THE UNIVERSITY OF CHICAGO

SHARED INTANGIBLES & TECHNOLOGY REVOLUTIONS: EVIDENCE FROM THE ENERGY SECTOR

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
NAM VERA CHAU

CHICAGO, ILLINOIS
JUNE 2022

This dissertation is dedicated to my father and his unwavering faith as well as my mother who once told ten-year-old me that the way to stay in school forever was to get a PhD. Most of all, this dissertation is dedicated to my sister who is a daily reminder to me that true intellect is bold, creative, and unafraid.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful to my committee, Steven Kaplan, Pascal Noel, and Constantine Yannelis for helpful comments and support throughout the years. A special thanks goes out to committee member Lars Hansen who has guided this project from the beginning through his working group. Most importantly, I would like to acknowledge my chair, Amir Sufi, and all the work he has put into advising me from the curriculum paper to the dissertation. I also want to thank the entire finance faculty at Booth for their advice throughout this process.

I acknowledge the Energy Policy Institute at the University of Chicago and Ryan Kellogg for providing the data from Enverus, Inc. used in this paper. The hallmark of a Booth PhD is our stellar colleagues. I want to acknowledge participants in Amir Sufi's working group, Lars Hansen's economic dynamics working group, and the finance brownbag for helpful comments. I would not have gotten to this point without my cohort-mates and homework partners Carter Davis and Mihir Gandhi. I also benefited from numerous discussions with Greg Buchak, Ana-Maria Tenekedjieva, Kelly Posenau, Gursharan Bhue, Agustin Hurtado, and Ehsan Azarmsa.

Finally, to the people who gave me the strength to make it here. First and foremost, Samuel Hirshman. To quote "Boy meets world", the best version of a partner that I could dream up would pale in comparison to you. I have relied on both your unfailing kindness and stubborn intellectualism and I look forward to more to come. To Connor Dowd, you were never too busy to discuss everything from econometric specifications to sports data with me and that has made me a better researcher. To Yewon Kim and Olivia Natan, who were always ready with an encouraging word or a reality check. To Sondre Skarsten who challenged my notions of what science is. To April Collaku and Sampoorna Dasgupta, I'm amazed at the paths our lives have taken and the way we stuck through it together. To Rachel Shapiro,

you've been more excited for every milestone of this journey than I have been and that is a light in a tunnel that gets very dark at times. To Lorie Chaiten and Harold Hirshmam who are my daily reminders of what perseverance and generosity can accomplish in the world.

To my father who never doubted what a mechanic's daughter can accomplish. To my mother who is a living testament that joy always comes with the morning. To my sister who is never constrained by what others think is possible.

ABSTRACT

Knowledge and experience learned from investing in physical, tangible capital is sometimes non-excludable between firms. I explore how this simple externality affects technological growth. Using the recent fracking revolution in American oil & gas, I develop a two-stage empirical procedure to 1) provide evidence that this shared intangible exists and 2) show that firms value this knowledge externality when making investment decisions. I use a spatial panel model that is a natural network structure for localized knowledge diffusion to identify cross-sectional differences in intangible capital value across counties. Then, I use global oil prices as a plausibly exogenous source of variation in investment levels. I find that one extra investment made by other firms in strong knowledge network areas is associated with a 13% increase in monthly productivity. I show that this effect is particularly important for growing technologies; tests using older methods of production do not have the same impact. In a heterogeneous firm model, I formalize how this mechanism drives growth cycles by effecting technology improvement and widespread adoption jointly. As more firms learn by doing, the technology improves for everyone. As the technology improves, more firms invest. Technology transition is the result of firms optimally re-weighting their capital portfolios towards newer technology. Because tangible capital investment is the primary mechanism for technology growth, the distribution of firms in the economy becomes an important state variable which determines the rate of technology adoption and the phase-out of old technology.

CHAPTER 1

1.1 Introduction

Widespread adoption of a new technology as well as observed improvement in that technology often occur simultaneously. These episodes of technological growth also have important implications for broader economic growth. Consider for example, the introduction of horizontally fracked wells into the American oil & gas exploration and production industry. This development had a profound impact on geopolitics, moving the country from a net importer to a net exporter of fossil fuels. Despite its importance, the literature on technological change and endogenous economic growth is largely focused on discovery or innovation as opposed to improvement and diffusion. As Stokey noted in a recent working paper, "The diffusion of technological improvements, across producers within a country and across international borders, is arguably as critical as innovation for long run growth."[1]

