50	>50	>50
Pixar	x	x	x	x	x	x	x	x	x	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
Spyglass Entertainment	x	x	x	10	10	10	10	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Notes: Production companies that experienced a vertical merger between 1997 and 2019. In each cell there is the ownership share of the downstream studio for the row-specific developer. The years when the developer operated as independent are marked by an "x". By contrast, "." indicates those years when the developer had not been founded yet. Majority shares are reported with "> 50" if the Orbis Database does not provide the exact percentages. *Relativity Media* acquired *Focus/Rogue*'s distribution assets and started to release movies to theaters. As a result, the underlying production arm became vertically integrated.

vertical mergers where the downstream studio acquires more than 50% of the shares of the production company. Majority deals grant the acquiring firm full control over the upstream asset. Therefore, I retrospectively analyze the investment outcomes of eight vertical mergers. [7]

Based on ownership and input patterns, I divide the data as follows. First, I select movies financed by production companies that underwent a vertical merger during the time window covered by the data. Then, I divide post-merger movies distributed by the integrating studio (in-house movies) from those distributed by unintegrated studios (outward movies). This procedure generates two distinct subsamples. The first subsample includes *in-house post-merger movies* and pre-merger movies (*in-house sample*). Pre-merger movies are those produced in an "unintegrated" fashion because in absence of a distribution arm, the production company had to sign distribution deals with studios.

[7] The final list of majority vertical mergers analyzed in the article is Lionsgate-Mandate, Lionsgate-Summit, Paramount-MTV, Relativity-Focus, Disney-ImageMovers, Disney-Lucas, Disney-Marvel and Disney-Pixar.

I excluded these mergers from the in-house sample for the following reasons. Mechanically, Mandate kept distributing movies only through different distributors, falling in the outward category. I excluded Relativity to maintain homogeneity in the economic rationale behind integration and the interpretation of results. Relativity differs from the other vertical mergers because it internally integrated by opening its own distribution branch. Therefore, there was not a potential hold-up issue between an existing studio and an upstream producers that called for a reallocation of property rights between two existing firms.

Finally, the second subsample includes *post-mergers outward movies* and pre-merger unintegrated movies (*outward sample*). The three mergers for which I observe post-merger outward movies are Marvel, which had deals with Sony, Paramount, and Universal, Relativity Media and Mandate. Mechanically I had to exclude other production companies, such as Pixar, because they kept distributing solely with the integrating studio. Despite the altogether foreclosure of rival studios leading to potentially anticompetitive outcomes per se, I lack an appropriate counterfactual to measure these effects.

The *in-house sample* includes a total of 133 titles, and the *outward sample* includes 98 titles. I augment each subsample with a group of "never treated" titles, i.e., movies originating from production companies that did not change ownership over time (either they had always been vertically integrated or were unintegrated from 1997 to 2019). I provide more detail on the selection criteria in the following section. Table 3.2 concludes this section, displaying the summary statistics of the two augmented subsamples.

TABLE 3.2. Augmented In-house and Outward Sample, Descriptive Statistics

	In-house Sample				
	N	Mean	SD	Min	Max
Production budget ($ million)	621	74.84	61.18	5.00	356.00
Domestic box office ($ million)	621	104.49	117.50	0.01	936.66
International box office ($ million)	621	150.36	202.84	0.00	1842.81
DVD & Blu-ray units (million)	468	2.63	3.24	0.00	24.11
DVD & Blu-ray revenues ($ million)	385	44.46	55.96	0.08	341.51
Opening weekend theaters	621	2744.83	1209.02	0.00	4662.00
Maximum theaters	621	2923.98	967.69	0.00	4662.00
IMAX	621	0.18	0.38	0.00	1.00
Screenplay	621	0.51	0.50	0.00	1.00
Sequel	620	0.20	0.40	0.00	1.00
Wide	621	0.88	0.33	0.00	1.00
PG-rated	620	0.17	0.37	0.00	1.00
Vertical mergers	6				
	Outward Sample				
	N	Mean	SD	Min	Max
Production budget ($ million)	345	66.59	54.56	5.00	275.00
Domestic box office ($ million)	345	88.07	89.94	0.04	533.72
International box office ($ million)	345	120.44	158.02	0.00	960.50
DVD & Blu-ray units (million)	276	2.17	2.75	0.00	22.28
DVD & Blu-ray revenues ($ million)	216	39.87	57.66	0.02	548.89
Opening weekend theaters	345	2723.22	1144.38	2.00	4404.00
Maximum theaters	345	2905.43	909.19	5.00	4404.00
IMAX	345	0.13	0.34	0.00	1.00
Screenplay	345	0.56	0.50	0.00	1.00
Sequel	345	0.16	0.36	0.00	1.00
Wide	345	0.87	0.33	0.00	1.00
PG-rated	345	0.12	0.32	0.00	1.00
Vertical mergers	3				

Notes: Both samples include data on movies released in theaters from 1997 to 2019 and having a production budget above $5 million. Production budget, box office data and DVD revenues are in millions of dollars. The production budget does not include marketing costs and advertising expenditures incurred to promote the movie. The Marvel acquisition belongs to both subsamples, though two disjoint subsets of projects are included in the post-merger scenario.

3.3 The Causal Effect of Vertical Mergers on Investments

This section estimates the causal impact of vertical acquisitions on upstream investments. I show that after vertical integration, the production budget substantially increases for movies produced by the integrated company and distributed by its downstream counterpart. Conversely, movies distributed by rival studios experience a sharp and persistent decrease in their production budget after vertical integration. I estimate a difference-in-difference model which identifies the average treatment effect of vertical mergers on investments. I separately run it on the *in-house* and *outward* sample. Thus, the *in-house* sample identifies changes in investments benefiting the integrating unit, whereas the *outward sample* identifies changes in investments affecting the integrating unit's rivals. Identification posits two main challenges. First, integrated production companies might be on a different investment trajectory than the full population. To address this challenge, I match each upstream production company that underwent vertical integration to a placebo firm that did not change ownership. Second, vertical mergers are endogenous. To mitigate this concern, I show that outcomes diverge only after the merger.

A preliminary step to implement the research design is defining an appropriate control group. I leverage the large number of upstream companies that did not change ownership to identify the counterfactual. The selection procedure isolates a control company, minimizing the distance on average investments between the assigned treated company and the control group. This procedure is described in detail in Appendix B.3. After selecting the control group for each sample, I specify a movie-level model where the passing of calendar time is pinned down by movies' production years. Specification 3.1 measures the dynamics of the effects, while at the same time probing the validity of the research design by testing whether there appears to be any effect of vertical mergers before the acquisition actually occurs [8].

$$Y_{mjt} = \sum_{k=-l}^{l} \beta_k^{All} \mathbb{I}_{\{L_{mjt}^{All}=k\}} + \sum_{k=-l}^{l} \beta_k \mathbb{I}_{\{L_{mjt}=k\}} + \zeta_t + \zeta_j + \beta_z \mathbf{Z}_{mjt} + \epsilon_{mjt}. \tag{3.1}$$

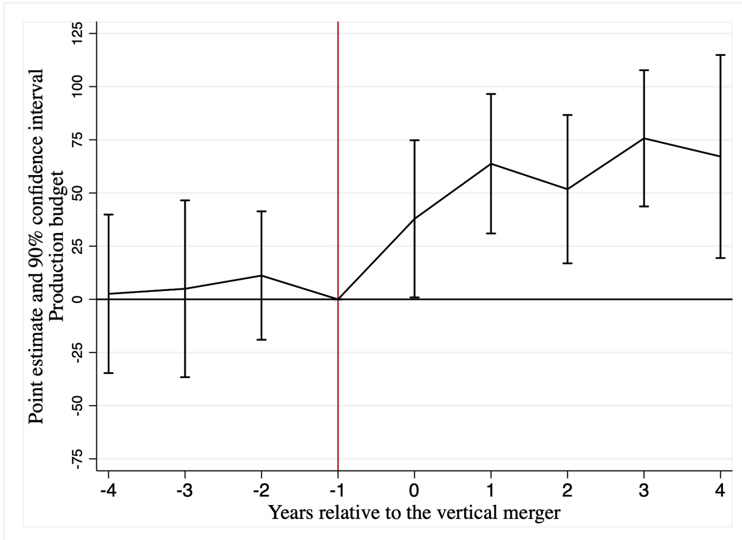
Here Y_{mjt} is the production budget of movie m, financed by the upstream company j in year t. This outcome variable is a function of a full set of leads and lags around each acquisition, $L_{mjt} = t_m - M_j = k$, where M_j is the year company j became vertically integrated and t_m is the production year of movie m produced by the treated firm. The specification includes a set of leads and lags $L_{mjt}^{all} = t_m - M_j = k$, where M_j is the year of the vertical integration

[8] This specification follows Jaravel et al. (2018)

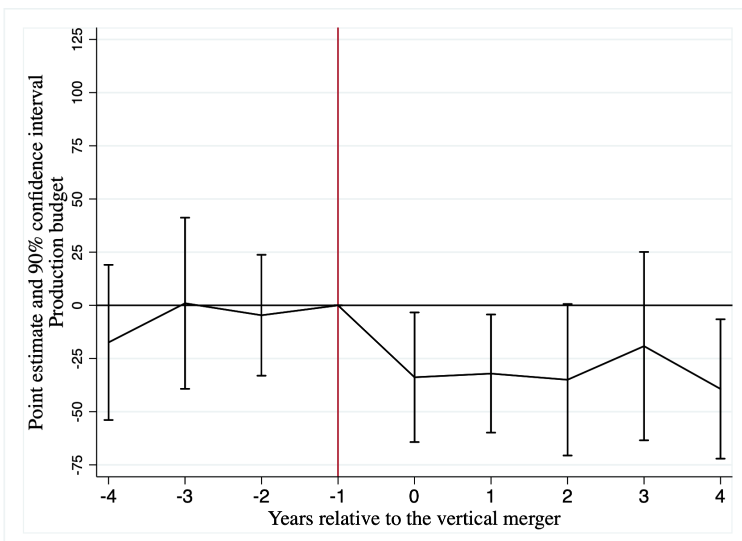
of company j and t_m is the production year of movie m produced by the *treated firm* and its *placebo firm*. The additional leads and lags serve the purpose of better capturing the time trends common to the treatment and control groups around the merger years. Moreover, following Palia et al. (2008), I add a sequel indicator and MPAA rating control for project-specific risk. The indicators for the production method and opening pattern complete the set of covariates. These variables are denoted by $\mathbf{Z}_{mjt}$. Lastly, ζ_t and ζ_j denote production year and production company fixed effects, respectively. The coefficients of interest are $\{\beta(k)\}_{k=-l}^{l}$. The effects of interest are identified as long as $E[\mathbb{I}_{\{L_{mpt}=k\}}\epsilon_{mpt}|X_{mpt}, t, j] = 0$.

Figures 3.5a and 3.5b report the point estimate and the 90% confidence intervals for the coefficients β_k obtained from specification (1). The lack of a pre-trend for any of the subsamples supports the identifying assumption. The set of coefficients drastically differ in the two samples. Following the merger, projects distributed by the integrating unit experience a sudden and persistent increase in their production budget compared to pre-merger levels. By contrast, after the merger projects distributed by the integrating unit's rivals experience a sudden and persistence drop below pre-merger levels.

Figure 3.5: Identification Assumption and Dynamic Effects



(a) Inhouse pre-trends



(b) Outward Sample

Notes: Production budget is expressed in million dollars. The red vertical line is the year before the production company's takeover. The coefficient $\beta(-1)$ is normalized to zero. Standard errors are robust to heterogeneity.