This paper studies these two processes, improvement and adoption, through the lens of one such episode, the American fracking revolution. Importantly, both improvement and adoption are difficult to study empirically. Thus, even as individual mechanisms, little is understood about them.[2] This paper tests an idea which sheds light on the relationship of growth and diffusion to each other and the joint effect they have on periods of marked technological growth. Firm investment in physical capital creates knowledge and experience which is not always excludable from other firms. For example, oil & gas firms have to submit permits to drill which makes some information regarding their production process publicly available. This simple externality, the accumulation of shared intangible capital through physical capital investment, results in a novel mechanism which sheds light on how technology revolutions occur. Firms learn as they invest in new technology but due to sharing, that

1. [63]

2. [63] reviews the literature on diffusion for the cases where data are available.

learning improves the technology for everyone. At the same time, as the technology improves, the physical capital investment decision for all firms is affected by the creation of valuable, shared intangible capital. This creates a self-sustaining endogenous growth mechanism which is explored in a theoretical companion paper.[3] In this paper, I use this framework to build reduced-form empirical tests of both sides of this feedback loop. I show that physical capital investments results in the creation of valuable shared intangibles evidenced in productivity differences. I also show that investment activity by other firms have a larger impact on firm investment decisions when shared knowledge is valuable.

The American fracking revolution in oil & gas presents an interesting empirical test case of this mechanism. The discovery and subsequent growth of hydraulic fracturing technology is one of the most recent examples of an influential technology growth episode. Additionally, the American oil & gas sector has close institutional similarities with renewable energy sectors such as wind and solar. Given the challenges of climate change, the oil & gas sector presents an interesting opportunity to study the economics of technology growth with applications to similar, prescient sectors. In particular, sectors which extract natural resources are monitored by permits and the land leases for oil wells and solar panels both require contracts which are recorded by local municipal authorities. While these are important policies for resource management, they also make knowledge and experience partially available to all firms in the industry. Different versions of these externalities exist in other industries of interest as well. In life sciences and biotechnology, it is not uncommon for firms to publish peer-reviewed research to test and prove their technology. This is a process by which other firms may learn about technology developments industry-wide. In consumer finance, many firms are building products with analyses from the same set of credit-bureau data (ex: Equifax and TransUnion). This shared data pool also results in shared knowledge and

3. This theoretical companion uses a generic production function which is not energy specific.

experience.

I develop a novel empirical strategy to estimate this knowledge externality from physical capital investment. The procedure is two-fold. I first identify cross-sectional variation in the usefulness of this shared knowledge or experience by estimating a network model. Then, I find exogenous changes to investment levels which impact all firms in the economy. If there are areas where the creation of this shared knowledge is more useful, then the decline in investment levels should disproportionately affect the productivity in those areas since investment by other firms in the economy contribute disproportionately to the effectiveness of technology in those areas.

To implement this framework, I rely on natural network structures which intuitively govern the propagation of shared knowledge. In natural resources, geological differences suggest that knowledge proliferation is not friction-less industry-wide[4]. Information produced by another investment may not be as valuable if it is too far away or if it is in a geology that is so complex that there is significant heterogeneity in technical design[5]. Other examples of natural networks might be trade flows which are determined by historical political agreements instead of economic benefit. I implement the network structure by estimating a spatial autoregressive model which weights the investment effect from other firms based on their inverse distance[6]. This measure is estimated in the first stage using detailed capital unit-

4. Note that I do not assume that there are no global knowledge transfers. Labor movements can result in knowledge transfers that are not limited by distance. Rather, this procedure precisely captures the local effect and differentiates it from the global one.

5. Importantly, this network is exogenously formed. While firms may choose to invest near other firms to take advantage of the network, the fact that the networks differ in their efficacy remains exogenous. The fact that information from some investments are more useful in some areas is due to exogenous factors such as geology. Said differently, while the size of networks may be endogenous due to the investment choices firms make, the heterogeneity in network effectiveness is not.

6. This model is studied in-depth by [15] and has been implemented in house price contagion studies and environmental impact analyses.

month level data and the model is implemented separately for each oil producing county in the United States to capture cross-sectional variation in network strength. To avoid measuring other sources of contagion such as geological productivity, I use indirect measures as explanatory variables. Rather than studying the network effect of productive investments on productive investments, I study the effect of proximity to skilled firms on productive investments. I also consider the impact of more information about nearby investments on the success of my own well. If investments do not rely on the experiences generated in the network then the skill of firms nearby should not explain variation in other wells.