To quantify the average impact of vertical mergers on investments, I employ a second specification with a dummy that becomes 1 after a vertical merger. Specifically,

$$Y_{mjt} = \beta^{All} AfterMerger^{All}_{mjt} + \beta AfterMerger_{mjt} + \zeta_t + \zeta_j + \beta_z \mathbf{Z}_{mjt} + \epsilon_{mjt}. \tag{3.2}$$

Under the identification assumption, β captures the average causal treatment effect of vertical mergers on the production budget. The *AfterMerger* dummy's coefficient accounts for a maximum of 10 years following the acquisition[9]. Vertical integration significantly raises investments benefiting the integrating firm.

The investments in the production budget increase on average by \$72.4 million for projects distributed by the integrating studio compared to pre-merger levels, as reported in column (3) of Table 3.3. Comparing the estimated effects with the average budget of projects produced before the mergers shows the economic relevance of these estimates. Indeed, the outcome corresponds to an average increase of 73.5% compared to the pre-merger level. Conversely, vertical mergers negatively affect investments in projects distributed by rival studios. The investments in the production budget decrease on average by \$26.6 million for projects distributed by the integrating unit's downstream rivals. This corresponds to an average decrease of 47% compared to the pre-merger level. The effects are robust across the specifications of Table3.3, bolstering confidence in the magnitude and economic importance of the results.

The increase in investments for the in-house sample is consistent with the internalization of larger profits by a vertically integrated firm. The theoretical foundation for these effects lies in Hart and Moore (1990), Bolton and Whinston (1991) and, in general, the property right theory. Additionally, these outcomes support a recent article by Yang (2020), who studies complementary innovation in vertical industries. Here, integration leads to a joint maximization problem that solves under-innovation stemming from firms' failure to internalize complementarities when separated. Counterfactual simulations suggest that integration increases coordinated investments. In the movie industry, upstream producers invest in product quality (production budget), and downstream distributors complement the investment with expensive advertising campaigns (which in many instances exceed the production budget). Both investments are key for the project success. Therefore, my results complement

[9] Appendix B presents the short-term outcomes computed restricting the leads to a maximum of five years after the acquisition.

TABLE 3.3. Causal Effects of Vertical Mergers on Production Budget ($ million)

Dependent Variable:	Production Budget ($ million)					
	(1)	(2)	(3)	(4)	(5)	(6)
Specification	In-house	In-house	In-house	Outward	Outward	Outward
AfterMerger	73.400***	75.206***	72.354***	-19.203**	-21.136**	-26.609**
	(20.293)	(20.913)	(22.847)	(8.895)	(8.470)	(7.586)
AfterMerger (All)	✗	✓	✓	✗	✓	✓
Time-varying controls	✗	✗	✓	✗	✗	✓
Company FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Mean pre-merger outcome	105.263	105.263	105.263	88.600	88.600	88.600
Observations	622	622	620	345	345	345
Adjusted R-squared	0.435	0.435	0.489	0.258	0.258	0.358

Notes: Results are computed using all available observations. Standard errors in parentheses; *** p<0.01; ** p<0.05;* p<0.1. Standard errors are block-bootstrapped at the production company level with 200 replications. Each specification uses movies' production budget as the dependent variable expressed in million dollars. The *AfterMerger* dummy includes 10 leads after the year of merger.

and support Yang (2020), providing a retrospective analysis and data from vertical mergers consummated in the entertainment industry.

The reduction in the production budget instead resonates with supply assurance concerns. Specifically, it speaks in favor of a change in competitive incentives in line with the theory of Bolton and Whinston (1991). The reduction in investments decreases the attractiveness of projects distributed by the integrating unit's rivals, softening competition for the integrating firm on the output market.

At the same time, entering a studio can alleviate the credit constraints for the upstream producer. Studios are diversified firms with access to equity and bond markets and a vast amount of pledgeable collateral. This mechanism conflates with the change in profit internalization, complicating the interpretation of the increase in investments. Similarly, the higher debt burden together with financial frictions could be partially responsible for the drop in the production budget. This is a relevant threat to the interpretation, in particular in light of the fact that a large part of the variation in the outward sample comes from Relativity Media, who vertically integrated by opening a distribution branch. The initial debt burden might curtail initial investments, conflating with the difference-in-difference estimates. Therefore, in what follows I partially mitigate these concerns by explicitly testing the relevance of credit constraints and isolating other competing mechanisms explaining the outcomes of interest.

3.4 A Model of Within-firm Resource Allocation

This section presents a model of optimal within-firm resource allocation to illustrate how ownership configurations, technology and financial constraints affect upstream investments. First, I describe the industry structure where firms make investment decisions. There are a set $\mathcal{N}$ of upstream production units, $\{U_n\}_{n=1}^{N}$, and a set $\mathcal{M}$ of downstream distribution units, $\{D_m\}_{m=1}^{M}$. The set of active firms in the industry, $\mathcal{F}$, can be partitioned in two categories, studios and stand-alone companies. Each studio controls a subset $\mathcal{J}$ of upstream production units, $\{U_n\}_{n\in\mathcal{J}}$, and has access to the output market through the ownership of a distribution unit D_m. [10] In contrast, stand-alone companies operate only a single upstream production unit U_n. To access the output market, a stand-alone company must sign a distribution deal with the studio.

I focus on the upstream investment decision problem. That is the choice of how many resources to allocate to movies' production budget. The problem is static, and input choices are centralized at the *firm* level. At the beginning of each period t, each production unit is endowed with resources C_n. The firm pools resources and reallocates them among production units. I assume that the divisions along the supply chain do not compete for resources. This subsumes that production budget investments do not crowd out resources to advertisement expenditures. The assumption is justified by the complementarity of the two investments.[11]

[10] In reality, studios contract with movie theaters and obtain a cut of the box office. In this article, I ignore the strategic price interaction of motion picture chains. This is motivated by the industry practice to split the tickets' revenue in half between studios and motion picture chains. Bargaining influences mostly the number of opening screens, which I observe in my data and can control for. In the difference-in-difference specification, the opening indicator proxies this margin. In the robustness check section I will directly control for the number of opening theaters, to further neutralize this channel.

[11] A movie with a very high production budget and no marketing campaign is likely to suffer large losses. Similarly, large expenditures in marketing must be backed up by the high quality of the movie (reflected in the production budget).

The industry produces a homogeneous good, movie tickets, using a technology that transforms the production budget L_n invested in t into $t+1$, according to

$$Y_n = f_n(L_n) = e^{\omega_n}(L_n)^{\alpha}, \tag{3.3}$$

where the time index is omitted in virtue of the static nature of the model. Here ω_n is the unit's productivity and α is the constant investment elasticity of production. In the following section I provide an additional micro-foundation to this functional form. Revenues from tickets' sales are expressed as $pf_n(L_n)$, where p is the price of a movie ticket. To simplify notation, each production unit U_n produces one movie per period. Extending the model to accommodate a production unit to produce multiple projects does not affect the investment choices, as long as the budget constraint and resource allocation remain at firm level. Therefore, L_n can be interpreted as the production budget of a movie.

Contractual features enter as an ownership-dependent *primitive* of the model. The ownership status of the upstream production company affects the firm's revenue internalization rate, and in turn this affects the input allocation choices. Specifically, the internalized revenues are $\tau(pf_n(L_n))$, where $\tau \in (\{\tau(v), \tau(s)\})$, and v and s indicate the ownership status of the upstream production unit, vertically integrated or separated, respectively. The parameter τ can be rationalized as summary statistics of rent-seeking behavior, complementarities of investments and anticompetitive incentives. In presence of hold-up problems when making investment decisions, or anticompetitive incentives, I expect τ to be lower than one. Similarly, if integration solves the hold-up issues for the integrating firm, τ increases, because the firm appropriates all the returns from investments.

Firms invest in production units to maximize profits. A stand-alone company f controls a single production unit U_j, and thus produces only one movie in the baseline model, solving

$$\max_{L_j} \ \tau[pf_j(L_j)] - L_j$$
$$s.t.$$
$$L_j \leq C_f,$$

where $\tau = \tau(s)$. Denote λ the Lagrange multiplier associated with the budget constraint. The firm maximization problem is equivalent to

$$\max_{L_j,\lambda} \ \tau[pf_j(L_j)] - L_j + \lambda[C_j - L_j]. \tag{3.4}$$

The Kuhn-Tucker conditions are

$$\tau(s)pf_j'(L_j) = (1 + \lambda) \tag{3.5}$$
$$L_j \leq C_j \tag{3.6}$$
$$\lambda[C_j - L_j] = 0; \quad \lambda \geq 0. \tag{3.7}$$

When the resource constraint binds, $\lambda > 0$, then $L_j = C_j$. An increase in the production unit's initial resources perfectly correlates with an increase in the unit's production budget. Now I focus on the optimal allocation of inputs to the same production unit if it were controlled by a studio. There are two major changes. First, the production unit is vertically integrated. Second, the production unit gains access to larger internal capital markets. Denote $\mathcal{I}_g$ the set including the production units controlled by the studio before the merger and the acquired production unit j. In each period after the merger the studio maximizes profit solving

$$\underset{\{L_i\}_{i\in\mathcal{I}_g}}{max} \quad \sum_{i\in\mathcal{I}_g}\Big\{\tau(pf_i(L_i)) - L_i\Big\}$$

$$s.t.$$

$$\sum_{i\in\mathcal{I}_g} L_i \leq \sum_{i\in\mathcal{I}_g} C_i.$$

Denote λ the Lagrange multiplier associated with the budget constraint. The firm maximization problem is equivalent to

$$\underset{\{L_i\}_{i\in\mathcal{I}_g}}{max} \quad \sum_{i\in\mathcal{I}_g}\Big\{\tau(pf_i(L_i)) - L_i\Big\} + \lambda[\sum_{i\in\mathcal{I}_g} C_i - \sum_{i\in\mathcal{I}_g} L_i]. \tag{3.8}$$

The Kuhn-Tucker conditions are

$$\tau(v)pf_i'(L_i) = (1+\lambda) \quad \forall i \tag{3.9}$$

$$\sum_{i\in\mathcal{I}_g} L_i \leq \sum_{i\in\mathcal{I}_g} C_i \tag{3.10}$$

$$\lambda[\sum_{i\in\mathcal{I}_g} C_i - \sum_{i\in\mathcal{I}_g} L_i] = 0; \quad \lambda \geq 0. \tag{3.11}$$

Equations 3.5 and 3.9 show that the optimal amount of resources invested in unit U_j increases with the rate of revenue internalization, τ, and diminishes with the severity of the financial constraints, λ. Before leveraging the first-order conditions to directly estimate τ, I derive the following result: When the firm's resource constraint binds, investment L_j in production unit j is less sensitive to the unit's cash flows, C_j, if the firm controls multiple upstream production units than if it controls a single upstream production unit. The proof is in Appendix B.4. If the stand-alone company controlling j is financially constrained and gets acquired by a studio, the model predicts that the investments in the production unit j become less sensitive to the initial unit cash flow C_j. Intuitively, a studio allocates its investment across multiple upstream assets to smooth idiosyncratic productivity shocks. In contrast, a financially constrained stand-alone production company is fully dependent on its productivity sequence. Instead, if companies had perfect access to external funding, ownership would not impact credit constraints. Regardless of its size, a firm would borrow the amount necessary to satisfy the unrestricted optimality condition ($\lambda = 0$). I use this result to directly test for the effect of vertical mergers on credit constraints.

3.5 Mechanisms

The final contribution of this article is to quantify the different mechanisms behind the change in investments. To this end, I proceed in three steps. First, I perform a reduced-form test to assess the relevance of credit constraints. Second, I estimate the production function of movie tickets to pin down α, i.e., the conversion rate of investment into output. Third, I isolate and estimate the change in the ratio $\frac{\tau_j(i)}{\tau_j(s)}$. The last two steps serve to distinguish the change in marginal returns from investments from the industry specific-technology. This is important in light of the multiple combinations of α and $\frac{\tau_j(i)}{\tau_j(s)}$ that can originate the reduced-form evidence. Different combinations of these objects can originate the same investment effects in the data.

In principle, α and the ratio $\frac{\tau_j(i)}{\tau_j(s)}$ are not separately identified from the optimality conditions. With only data on budget, it would not be possible to separately identify the technological parameter alpha from the difference in marginal returns between integrated and non-integrated production companies. However, I also observe output, which is proxied by box office revenue. Thus I overcome the identification problem in two steps. First, I estimate the production function. Then I leverage the investment first-order conditions to pin down the difference in internalized marginal returns.

Identifying both components is crucial for policy implications. Intuitively, suppose the investments' effect arises because of a very large alpha and a negligible change in internalized returns. The former mechanism is industry-specific. Therefore integration would have sizeable effects in the motion picture industry, but it would be unlikely to affect investments in other industries or production processes. By contrast, the change in marginal returns from investment is a function of the economic primitives of the markets. Hold-up problems lie in contractual incompleteness, which plagues different contracting environments. Therefore, if the predominant force explaining the estimated investment outcomes is a change in the internalized return on investment, the results of this article will extend to vertical mergers outside the entertainment industry.

3.5.1 Credit Constraints

Financial frictions posit the basis for vertical integration to affect investments through borrowing constraints. With imperfect access to external capital markets, entering a diversified studio could improve access to funds. At the same time, vertical integration could increase the debt of the firm curtailing investments. Then, reducing borrowing frictions affects investment decisions. When the borrowing constraint binds, Lemma 3.4 proves that the production budget correlates less with cash flows after a vertical merger, providing a testable implication for the relevance of borrowing constraints.

Therefore, I aggregate the revenues from domestic box office at year- and production-unit level. I use box office revenues as a proxy for available cash flows. The reason is twofold. First, it maximizes the power of the test as I have domestic box office revenue for every observation. Second, both the international box office and the additional revenue stream are directly proportional to domestic box office. Then I specify the production budget in year t as a function of cash flows, proxied by the domestic box office in year $t-l$, ownership status, and the interaction between cash flows and ownership:

$$Y_{mjt} = \alpha BO_{j,t-l} + \gamma AfterMerger_{mjt} + \beta AfterMerger_{mjt} * BO_{j,t-l} + \beta_z \mathbf{Z}_{mjt} + \epsilon_{mjt}. \quad (3.12)$$

The null hypothesis is that the coefficient on the interaction variable is not different from zero, implying that either credit constraints do not bind or they are orthogonal to the ownership status. This suffices to rule out any confounding effects related to vertical mergers affecting investments through changes in firms' borrowing constraints. I run specification 3.12 separately for the in-house sample and the outward sample.

To robustify the results, I account for the possibility that the relevant cash flows for investments are from domestic box offices collected one, two or three years before the production year of the movie ($t - l \in \{1, 2, 3\}$). Cash flows are correlated to investments, supporting the presence of borrowing constraints in the industry. The coefficient on the box office variable in Table 3.4 is statistically significant and economically relevant in half of the specifications. Notwithstanding, vertical integration does not impact access to funds. In all the

TABLE 3.4. Test the Impact of Vertical Mergers on Borrowing Constraints

Dependent Variable:	Production Budget ($ million)					
	(1)	(2)	(3)	(4)	(5)	(6)
Specification	t-l=1	t-l=2	t-l=3	t-l=1	t-l=2	t-l=3
Box-Office	0.235***	0.104	0.173***	0.142***	0.100	0.065
	(0.045)	(0.064)	(0.047)	(0.047)	(0.075)	(0.221)
AfterMerger*Box-Office	-0.056	0.065	-0.055	-0.082	-0.043	0.051
	(0.060)	(0.078)	(0.073)	(0.063)	(0.082)	(0.226)
Observations	96	93	84	80	71	59
Adjusted R-squared	0.379	0.224	0.199	0.298	0.208	0.310

Notes: Standard errors in parentheses; *** p<0.01; ** p<0.05;* p<0.1. Standard errors are robust to heteroskedasticity at company level. Production budget, the dependent variable, and cash flows are measured in million dollars. Results in column (1)-(3) are based on the in-house sample, and results in columns (4)-(6) are based on the outward sample.

specifications I do not reject the null hypothesis that the correlations between cash flows and investments change after integration. Therefore, changes in ownership status do not reduce financial frictions for stand-alone companies. Based on these results, in the rest of the article I assume that credit constraints are independent from vertical mergers.

3.5.2 Technology

I estimate the technology transforming investments, measured by production budget, into output, measured by movie tickets. The parameter governing this relation is industry specific. Conversely, the changes in investments' marginal returns due to ownership apply to a multiplicity of markets and industry. Quantifying the technological rate of transformation allows separating the portion of effects explained by industry-specific mechanisms from those explained by hold-up problems and the anticompetitive rationale. The presence of these latter mechanisms informs antitrust decisions outside the entertainment industry.

The industry produces a homogeneous good: movie tickets. Two types of investment are required to bring a movie into theaters: production budget and advertising expenditures. Those investments take place sequentially, and are incurred by divisions operating in different levels of the industry. Upstream production divisions invest in the production budget, whereas downstream distribution divisions invest in advertisement. These investments show high complementarities, but are also partially substitute. A very famous actor, with a very high salary, potentially offsets an ineffective marketing campaign. Accordingly, I model tickets' production according to a Cobb-Douglas production function:

$$Y_{mji,t+1} = e^{\omega_{mjt}}(L_{mjt})^{\tilde{\alpha}}(A_{mi,t+1})^{\tilde{\beta}}.$$

Here ω_{mjt} represents the productivity shocks that are potentially observable or predictable by the production companies when they make input decisions; L_{mjt} indicates the production budget invested by the upstream division j in year t; and $A_{mi,t+1}$ represents the advertising expenditures incurred by the distributor i.

Intuitively, the production budget can be thought of as a summary statistic of the physical input quality. The marketing investment expands the market creating awareness of the project among potential consumers. The timing reflects the fact that an advertising campaign generally starts in the post-production phase of the movie, whereas budget expenditures are set at the end of the development stage. Therefore, the distribution divisions observe the production budget before making advertising expenditures. It is industry practice to calibrate the advertising expenditures on the project's production budget. Motivated by these facts, I assume that advertising is a function of production budget:

$$A_{mi,t+1} = h(L_{mjt}) = (L_{mjt})^{\gamma}$$
$$Y_{mj,t+1} = e^{\omega_{mjt}}(L_{mjt})^{\alpha},$$

where $\alpha = \tilde{\alpha}+\tilde{\beta}\gamma$. Therefore, log-advertisement expenditures in $t+1$ are directly proportional to log-production budget investment in t. This transformation subsumes that there are no underlying differences at distribution level. That is, the distribution units have the same access to resources and ability to implement marketing campaigns. Generally, this is a strong assumption. However, two characteristics of the industry mitigate this assumption's implications on my estimates. First, in the outward sample the distributor is held fixed before and after the merger by construction. In the in-house sample, except for the Marvel merger, the distributor remains constant before and after the vertical mergers. Second, the distribution side of the industry is an oligopoly where the five major companies have similar market shares. Thus, even for the subset of Marvel movies in the in-house sample the extent of the unobserved marketing confounders is limited.

My goal is to identify and estimate α. To this end, I leverage the information on movie-level box office revenues in the data. Studios are effectively price takers at the box office and charge a common price $\tilde{P}$. As argued by Einav and Orbach (2007), one reason lies within the contractual features implemented to discipline movie theaters. However, box office revenues are only a fraction of studios' revenues. Nonetheless, ancillary revenues are proportional to the box office success of a movie. Streaming video platforms, such as Netflix, acquire the rights to stream the movie paying an amount proportional to the movie's box office success. Similarly, the revenues from theme parks are increasing in the theatrical success of the film. Given that ancillary revenues are directly proportional to theatrical sales, $P \propto \tilde{P}$, project-level revenues can be written as

$$R_{mj,t+1}(L_{mjt}) = Pe^{\omega_{mjt}+\epsilon_{mjt}}(L_{mjt})^{\alpha}, \tag{3.13}$$

where ϵ_{mjt} is a log-normal idiosyncratic term. I estimate the trans-log of equation 3.13

$$r_{mj,t+1} = p + \omega_{mjt} + \alpha l_{mjt} + \epsilon_{mjt}, \tag{3.14}$$

where p is the constant stemming from the OLS specification and represents the average return from investment.

The technology parameter of interest, α, is recovered by variation in the observed production budget. The main endogeneity threat in the production estimation comes from the production-division unobserved productivity term. Thanks to the granularity of the data I sidestep the restrictive assumption of time-invariant productivity, without having to recover the unobservable productivity by inversion such as in Olley and Pakes (1996). Project-level covariates, such as the sequel indicator, capture part of the time-varying productivity shock, whereas I assume that the remaining productivity is time-invariant, and at the division level. This reflects the talent present in the production division. Indeed, many of these companies were built around directors or producers who stayed and identified with the company along its production activity. Formally, the key identifying assumption is that the unobserved productivity term is separable additive in the production company fixed effects and project-specific covariates:

$$\omega_{mjt} = \omega_j + \rho_z \mathbf{Z}_{mjt}. \tag{3.15}$$

Finally, identification might be still threatened by a concern of functional dependence similar to the one pointed out by Ackerberg et al. (2015). Indeed, the first-order conditions in specification 3.9 show that the optimal production budget depends on credit constraints, the technological parameters, the rate of profit internalization and ω. Hence, if credit constraints are constant over time, there is no variation left to identify α, except the change in profit internalization rates.

One way to neutralize the functional dependence problem is to assume an optimization error. That is, it exists an optimal level of the production , L_{mjt}, but the firm chooses the optimal level plus noise. The sensibility of this assumption is industry-specific. Notwithstanding, the motion picture industry appears well-suited for this data-generating process. The uncertainty behind a movie's success is extremely elevated, especially for new projects. Moreover, the production budget is chosen before the production stage and numerous are the instances when projects go far beyond the ex ante planned budget because of a director's artistic requests. Under such optimization error, estimating equation 3.14 provides consistent estimates for the technological parameter of interest. And, again, this data-generating process is the same as the one required for Olley and Pakes (1996).

To further reduce the concerns about unobserved heterogeneity at the distributor level I use out-of-sample data that is, projects produced by the main production unit of each studio and distributed through their distribution division. Thus, the firm-level heterogeneity is constant along the supply chain, justifying the use of production-unit fixed effects. The technological rate of transformation of the variable input into output lies in a narrow interval when controlling for movie-level controls. The estimates in Table 3.5 range from 0.645 to 0.580, lending credibility to movie-level controls for capturing a large part of the unobserved endogeneity. In what follows, I use the estimate of the within model to back out the implied change in revenue internalization rates. Intuitively, a 10% increase in the production budget is associated with a 6.07% increase in the ticket unit price. Table 3.5 also reports the results from a standard OLS specification and a model with time-fixed effects.