The network strength estimated using this spatial autoregressive model forms the basis of "treatment" and "control" distinctions in a differences-in-differences ("diff-diff") framework. I then use the decline in global oil prices starting in 2014Q3 to instrument for sustained investment declines industry-wide. The full empirical specification is a hybrid between a "diff-diff" and an instrumental variable specification. In a traditional diff-diff, one studies how the treatment group differs in their reaction to an exogenous policy change. Here, I study how the treatment group differs in their reaction to investment changes as instrumented by the oil price decline in a two-stage least squares specification. I measure the effect of lower investment activity by other firms on both lifetime output measures and other performance indicators. The baseline results show that investment activity by other nearby firms have a larger effect on the productivity of firms in strong network areas as compared to weaker network counties. The marginal effect of one additional well drilled nearby in high knowledge sharing regions is a 13% improvement in well output *every month of its life.* The analogous effect in low sharing areas is negative.

I employ this empirical strategy for both new and old type technology. Each county is sorted once based on its new technology network strength and its old technology network strength. I then test the impact of decreasing investment levels for both production methods. Notably, the disproportionate impact from other firms making physical investments only hold in the new technology. Despite measuring a strong productivity network in the old type of technology, the results show that the coefficient on investment activity is *decreasing* in network strength. This indicates that shared knowledge as a spillover is particularly important for growing technology. This is distinct from other sources of productivity spillovers which may be relevant in a number of scenarios. For example, increased efficiency from supply chain effects because a similar plant is opened nearby may result in productivity spillovers even for mature technology.

I formalize the mechanism in a heterogeneous firm model which incorporates both the shared knowledge and the investment choices. Using aggregate transition dynamics of investment and technological sophistication, I develop intuitive results which highlight the theoretical contribution of this framework. In particular, the mechanism for technological change is shown to be changing incentives for firms to improve their technology sophistication rates as compared to making larger level investments in new technology. Further, the results show that the distribution of firms operating in the economy is important for the rate of technology transition as well as the ultimate productivity of that new technology. Empirical predictions from this theoretical discussion is then used to study the other side of the feedback loop. If there's shared intangible capital creation, then increased investment activity by other firms should result in higher investment levels since I can take advantage of the knowledge. Because the oil price shock affects all investment levels, it is not useful for this analysis. Instead, I turn to a different source of exogenous variation in investment levels, the expiration of leases which give firms the option to drill in a given area. This expiration date has been shown in the literature to correspond with increased investment activities as firms try to retain the option. The results show that increased investment by other firms corresponds to higher firm-level investments in strong estimated network areas as compared to weaker ones. Investment activity broken down by firm characteristics are consistent with predictions from the model. While larger firms make larger level investments in new technology because of the size effect, it is smaller firms that make larger technology adjustments.

Finally, I study the broader economic effects from this shared knowledge effect. Relative to investments in other regions, firm investments in strong network areas tend to outperform when aggregate investment by other firms is high. Over time, the relative outperformance in strong network areas parallels industry-wide growth trends. In other words, when high knowledge sharing areas stop benefitting from the creation of shared intangibles, industry-wide technology growth also slows down.

The paper proceeds as follows. This introduction concludes with a discussion of related literature. Section discusses the empirical framework including the empirical specification, discussions of the strategy and thought experiment, institutional details, and descriptions of the data. Section 1.3 contains the baseline productivity results including mechanism and robustness tests, section 1.4 discusses the theoretical model and intuitive results, section 1.5 studies the investment and adoption side of the feedback loop, section 1.6 shows the effect on industry-wide trends, 1.7 discusses additional mechanisms such as learning, and section 1.8 concludes. More technical details and discussions are left to the appendix. [7]

1.1.1 Related Literature

My paper approaches two large strands of literature from the corporate finance perspective, endogenous growth and technological change. To do this, I draw on insights made in the

7. The appendix materials and model appendix are available at https://www.nverachau.com/research

productivity and capital allocation literature.

The endogenous growth tradition is focused broadly on characterizing the relationship between technological change and broader economic growth. Works such as [62], [46], [9], [47], and [8], explore different ways in which the actions of agents within the economy can influence growth. This literature makes use of the economics of ideas and integrates it into technology change and economic growth. My paper also relies on a key characteristics of ideas, non-rivalry and non-excludability. However, a more recent set of papers focusing on the process of technology change specifically is probably closer to this paper. The early endogenous growth papers analyzed idea creation more generally and tended to prescribe educational reform policies and other ways to push out the technology frontier. A sub-genre of the growth literature became more interested in understanding the mechanisms which not only lead to technology change but allows that change to infiltrate the economy. These include [3], [14], [13], [11], [12], as examples. These papers more closely examine the process of technological growth. Considerations include the tax regime, the market structure, regulatory protection of innovation, and financing. My paper is differs in studying technology improvement through learning as opposed to the process of new technology discovery. Economically, I move away from the idea that direct innovative investment activity is necessary for technology change and study inherently innovative production.

Two closely related mechanisms have been studied alongside the literature on technology growth. The first is adoption and diffusion. While it is distinct from the technology change model, adoption and diffusion is important for understanding how technology change can lead to economic growth. Empirically, this has been a difficult topic to study because the data is hard to acquire. [42] was a seminal paper looking at diffusion in a specific industry. [63] surveys the handful of other papers that have conducted this exercise in other industries and [43] gives a good overview of the literature. [23] is a recent example of studies in this space. Finally, [29] is an example of how this technology diffusion idea has been used to analyze inter-country growth. Most of these papers focus on understanding how technology growth leads to adoption of new technology. In my paper, I also allow for the possibility that changing adoption rates impact technology growth in return. Further, in applying this to the oil and gas industry, I add to the set of empirical studies trying to estimate this effect.

The second mechanism which is closely related to the technology growth literature is that of spillovers and imitation. [60], [66], and [20] are three very different approaches to modeling imitation and social learning. In the spillovers space, much of the work has been applied to either R&D spillovers as in [25] and [56] or productivity spillovers as in [41]. Although distinct from the work on productivity spillovers, the large literature on agglomeration often incorporate similar ideas so I note them here. For example, [18], [28], [57], and [53].

Knowledge creation, sharing, and adoption have also been a critical input in a variety of contexts. I draw on this diverse set of literature in thinking about how to apply the economics of knowledge. I contribute empirically by studying the energy sector specifically but also theoretically but using the spillover mechanism in a model of heterogeneous firms. The implications for aggregate technology growth differ slightly as a result. A paper which studied knowledge sharing in a specific industry is [64] whereas [58] analyze adoption of new technology in the consumer finance industry. There is also growing interest in data as a capital input and the economics of knowledge given the prevalence of big data. For example, [24], [36], [2], and [37]. Finally, there is a large set of literature studying how the ability to create knowledge given different macroeconomic conditions impacts economic growth. Examples of this include [30], [16],[22],[10]. My paper distorts the knowledge creation process and develops a tractable equilibrium model which could be incorporated to study larger macroeconomic effects.

While the main economic effect is related to technology change, the empirical application in energy is directly related to papers trying to understand the role of technology change in energy and environmental economics. In particular, my paper contributes to studies of the energy transition which is a natural application of the adoption problem solved by firms in my model. Further, my empirical work showing the impact of aggregate investment levels on technology growth have implications for the rate of growth and adoption in a wide variety of critical industries such as renewables. Papers along this vein include, [4],[61], [5], and [59]. On the other hand, a number of papers have used the energy space as a test case for broader economic concepts. Examples include: [51], [50], [31], [40], [32]. In particular, [32] and [33] both study peer effects in real option exercise using this same institutional setting as a case study. These paper show that firms exercise their real options to drill strategically with these peer effects in mind. My paper shows what those peer effects are and how they manifest in production.

Finally, I drew heavily on existing literature in corporate finance, capital allocation, and firm life-cycle studies in building both the empirical and theoretical framework for this paper. While these are not directly related to technology change, my paper incorporates technology into the problems these papers studied. In return, my work also has implications for how we study firm valuation and external financing. On the external financing front, papers like [55] and [54] study the role that private equity and venture capital have played in innovation. Regarding models of investment, I drew on a large literature including [35],[27],[44],[45],[26],[1],[21]. My model of technology adoption was heavily influenced by both macro models of capital allocation and two-sector models such as [34] and [6] as well as internal capital allocation models such as in [19]. The idea that knowledge spillovers may rely on aggregate investment changes was drawn in part by the literature on firm life-cycles and how their investment and innovation propensity may be different. These include, [52],[48],[65],[17], and [49].Finally, my paper is closely related to a set of papers that study productivity more generally such as [39] and [38]. My study of technology improvement bridges the gap between technology discovery and productivity growth.

1.2 Empirical Framework

The empirical strategy of the paper can be intuitively described as a network stress test. The first step is to estimate the network strength in each geography that produces oil using the spatial panel model. Then, I use an exogenous shock to investment levels which impact all areas. If there is a strong network effect then removing nodes of the network through lower investment levels should disproportionately affect the stronger network areas. This should manifest in larger performance depreciation if the network was in fact valuable. This approach presents an advantage. One can study the precise channel through which this shared experience leads to technology growth. By simply changing the dependent variable used to estimate the first stage network effects, I can study whether the shared experience proliferates through technical design or methods or even cost reduction. Additionally, it's possible to examine instances where this shared knowledge is not valuable. For example, I show that this network effect isn't strong in older, mature technology. A challenge to this method is to find settings where shared knowledge value might vary exogenously. In the institutional detail section 1.2.2 below, I motivate this variation in the oil & gas example.