TABLE 3.5. Out-of-Sample Production Function Estimates

Dependent Variable:	Log Revenues			
	(1)	(2)	(3)	(4)
Specification	OLS	OLS	Within	Total
Log-production budget	0.740***	0.645***	0.607***	0.580***
	(0.047)	(0.050)	(0.052)	(0.054)
Time-varying controls	✗	✓	✓	✓
Company FE	✗	✗	✓	✓
Year FE	✗	✗	✗	✓
Observations	655	654	654	654
Adjusted R-squared	0.385	0.408	0.409	0.420

Notes: Results are computed using projects from the studios' original upstream division, distributed by their own distribution branch. I focus on productions where the studios' division is the main holder of residual profit claims based on company credits. Standard errors in parentheses; *** p<0.01; ** p<0.05;* p<0.1. Standard errors are robust to heteroskedasticity.

3.5.3 Contractual Incompleteness and Hold-ups

I use the previous results to back out the implied change in marginal returns from investment. To estimate the change in revenue internalization due to vertical integration I use the investment optimality conditions 3.9. After quantifying the technological parameter, α is observed. Moreover, the orthogonality of between vertical integration to borrowing constraints implies that $\lambda_{j,t}(o) = \lambda_{j,t}$. Denoting the constant term $\frac{ln(P\alpha)}{1-\alpha} = \gamma$ and replacing the productivity as a function of the project's characteristics and company fixed effects I write

$$l_{mjt} = \gamma + \frac{1}{1-\alpha}[ln(\tau_{jt}(o)) - ln(1+\lambda_{j,t}(o))] + \frac{\omega_{mjt}}{1-\alpha} \tag{3.16}$$

$$l_{mjt} = \gamma + \frac{1}{1-\alpha}[ln(\tau_{jt}(o))] + \tilde{\omega}_j + \tilde{\rho}_z \mathbf{Z}_{mjt} + \tilde{\lambda}_{j,t}, \tag{3.17}$$

where the tilde denotes terms scaled by $\frac{1}{1-\alpha}$. At last, I assume that $\tau_{jt}(o) = \tau(o)$. That is, the revenue internalization depends only on the company ownership status. The motion picture industry's contracting environment justifies this assumption. Stand-alone production companies enter into long-term distribution contracts with defined revenue shares. The contract defines the output, and sets minimum clauses about advertising and distribution effort on the studio's account. Therefore, except for ownership changes, there is a narrow margin for changes in revenue internalization over the course of the distribution deal. Using an indicator for the ownership status, $\mathcal{X}_{mjt}(i)$, and the fact that credit constraints are absorbed by the measurement error term, equation 3.17 is equivalent to

$$l_{mjt} = \gamma + \mathcal{X}_{mjt}(i)\frac{1}{1-\alpha}[ln(\tau_{jt}(v)) - ln(\tau_{jt}(s))] + \frac{1}{1-\alpha}[ln(\tau_{jt}(s))] + \tilde{\omega}_j + \tilde{\rho}_z \mathbf{Z}_{mjt} + \epsilon_{j,t}. \tag{3.18}$$

As a consequence, a standard difference-in-difference estimator captures $\hat{\beta} = \frac{1}{1-\alpha}[ln(\tau_(i)) - ln(\tau_(s))]$. As the production function estimation recovered α, I can invert this equation, taking the exponential on both sides, to derive

$$\frac{\tau_j(v)}{\tau_j(s)} = exp((1-\hat{\alpha}) * \hat{\beta}). \tag{3.19}$$

This quantity embodies the change in internalized marginal returns after vertical integration. A sizeable increase in this ratio is consistent with the predictions of the property right theory, corroborating the importance of firms' boundary for investments decisions. In presence of hold-up concerns, bringing production in-house increases investment for the integrating firm. This is because both parties do engage in rent-seeking behavior and ideally solve a joint maximization problem when making investment decisions. In contrast, a post-merger decrease in the ratio can be driven by anticompetitive incentives to reduce input quality. Another rationale is the increased coordination issues between multiple firms. However, the conceptual boundary between these two mechanisms is thin. Crucially, the investment outcome is the same together with a reduction in welfare in the short run.

The inversion step highlights the importance of the production function estimation step. Indeed, the ratio is not identified by a simple difference-in-difference regression. Multiple combinations of $ln(\tau(v)) - ln(\tau(s))$ and $\hat{\alpha}$ could lead to the same $\hat{\beta}$. Moreover, α is an industry-specific parameter. Conversely, the change in investments' marginal returns generalizes to a multiplicity of markets and industries where long-term distribution contracts are used. Prominent examples are airplane manufacturers, electric utilities and cable TV. As mentioned above, the main results subsume $\hat{\alpha} = 0.607$ as technological parameter. This is the value estimated by the within model in Table 3.5.

Integration solves hold-up issues for the integrating firm. In contrast, it raises anticompetitive concerns for its rivals. Indeed, Table 3.6 exhibits a sizeable change in marginal returns from investments in both samples, but of opposite sign. The integrating firm internalizes 37% higher revenues after vertical integration compared to the pre-merger scenario. By con-

TABLE 3.6. Implied Change in Internalized Marginal Return on Investment

	In-house			Outward		
	(1)	(2)	(3)	(4)	(5)	(6)
Estimates	Full	Full	Full	Full	Full	Full
$\widehat{\frac{\tau_j(v)}{\tau_j(s)}}$	1.400	1.370	1.355	0.889	0.859	0.851
	(.164)	(.185)	(.179)	(.041)	(.042)	(.044)
Specification						
Time-varying controls	✗	✓	✓	✗	✓	✓
Distribution pattern	✗	✗	✓	✗	✗	✓
Company FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓

Notes: Here $\frac{\tau_j(v)}{\tau_j(s)}$ represents the ratio of revenue internalization parameters; $\hat{\beta}$ is the coefficient on the ownership dummy stemming from the difference-in-difference estimation with log production budget as the dependent variable; and α is the technology parameter that governs the transformation of inputs (production budget, as summary statistic for input quality) into output (domestic box office tickets). The columns labeled "Full" use the entire panel. The distribution pattern includes a dummy for movies who got a wide release, i.e., more than 600 theaters in the opening weekend, compared to those that obtained a limited release. The standard errors in parentheses are block-bootstrapped at the production company level with 200 replications. To compute the bootstrapped standard errors I used the Within α estimated from in-sample observations reported in Table 4.7 in order to maintain the same population during the bootstrap procedure.

trast, the upstream production company internalizes 14.1% lower returns when it anticipates that its rivals distribute the movie. Thus, firms' boundaries appear to be a relevant determinant for investments. In turn, this leads to higher investments favoring the downstream counterpart, while reducing investments towards its downstream rivals. The existence of a sizeable change in marginal returns from investment confirms that this effect is general to vertical integration. Therefore, they are likely to arise and to have a sizeable magnitude for mergers taking place in industries plagued by contractual incompleteness.

3.6 Conclusions

In this article I combine reduced-form techniques with a model of within-firm resource allocation to study the impact of vertical integration on investments. I apply these techniques to the US motion picture industry, analyzing investments in the upstream production. A key ingredient of my analysis is the construction of a unique data set, assembled from multiple sources, that includes detailed project-level cost and revenue data combined with information on firms' ownership structure over the last 25 years. This allows me to identify eight vertical mergers and evaluate their impact on the investment patterns in the production budget.

Vertical integration has a large and significant impact on investments. In the US motion picture industry, the investments in projects distributed by the integrating distributor increase on average by $75 million compared to pre-merger levels. In contrast, the investments in projects distributed by the integrating distributor's downstream rivals decrease on average by $21 million compared to pre-merger levels. To isolate the economic mechanisms behind these effects, I combine the reduced-form estimates with a resource allocation model in the spirit of Giroud and Mueller (2019), modified to account for the role of vertical integration. The advantages of the model are threefold. First, it provides a testable empirical implication to test for the role of financial constraints on investments. Second, it specifies a technology function that combines investments to produce the industry output, movie tickets. Estimating the production function measures the parameter converting input into output. Third, it allows me to distinguish the role of an industry-specific technology from the solution of contractual incompleteness due to integration. Firm boundaries matter. When investing in projects distributed by the downstream counterpart, a production company internalizes 37% higher revenues after integration compared to the pre-merger scenario. In contrast, when the project is distributed by a downstream rival, the upstream production company internalizes 14.1% lower returns. In turn, this leads to higher investments favoring the downstream counterpart, while reducing investments towards its downstream rivals. This might have relevant aggregate welfare consequences.

From an antitrust standpoint, these results call for an in-depth assessment of the investment outcomes of vertical integration.

CHAPTER 4

Appendix

A Chapter I and II

A.1 Data

In this subsection I provide a detailed description of the key variables I measured and inferred from the raw data. Then, I provide a brief description of the data sources other than the infrastructure and drilling ones.

Variable Definitions

development well. I define development wells those entities that are drilled within two miles of a previously drilled well with proven positive production. Furthermore, the data identifies those prospect drilled in "Wildcat" fields. Those wells are exploratory wells that turned out dry holes. I refine the precision of my inference of development wells excluding these prospects from the data.

drilling prospect. I use wells that have been spudded between 2008 and 2022 to identify drilling prospect. That is, a drilling prospect is defined as a point in space where a well has been spudded between 2008 and 2021. Then, I assume that each drilling prospect enters the producer's choice set from its "discovery date".

discovery date. I trace back the discovery date of each well leveraging information on the past drilling activity of the operator holding the prospect. I assume that the discovery date of a prospect coincides with the completion date of the first producing well held by the same operator within a two miles radius. This is justified by my focus on development wells.

Additional Data Sources

Infrastructure I complemented this data with public available data on the location of crude transmission pipelines, crude oil refineries, crude oil storage terminals, shale basins available from the U.S. Energy Independent Administration (EIA). Figure 4.4 provides a snapshot of the current infrastructure network.

Production Data. I measure firms monthly sales using a data-set of monthly crude oil production that I obtained from Enverus, a private data provider. These data spans from 2000 to 2020 and each observation corresponds to an entity, a lease or a well, controlled by a firm in charge of making production decisions. Then I use the company names and time of production to identify entry, exit and incumbents in a given location.

Inventories and barrel disposition methods I measure infrastructure usage and inventories at field-level with data on monthly methods of disposition of barrels from the field. The data are publicly available in pdfs format from the railroad commission of Texas from May 2013 to December 2020. I manually digitized those data. Barrels are shipped from the wellhead to short-term storage facilities to eventually reach the downstream refineries. The method of dispositions indicates how barrels are removed from the well-head, if by pipeline, trucks or rail.

The data are at the sub-field level, and crucially the report the flow of barrels leaving the field. They include the initial number of barrels present on the field at the beginning on the month, the number of barrels produced, the number of barrels taken away, and the number of barrels remaining at the end month. I match this data with production data based on field-names, month and year. This allows me to identify inventories and initial transport choices at a granular spatial level.

At the onset of the shale revolution, the pipeline network fell short in absorbing the crude oil production. Trucks dominate over pipelines as predominant crude disposition method. Indeed, the number of barrels disposed by pipeline exceeded the one of oil disposed by trucks starting mid 2015. This is consistent with the lagged expansion of the pipeline network. The gap constantly widened from that point onward, despite the surge in supply. Figure 4.3 clearly show the substitution between pipelines and trucks.

Crude Oil Prices At last, I downloaded the price bulletin posted in the marketing bulletin of the marketing division of *Plains All American Pipelines*. This include upstream purchasing prices for specific region and crude type within Texas and other producing states. This is the (discounted) price shipper pay at the well-head. Differences in these prices is used to isolate aggregate transport costs at region level.

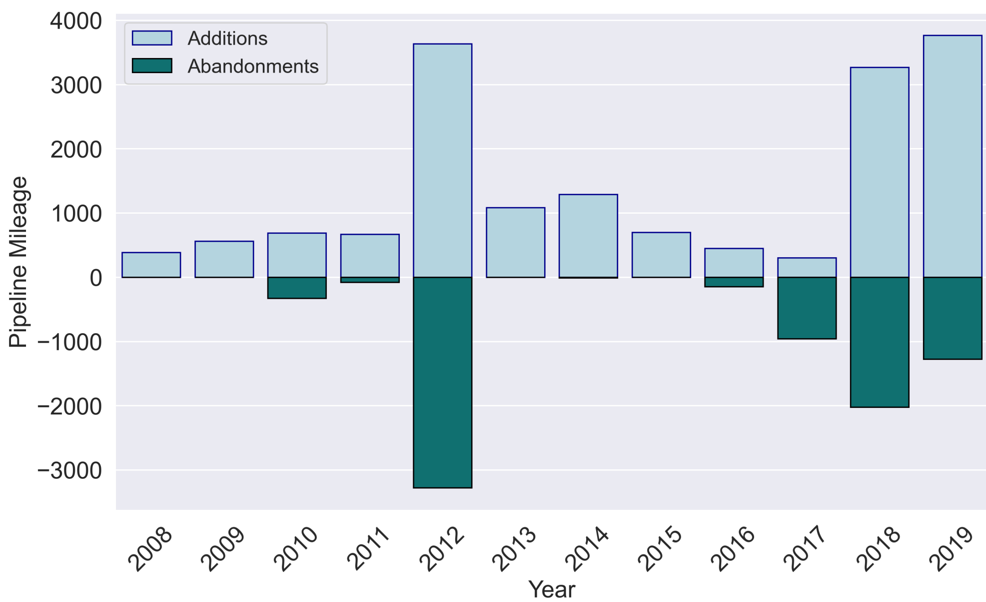
A.2 Descriptives: Dynamics

In this subsection I provide additional details about the industry's evolution in the recent decade. First, Figure 4.1 and 4.2 show how the pipeline network changed over time and across space. Following the increase in the drilling activity, and thus oil supply builders heavily invested in new pipeline. The rate of new constructions partly decelerated after the 2014 oil crash, to pick up again in 2018 and 2018.

At the same time, a large portion of the old pipeline network has been abandoned. This is in line with the change of the geography of supply basins. The discovery of novel shale basins changed to spatial distribution of oil in Texas. At the same time, some vertical wells dried up. These trends explain why part of the infrastructure has been abandoned.

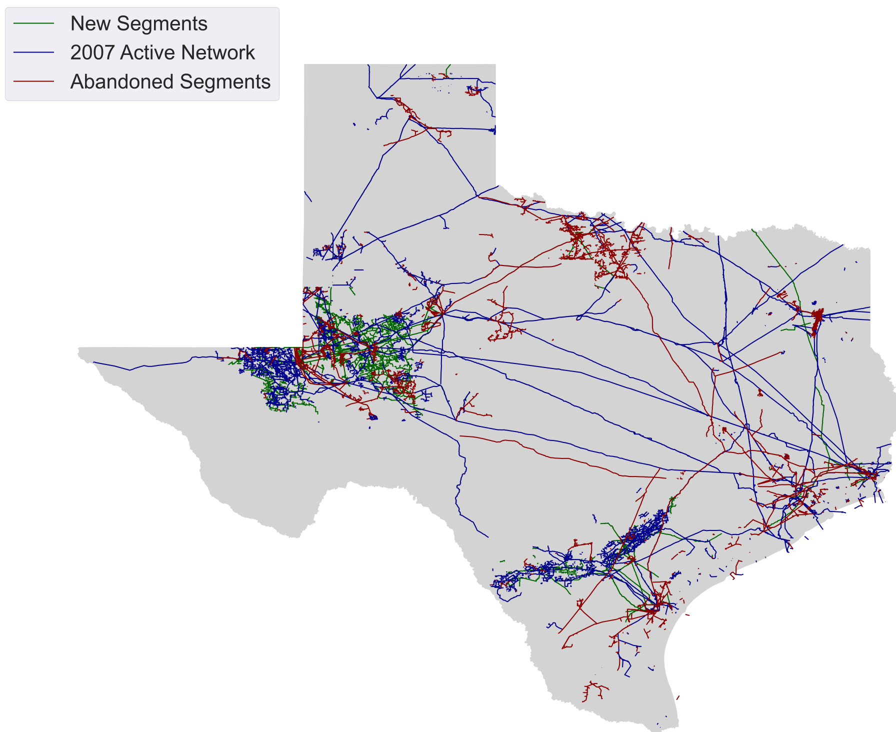
Despite the construction of new infrastructure, producers heavily rely on trucks to move the crude away from the leases. However, the share of crude disposed by pipelines share went from 35% to 55% of the monthly supply. Therefore producers substitute away from trucks when pipelines were available. This is consistent with pipelines' tariffs being lower than trucks'.

Figure 4.1: Timeseries of Pipeline Additions (2007-2020)



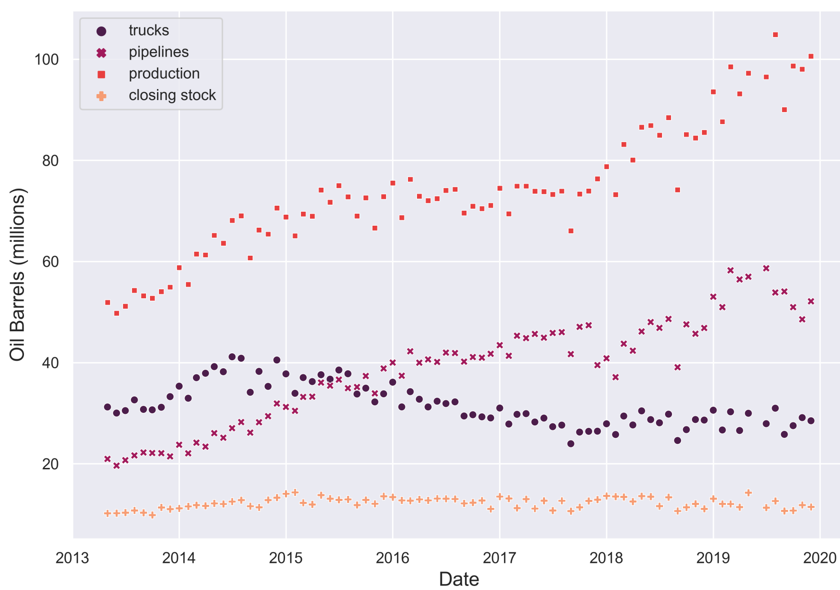
Notes: Additions and abandonments are measured in miles. The data includes pipeline segments longer than one mile.

Figure 4.2: Changes in the Texas Crude Oil Infrastructure Network



Notes: The data includes pipeline segments longer than one mile.

Figure 4.3: Production, Transportation Methods, Inventories



Notes: The data points show the aggregate monthly quantity of crude disposed from the Texas' oil fields. Production are the barrels of crude extracted each month. The closing stocks indicate the amount of crude do not moved away from the fields.

A.3 Estimation

In this section I show the formal derivations of the closed form solution for the producer's probability of drilling. This choice probabilities are at the basis of the Nested Fixed Algorithm. An oil producer decides to complete a drilling prospect if and only if the profits from drilling today exceeds the option value from delaying drilling to the next period. Formally, an oil producers decides to drill and complete a prospect if and only if:

$$\pi(x_i, p, \tilde{w}_i) + \epsilon_t(1) > \beta E[V(x_i', p', \tilde{w}_i, M', \epsilon')] + \epsilon_t(0)$$
$$\epsilon_t(1) - \epsilon_t(0) > \beta E[V(x_i', p', \tilde{w}_i, M', \epsilon')] - \pi(x_i, p, \tilde{w}_i)$$

The unobserved error component $\epsilon_{i,t}(I)$ is assumed to be identically and independently distributed according to a logit distribution. Hence, the probability that a drilling prospect is completed given the realized prices, the well's access status and the exogenous profit shifters coincides with the well known multinomial logit formula

$$\begin{aligned}
&Pr(I_d = 1|p, x_i, \tilde{w}_i, M) = \\
&Pr(\epsilon_t(1) - \epsilon_t(0) > \beta E[V(x_i', p', \tilde{w}_i, M', \epsilon')] - \pi(x_i, p, \tilde{w}_i)) = \\
&= \frac{exp(\pi(x_i, p, \tilde{w}_i))}{exp(\pi(x_i, p, \tilde{w}_i)) + exp(\beta * E[V(x_i', p', \tilde{w}_i, M', \epsilon')])}
\end{aligned}$$

A.4 Identification

I consider a generic prospect i that obtained a pipeline connection in period t. In period $t-1$ the prospect does not enjoy a connection to the pipeline. I discuss identification up to a normalization for the discount factor β and the logit variance σ_e.

I consider the time period when a prospect does not have access to the pipeline network.

The ratio between the model's choice probabilities can be expressed as:

$$\frac{Pr(I_{t-1}=1|\tau_0)}{Pr(I_{t-1}=0|\tau_0)} = \frac{exp((p_{t-1}q - \tau_0 q - \kappa(w))/\sigma_e)}{exp((\beta * \tilde{EV}_{t-1})/\sigma_e)} \tag{4.1}$$

$$log\Big(\frac{Pr(I_{t-1}=1|\tau_0)}{Pr(I_{t-1}=0|\tau_0)}\Big) = \sigma_e^{-1}[p_{t-1}q - \tau_0 q - \kappa_i(w) - \beta * \tilde{EV}_{t-1}] \tag{4.2}$$

Through inversion I can express τ_0 as a function of this ratio:

$$\tau_0 = \frac{\beta * \tilde{EV}_{t-1}}{q} + p_{t-1} - \frac{\kappa_i(w)}{q} - \frac{\sigma_e}{q} * log\Big(\frac{Pr(I_{t-1}=1|\tau_0)}{Pr(I_{t-1}=0|\tau_0)}\Big) \tag{4.3}$$

Where for simplicity I suppressed the dependence on i, the state variable vector, and parameters other than the shipping costs, in the probability objects notation. On the right hand side of the equations there are two unknowns, the scale parameter of the logit term and the fixed costs κ_i.

Given the scattered arrival of the timeline, supposed there exist another prospect j that has a connection to the pipeline in $t-1$, then I can recover τ_1 as:

$$\tau_1 = \frac{\beta * \tilde{EV}_{t-1}}{q} + p_{t-1} - \frac{\kappa_j(w)}{q} - \frac{\sigma_e}{q} * log\Big(\frac{Pr(I_{t-1}=1|\tau_1)}{Pr(I_{t-1}=0|\tau_1)}\Big) \tag{4.4}$$

Note that, in both formulations EV_{t-1} is observed following the inner fixed point routine of the estimation algorithm. Suppose that both prospects have the same drilling technology, such that $\kappa_i(w) = \kappa_j(w) = \kappa_0$.

Therefore, to jointly identify τ_0, τ_1 and κ_0 I need a third equation. To this end I leverage the discontinuity in the probability of drilling for the same prospect triggered by the connection to the pipeline.

$$\begin{aligned}\tau_1 - \tau_0 = {} & \frac{\sigma_e}{q}\Big[log\Big(\frac{Pr(I_t=1|\tau_1)}{Pr(I_t=0|\tau_1)}\Big) - log\Big(\frac{Pr(I_{t-1}=1|\tau_0)}{Pr(I_{t-1}=0|\tau_0)}\Big)\Big] \\ & - \frac{\beta}{q}[(\tilde{EV}_t - \tilde{EV}_{t-1})] + (p_t - p_{t-1})\end{aligned} \tag{4.5}$$

This steps make clear how identification is achieved up to a normalization for the discount factor and the variance of the unobserved error component. Normalizing the discount factor and the logit variance I obtain three equations in three unknown. Therefore the parameters are jointly identified via maximum likelihood. Alternatively, to recover the logit variation from the model I need to impose a normalization over the fixed costs one of the prospects. This normalization is required since q is constant over time. This is the expected extracted crude from the drilling prospect, which does not vary over time.

The identification of the cost parameter associated with horizontal well comes from the spatial variation in drilling rates between prospects with different drill types, conditioning on the pipeline state. Having multiple year in the sample allows mw to identify the linear time trend, normalizing to zero the shock in the first year of the sample.

Algorithm details

Baseline Model. I solve the value function 2.6 on a grid of points in $(p, x_p, q, \tilde{w} = Horizontal)$ space using standard value function iteration. I extend the state space beyond realized price and production values. The state space I used extends one-tenth from the lowest realized price and well's production to one-tenth from the highest realized price and well's production. The crude extracted from each well is expressed in million of barrels. The grid I use has 14,400 points: 60 price states, 60 production states, 2 connection to the pipeline states and 2 well's type states. Starting from this density, the estimated results are insensitive to increase or decrease in the number of grid points. In the full estimation routine, the initial value function used for each guess of parameters is the value function from the previous guess. For the first parameter guess, the initial values is zero in all states. The convergence criterion is 10^{-6} on the sup norm of the value function.

With the value function solved, I can match the value from delaying completion to any give p, q, x_p and $\tilde{w}$. However, the price and production points in the grid do not coincide with the data realizations. Therefore, I use linear interpolation to find the value function at each revenue state. At each price-quantity grid point, (p, q) I calculate the value function at the realized $(\hat{p}, \hat{q})$ by linearly interpolating the value function between the states immediately above and below the grid's point.

Robustness checks. The value function iteration is similar to the baseline model, however the grid I use is more dense. In the augmented model I have 50 price states, 50 production states, 50 probability states, 2 connection to the pipeline states, 2 well's type states and 2 size states. Therefore, there are 750,000 grid points.

Starting from this density, the estimated results are insensitive to increase or decrease in the number of grid points. In the full estimation routine, the initial value function used for each guess of parameters is the value function from the previous guess. For the first parameter guess, the initial values is zero in all states. The convergence criterion is 10^{-6} on the sup norm of the value function.

In this formulation of the model, I expand the linear interpolation procedure to account for the predicted probability of having a connection to the pipeline. The predicted probability in the data, does not coincide with the grid points.

A.5 Graphs and Tables

Tables

TABLE 4.1. The Impact of Transport Infrastructure on the Timing of Investments

Dependent Variable:	Time for Completion				
Specification	(1)	(2)	(3)	(4)	(5)
Time for connection to pipeline	0.755***	0.564***	0.617***	0.639***	0.533***
	(0.053)	(0.045)	(0.017)	(0.036)	(0.030)
Well's cumulative oil		0.638	-0.087	-0.554*	-0.519
		(0.448)	(0.571)	(0.312)	(0.424)
Price 3 months before completion		-0.034***	-0.031***	-0.032***	-0.031***
		(0.008)	(0.005)	(0.004)	(0.005)
Discovery Year FE	✓	✓	✓	✓	✓
Oil Field FE	✗	✗	✓	✓	✓
Producer FE	✗	✗	✗	✓	✓
Pipeline Operator FE	✗	✗	✗	✗	✓
Observations	1,649	1,564	1,564	1,564	1,564
R-squared	0.885	0.912	0.935	0.908	0.911

Notes: Standard errors in parentheses; *** p<0.01; ** p<0.05;* p<0.1. Standard errors are clustered at the year of discovery level. The sample includes only single producing entities that have been completed and received a connection to the pipeline after 2008 and before 2020. Well cumulative production is measures in thousands crude oil barrel. Price at completion is measured using the daily West Texas Intermediate crude oil price. The time to complete the drilling prospect and the time to build the pipeline connections are measured in months. I impute 1 month as the time to build the pipeline connections for those drilling prospects who have always enjoyed the connection to a pipeline

TABLE 4.2. Estimates of the Baseline Model, with discount factor $\beta = .935$

	$\sigma_e = 1$	$\sigma_e = 2.5$	$\sigma_e = 5$	$\sigma_e = 7.5$	$\sigma_e = 10$	$\sigma_e = 12.5$
Cost Parameters						
τ_0 : variable costs without pipe	85.817	78.097	72.567	70.067	68.736	68.278
τ_1 : variable costs with pipe	71.251	66.740	59.66	53.828	48.575	43.841
κ_0: average sunk cost	6.294	13.443	25.104	36.861	48.657	60.462
κ_1: horizontal well	0.455	2.058	5.907	10.070	14.264	18.402
κ_2: linear time trend	-0.652	-1.132	-1.889	-2.682	-3.491	-4.306
Laws of Transition Parameters						
σ_ζ: log prices' change standard deviation	0.011	0.011	0.011	0.011	0.011	0.011
μ_ζ: log prices' change mean	0.000	0.000	0.000	0.000	0.000	0.000
δ_0: logit constant	-3.604	-3.604	-3.604	-3.604	-3.604	-3.604
	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)
ρ: prospect's cumulative production	10.112	10.112	10.112	10.112	10.112	10.112
	(.060)	(.060)	(.060)	(.060)	(.060)	(.060)

Notes:Standard errors in parenthesis. The discount factor across all specification is $\beta = .935$. Revenues are expressed in million of dollars. Fixed costs parameters are expressed in million of dollars. Transportation costs parameters are expressed in dollars per-barrel. The mean of the unobserved term logit distribution is calibrated at zero.

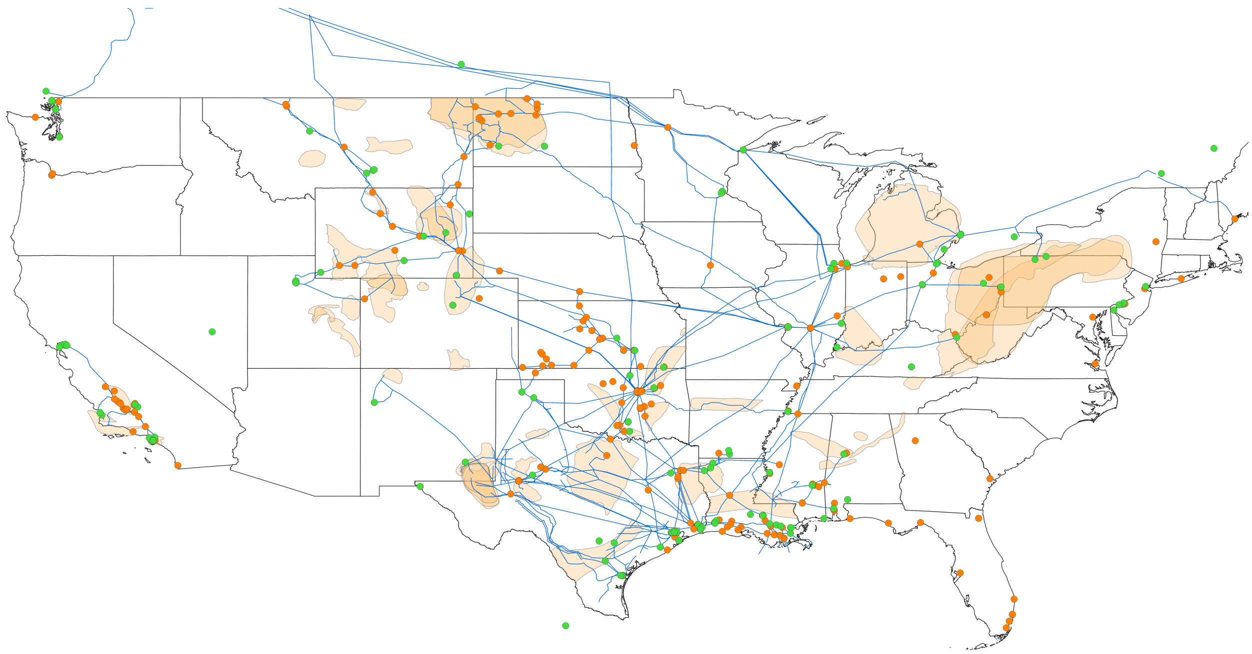
TABLE 4.3. Estimates of the Complete Model

	Size	Size Oil Supply Infrastructure Level
Cost Parameters		
τ_0: variable costs without pipe	78.636	75.85
τ_1: variable costs with pipe	67.879	65.250
τ_2: small producer (variable)	-11.215	-5.538
κ_0: average sunk cost	13.574	13.334
κ_1: horizontal well	2.438	2.581
κ_2: linear time trend	-1.205	-1.195
κ_3: small producer (sunk)	1.769	1.332
Laws of Transition Parameters		
σ_ζ: log prices' change standard deviation	0.011	0.011
μ_ζ: log prices' change mean	0.000	0.000
δ_0 logit constant	-5.840	-3.604
	(.005)	(.011)
δ_1: oil supply		0.038
		(.000)
δ_2: infrastructure level		0.761
		(.003)
ρ: prospect's cumulative production	10.112	5.832
	(.060)	(.075)

Notes: Standard errors in parenthesis. The discount factor across all specification is $\beta = .935$. Revenues are expressed in million of dollars. Fixed costs parameters are expressed in million of dollars. Transportation costs parameters are expressed in dollars per-barrel. The mean of the unobserved term logit distribution is calibrated at 0. The variance of the unobserved parameters is calibrated at 2.5

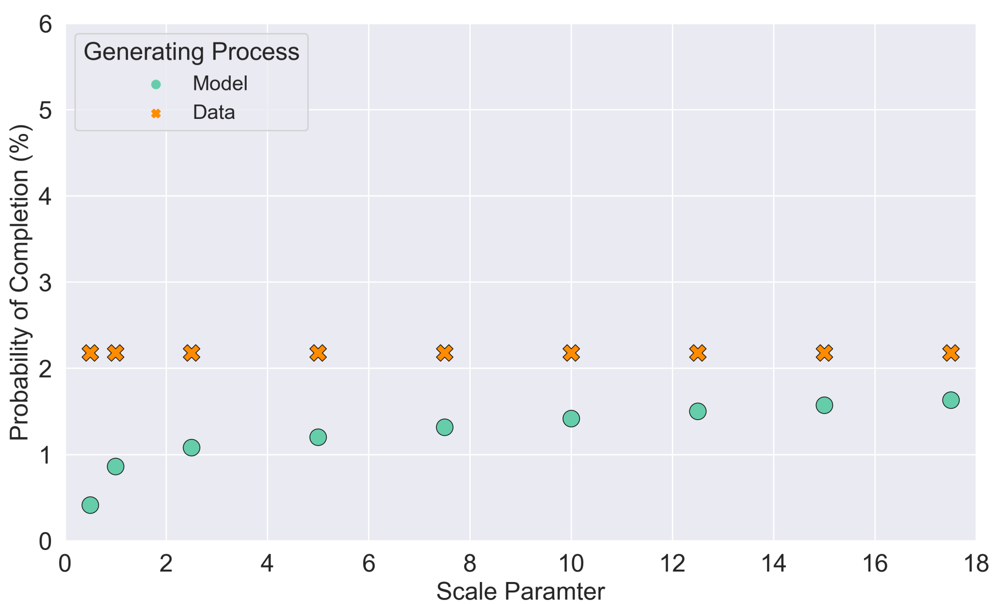
Graphs

Figure 4.4: US Crude Oil Infrastructure Network



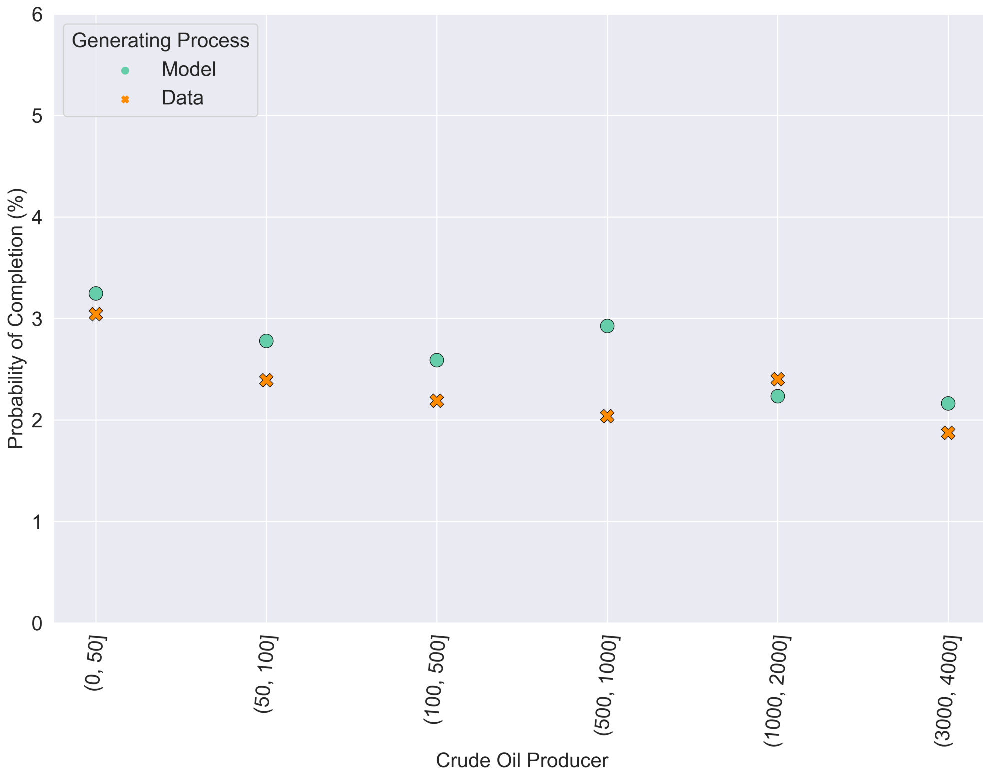
Notes: The blue lines represent crude pipelines. The green dots indicate crude refineries. The orange dots represent storage tanks. The light orange polygons indicate the crude oil basins that have been discovered on the U.S. territory.

Figure 4.5: Baseline Goodness of Fit without Time Trend



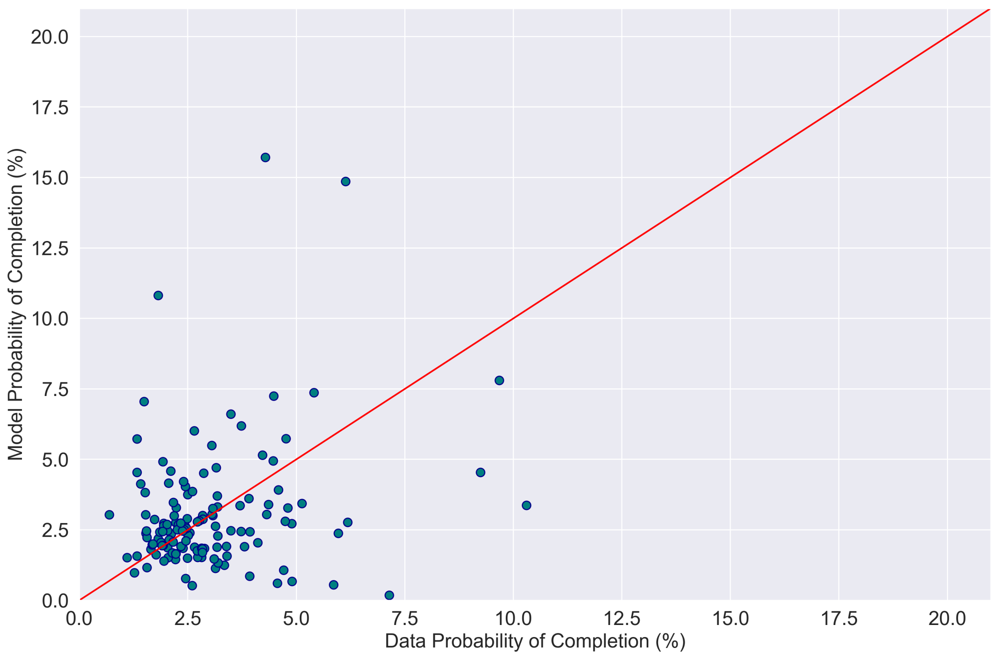
Notes: The discount factor used across all the specifications is $\beta = .935$

Figure 4.6: Augmented Model, Size-level heterogeneity



Notes: The model's probabilities come from the augmented model. The discount factor used across all the specifications is $\beta = .935$. The sample includes all the producers holding a drilling prospect during the years 2008-2021. The producers' size is computed using the total number of development prospects held by the firm in Texas.

Figure 4.7: Producer's Dispersion



Notes: The model's probabilities come from the augmented model. The discount factor used across all the specifications is $\beta = .935$. The sample includes all the producers holding a drilling prospect during the years 2008-2021.

B Chapter III

B.1 Robustness Checks

Empirical The previous estimates use all the available years after the merger. To mitigate concerns about long-run confounding factors, such as changes in market structure, I narrow the post-merger time window. I use for estimation movies produced within five years after vertical integration. The resulting estimates more accurately reflect short-term effects of vertical mergers. The impact of vertical mergers on investments remains large and economically relevant. Vertical mergers increase investments on average by the integrating firm by \$56.2 million, while reducing them by \$29.8 million for rivals, as reported in Table 4.4. At last, I replicate the estimation of the internalized returns on investment using the trimmed samples. The results are reported in Table 4.5. In the short-term, the estimates of the change in internalized revenues become 23.7% and 17.7% respectively. Standard errors ensure that

TABLE 4.4. Short-term Causal Effects of Vertical Mergers on Production Budget ($ million)

Dependent Variable:	Production Budget ($ million)					
Specification	(1) In-house	(2) In-house	(3) In-house	(4) Outward	(5) Outward	(6) Outward
AfterMerger (real)	62.599***	58.739***	56.188***	-27.438**	-29.668***	-29.803***
	(20.099)	(23.729)	(18.850)	(8.617)	(8.594)	(8.508)
AfterMerger (All)	✗	✓	✓	✗	✓	✓
Time-varying controls	✗	✗	✓	✗	✗	✓
Company FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Observations	287	287	287	215	215	215
Adjusted R-squared	0.502	0.503	0.573	0.309	0.309	0.418

Notes: Results are computed using movies produced within five years after vertical integration. Standard errors in parentheses; *** p<0.01; ** p<0.05;* p<0.1. Standard errors are block-bootstrapped at the production company level with 200 replications. Each specification uses movies' production budget as dependent variable expressed in $ million. The *AfterMerger* dummy includes 10 leads after the year of merger.

these estimates are substantially above one for the integrating counterpart and below one for rivals.

TABLE 4.5. Short-term Causal Impact of Vertical Integration on Internalized Marginal Return of Investments

	In-house			Outward		
	(1)	(2)	(3)	(4)	(5)	(6)
Estimates	5-years	5-years	5-years	5-years	5-years	5-years
$\widehat{\frac{\tau_j(v)}{\tau_j(s)}}$	1.270	1.240	1.229	0.828	0.819	0.801
	(.135)	(.127)	(.117)	(.048)	(.049)	(.056)
Specification						
Time-varying controls	✗	✓	✓	✗	✓	✓
Distribution pattern	✗	✗	✓	✗	✗	✓
Company FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓

Notes: $\frac{\tau_j(i)}{\tau_j(s)}$ represents the ratio of revenue internalization parameters. $\hat{\beta}$ is the coefficient on the ownership dummy stemming from the difference-in-difference estimation with log production budget as dependent variable. α is the technology parameter that governs the transformation of inputs (production budget, as summary statistic for input quality) into output (domestic box office tickets). Columns labeled "5-years" use movies produced five years after the merger. Distribution pattern includes a dummy for movies who got a wide release, more than 600 theaters in the opening weekend, compared to those that obtained a limited release.Standard errors in parentheses are block-bootstrapped at the production company level with 200 replications. To compute the bootstrapped standard errors I used Within α estimated from in-sample observations reported in Table 4.7 in order to maintain the same population during the bootstrap procedure.

B.2 Quality-Enhancing Investments in the Motion Picture Industry

In this article I use movies' production budget to measure quality-enhancing investments. Hence, to gauge the validity of production budget as a measure for quality enhancing investments, I list the expenditures it encompasses. The production budget is typically divided in four sections. *above the line* (creative talent), *below the line* (direct production costs) , *post-production* (editing, visual effects, etc.), and other (insurance, completion bond, etc.). Best-seller writers, and Oscar winning actors command higher salaries. By the same token, computer-generated imagery and elaborated visual effects demands significant investments[1]. These investments contribute to realize visually stunning shows. Therefore, the bigger the budget, the higher the movie's potential of capturing demand. Prima facie evidence supports the correlation between content expenditures and demand. I calculate the studios' market shares based on domestic box office in 2017. I aggregate data by studios' global ultimate owner, to capture distribution branches falling under the same company. Table ?? shows the correlation between total content expenditures and market shares. More importantly, the studios with the highest average production budget have the highest market shares. Interestingly, The Walt Disney Company captured 22% of movie-goers releasing only 7 movies, the same number as STX Entertainment, but with average budget of \$ 205 million, five times above STX Entertainment. Warner Media has the second highest market share, 18%. It deployed on average \$ 89 million per project. National Amusements and Twenty-First Century Fox go against the tide. However, the former is facing a radical re-organization attempting to merge Viacom and CBS Corporation, both its subsidiaries. Similarly, 20th Century Fox has been recently taken over by the Walt Disney Company. Ineffective content expenditures might signal corporate distresses that prelude a merger or exit from the market.

[1] For instance, (*Avatar* visual effects cost more than \$100 millions)

TABLE 4.6. Economic Performances of Active Studios in 2017

Studio	Market Shares (%)	Movies	Total Budget	Average Budget
Walt Disney Company	22.13	7	1437	205.3
Warner Media	18.85	18	1606	89.22
Comcast	15.31	21	1100	52.36
Tweny-First Century Fox	14.01	18	1110	61.69
Sony Corporation	11.57	19	775	40.79
Lions Gate	6.537	14	494	35.29
National Amusement	4.557	10	815	81.50
STX Entertainment	2.120	7	328	46.91

The table shows the economic outcomes of studios that released a minimum of one movie with budget above $5 million in the US in 2017. Market shares are calculated using domestic (US) theatrical box offices. "Movies" represents the movies distributed in 2017 by the studio. Total and average budget represents the total and average production budget calculated in $ millions.

B.3 Algorithm to Select the Control Groups

In this section I provide additional details on the procedure adopted to select the control groups. For each of the eight stand-alone company undergoing a vertical merger the procedure provides a placebo company matching trend in the outcome variable of interest. I proceed as follow.

I define the years of activity of all the upstream divisions observed in the data. A company is considered active in a given year if it has produced one movie that year. As a first step, I matched the treated upstream company to production divisions that were active in the same years. I focus on the five years preceding the vertical merger. This reduces the number of potential placebos for each treatment company. However, I obtained a set of potential controls for each vertical merger. That is, I have a set of candidate placebo production companies for each stand-alone company that changed ownership in my sample. Therefore, I refine the selection criteria using the yearly average production budget before the merger. I select the production company that minimizes the difference in outcome pre-trends with the treated company.

To reduce the threat of spillovers, I discarded division of the same studio from the set of potential placebos. Otherwise, the integration could directly affect the control, violating the assumptions of the research design.

B.4 Proof of Lemma 3.4

Suppose production unit is j controlled by a stand-alone company. Condition 3.11 implies that if $\lambda > 0$ $L_j = C_j$. Then $L_j C_j = 1$. Suppose production unit is j controlled by the studio. The first order condition 3.9 for investments in production unit $\forall\ i \in \mathcal{I}$ of studio g can be explicitly written as

$$\tau p \alpha e^{\omega_i} L_i^{\alpha-1} - (1+\lambda) = 0$$

After some algebra manipulation the optimal input level L_i can be expressed as

$$L_i = e^{\frac{\omega_i}{1-\alpha}} \left(\frac{p\alpha\tau}{1+\lambda} \right)^{\frac{1}{1-\alpha}}$$

Consider investments in production unit j. Differentiating,

$$L_j \lambda = -\frac{e^{\frac{\omega_i}{1-\alpha}}}{(1-\alpha)(1+\lambda)} (\frac{p\alpha\tau}{1+\lambda})^{\frac{1}{1-\alpha}}$$

When the firm's resource constraint binds, $\lambda > 0$, $\sum_{i \in \mathcal{I}_g} C_i = \sum_{i \in \mathcal{I}_g} L_i$. Substituting L_i

$$\sum_{i \in \mathcal{I}_g} C_i - \sum_{i \in \mathcal{I}_g} \left\{ e^{\frac{\omega_i}{1-\alpha}} \left(\frac{p\alpha\tau}{1+\lambda} \right)^{\frac{1}{1-\alpha}} \right\} = 0$$

Define the LHS as $G(\lambda, C_j)$. Applying the implicit function theorem and after some algebra

$$\lambda C_j = -\frac{(1-\alpha)(1+\lambda)}{(\frac{p\alpha\tau}{1+\lambda})^{\frac{1}{1-\alpha}} \left(\sum_i e^{\frac{\omega_i}{1-\alpha}} \right)}$$

At last, combining the previous expressions

$$L_j C_j = L_j \lambda \lambda C_j = \frac{e^{\frac{\omega_i}{1-\alpha}}}{\left(\sum_i e^{\frac{\omega_i}{1-\alpha}}\right)} < 1$$

And the result follows.

B.5 Graphs and Tables

TABLE 4.7. In-Sample Production Function Estimates

	In-house			Outward		
	(1)	(2)	(3)	(4)	(5)	(6)
Specification	OLS	OLS	Within	OLS	OLS	Within
Log-production budget	0.843***	0.635***	0.578***	0.998***	0.815***	0.795***
	(0.046)	(0.057)	(0.062)	(0.065)	(0.083)	(0.092)
Time-varying controls	✗	✓	✓	✗	✓	✓
Company FE	✗	✗	✓	✗	✗	✓
Year FE	✗	✗	✗	✗	✗	✗
Observations	622	620	620	345	345	345
Adjusted R-squared	0.346	0.411	0.428	0.409	0.440	0.435

Notes: Standard errors in parentheses; *** p<0.01; ** p<0.05;* p<0.1. Standard errors are robust to heteroskedasticity.

Bibliography

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Agerton, M. (2020, April). Learning where to drill: Drilling decisions and geological quality in the haynesville shale. Working Paper 20-439, SSRN.

Agerton, M. and J. Upton (2019, 07). Decomposing crude price differentials: Domestic shipping constraints or the crude oil export ban? *The Energy Journal 40*.

Allen, T. and C. Arkolakis (2019, January). The welfare effects of transportation infrastructure improvements. Working Paper 25487, National Bureau of Economic Research.

Anderson, S. T., R. Kellogg, and S. W. Salant (2018). Hotelling under pressure. *Journal of Political Economy 126*(3), 984–1026.

Atalay, E., A. HortaÃ§su, and C. Syverson (2014). Vertical integration and input flows. *The American Economic Review 104*(4), 1120–1148.

Banerjee, A., E. Duflo, and N. Qian (2012). On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics 145*, 102442.

Bernheim, D. and M. D. Whinston (1985). Common marketing agency as a device for facilitating collusion. *The RAND Journal of Economics 16*(2), 269–281.

Bolton, P. and M. D. Whinston (1991). The "Foreclosure" Effects of Vertical Mergers. *Journal of Institutional and Theoretical Economics (JITE) 147*(1), 207–226.

Borenstein, S., J. B. Bushnell, and F. A. Wolak (2002, December). Measuring market inefficiencies in california's restructured wholesale electricity market. *American Economic Review 92*(5), 1376–1405.

Brancaccio, G., M. Kalouptsidi, and T. Papageorgiou (2020). Geography, transportation, and endogenous trade costs. *Econometrica 88*(2), 657–691.

Bushnell, J. B., E. T. Mansur, and C. Saravia (2008, March). Vertical arrangements, market structure, and competition: An analysis of restructured us electricity markets. *American Economic Review 98*(1), 237–66.

Caoui, E. H. (2019). Estimating the costs of standardization: Evidence from the movie industry. *Working Paper*.

Chatterjee, S. (2019, March). Market Power and Spatial Competition in Rural India. Cambridge Working Papers in Economics 1921, Faculty of Economics, University of Cambridge.

Chipty, T. (2001). Vertical Integration, Market Foreclosure, and Consumer Welfare in the Cable Television Industry. *The American Economic Review 91*(3), 428–453.

Ciliberto, F. (2006). Does Organizational Form Affect Investment Decisions. *Journal of Industrial Organization 54*(1), 63–93.

Collard-Wexler, A. (2013). Demand fluctuations in the ready-mix concrete industry. *Econometrica 81*(3), 1003–1037.

Corts, K. S. (2001). The Strategic Effects of Vertical Market Structure: Common Agency and Divisionalization in the US Motion Picture Industry. *Journal of Economics & Management Strategy 2*(10), 509–528.

Covert, T. and R. Kellog (2017, September). Crude by Rail, Option Value, and Pipeline Investment. *working paper*.

Crawford, G. S., R. S. Lee, M. D. Winston, and A. Yurukoglu (2018). The Welfare effects of Vertical Integration in Multichannel Television Markets. *Econometrica* (3), 891â954.

Donaldson, D. (2018, April). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review 108*(4-5), 899–934.

Einav, L. and B. Y. Orbach (2007). Uniform Prices for differentiated goods: The case of the movie-theater industry. *International Review of Law and Economics* (27), 129â153.

Fajgelbaum, P. D. and E. Schaal (2020). Optimal transport networks in spatial equilibrium. *Econometrica 88*(4), 1411–1452.

Firth, J. (2017). Iâve been waiting on the railroad: The effects of congestion on firm production.

Gil, R. (2007). âmake-or-buyâ in movies: Integration and ex-post renegotiation. *International Journal of Industrial Organization 25*(4), 643–655.

Gil, R. (2008, 04). Revenue Sharing Distortions and Vertical Integration in the Movie Industry. *The Journal of Law, Economics, and Organization 25*(2), 579–610.

Gil, R. (2015, May). Does vertical integration decrease prices? evidence from the paramount antitrust case of 1948. *American Economic Journal: Economic Policy 7*(2), 162–91.

Giroud, X. and H. M. Mueller (2019, October). Firms' internal networks and local economic shocks. *American Economic Review 109*(10), 3617–49.

Grossman, S. J. P. and O. D. Hart (1986). The Costs and Benefits of Ownership: a Theory of Vertical and Lateral Integration . *Journal of Political Economy 92*(4), 691–719.

Hart, O. and J. Moore (1990). Property Rights and the Nature of the Firm. *Journal of Political Economy 98*(6), 1119–1158.

Hastings, J. S. (2004). Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California. *The American Economic Review 94*(1), 317–328.

Herrnstadt, E. M., R. Kellogg, and E. Lewis (2020, May). The economics of time-limited development options: The case of oil and gas leases. Working Paper 27165, National Bureau of Economic Research.

Hodgson, C. (2021, October). Information externalities, free riding, and optimal exploration in the uk oil industry. Working paper.

Hornbeck, R. and M. Rotemberg (2021, March). Railroads, market access, and aggregate productivity growth. *Working Paper*.

HortaÃ§su, A. and C. Syverson (2007). Cementing Relationshis: Vertical Integration, Foreclosure, Productivity and Prices. *Journal of Political Economy 115*(2), 250–301.

Jaravel, X., N. Petkova, and A. Bell (2018, April). Team-specific capital and innovation. *American Economic Review 108*(4-5), 1034–73.

Kalouptsidi, M. (2014, February). Time to build and fluctuations in bulk shipping. *American Economic Review 104*(2), 564–608.

Kellogg, R. (2014, June). The effect of uncertainty on investment: Evidence from texas oil drilling. *American Economic Review 104*(6), 1698–1734.

Lee, R. S. (2013). Vertical Integration and Exclusivity in Platform and Two-Sided Markets. *The American Economic Review 103*(7), 2960–3000.

Lim, C. S. H. and A. Yurukoglu (2018). Dynamic natural monopoly regulation: Time inconsistency, moral hazard, and political environments. *Journal of Political Economy 126*(1), 263–312.

Luco, F. and G. Marshall (2018). Vertical integration with multiproduct firms: When eliminating double marginalization may hurt consumers. *Available at SSRN 3110038*.

Mortimer, J. H. (2008). Vertical Contracts in the Video Rental Industry. *The Review of Economic Studies 75*, 165â199.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Palia, D., S. A. Ravid, and N. Reisel (2008). Choosing to Cofinance: Analysis of Project-Specific Alliances in the Movie Industry. *The Review of Financial Studies 21*(2), 483–511.

Preonas, L. (2019, December). Market power in coal shipping and implications for u.s. climate policy. *Working Paper*.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica 55*(5), 999â1033.

Timmins, C. (2002). Measuring the dynamic efficiency costs of regulators' preferences: Municipal water utilities in the arid west. *Econometrica 70*(2), 603–629.

Villas-Boas, S. B. (2007). Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data. *The Review of Economic Studies 74*(2), 625–652.

Waldfogel, J. (2017). Cinematic explosion: New products, unpredictabilty and realized quality in the digital era. *The Journal of Industrial Economics* (64).

Yang, C. (2020). Vertical structure and innovation: A study of the soc and smartphone industries. *The RAND Journal of Economics 51*(3), 739–785.