1.2.1 Empirical Strategy

The ideal specification would be a differences in differences ("diff-diff") framework where regions are divided into treatment and control groups based on their knowledge sharing propensity, ks or the value of shared knowledge. Contrary to many existing strategies which studies the effect from areas that receive the knowledge shock to ones that do not, this study compares areas which all receive the knowledge shock but that knowledge is more useful in some areas. If such a measurement existed and investment was exogenous, one could run,

$$Y_{w,t} = \alpha + \gamma I_{g,t} \times ks^{Treatment} + \epsilon_{w,t} \tag{1.1}$$

where Y is an outcome of interest such as production associated with well w. $I_{g,t}$ captures the investment level in the geography g at month t.[8] However, this specification is not usually available.

I use the period of sustained decline in oil prices which also corresponded to a period of investment decline industry wide. Figure A.3 shows time series plots of the WTI (West Texas Intermediate) oil price and corresponding monthly investment levels between 2010 and 2020. Oil prices begin to decline in the fourth quarter of 2014 and shortly afterwards, investment levels fall. I use this regime switch in oil prices as an instrument for changing investment levels in an instrumental variable framework.[9] Notice that this requires an additional modification to the specification in equation 1.1. In order to compare the impact of investments across knowledge sharing regions ("treatment"), I divide the knowledge sharing measurements described below into quartiles, n, and run the following 2sls specification individually in each bucket.

$$\begin{aligned} log(O_{w,t}) &= log(Age_{w,t}) + \gamma^{ks_n} \hat{I}_t + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage) \\ I_t &= \eta + \beta^{ks_n} \mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage) \end{aligned} \tag{1.2}$$

8. Recall that knowledge spillovers assume any investment could potentially impact all firm investments.

9. A more direct implementation would be to use the oil price regime as an exogenous policy shift in a traditional diff-diff setup. However, this specification is more precise in that it captures the impact on Y due to variation in investment as opposed to a simple pre-post period treatment. Also, I do not want to assume ex-ante that the investment elasticities are the same across knowledge sharing quartiles. I want to empirically estimate that.

$\mathbf{X}_{g,t}, \mathbf{X}_{i,t}$ is a vector of geography and firm level controls described in the appendix. The controls capture factors such as geological formation quality and firm skill which may impact production in addition to knowledge sharing. Each specification is run within each quartile n bucket of estimated network strength, ks.

The intuition behind the strategy is simple. Higher estimated network areas can be considered "treatment" groups while weaker network areas act as controls. Both areas are subjected to the same exogenous decline in physical investment level. If the networks are indeed valuable then removing a node of the network through lower investment levels should disproportionately impact areas where the network connections are strong. Figure A.4 illustrates the overview of the main empirical thought experiment.

A significant threat to identification is that the oil price instrument violates the exclusion restriction. The key assumption here is that while oil prices impact the extensive margin decision to drill a well, it has limited impact on the intensive margin production variables once that well is drilled. Conditional on paying the fixed cost of exploration and drilling a well in the first place, firms will want to drill the most productive wells possible.[10] The second panel of figure A.3 also contains a panel showing the average well productivity of wells in the dataset. Over the same period of declining oil prices and investment, there is not a corresponding decline in well productivity *on average*.

The remaining hurdle is finding variation in the value of shared knowledge. The discussion below proposes that natural networks which are the result of institutional details present a

10. Natural resource firms using more traditional methods in places with large oil reservoirs make use of various practices to keep the oil in the ground. This is to take advantage of the real option value in waiting during times of high oil price uncertainty. Until the recent Covid pandemic when oil prices dropped below the break-even price for most American shale producers, there is limited evidence that American firms use the practice in meaningful numbers.

useful method of measuring this variation. First, I discuss the institutional details of the energy sector which offers one such natural network through geology.

1.2.2 Institutional Details

The discovery that horizontally drilled wells can be combined with the hydraulic fracturing technique known as "fracking" unlocked vast resources in the continental United States and revolutionized American oil and gas. The first panel of figure 1.1 below shows the progress of that technology since its inception. The plot shows relative productivity of horizontally drilled wells as compared to its traditional vertical counterparts. In addition to being more productive at discovery, the new technology has continued to grow relative to the old technology[11]. The second panel shows the average count-level investments in new technology measured by horizontal wells drilled. The third shows industry-level intensive margin adoption rate. Over the time period when individual investments exhibit substantial technology efficiency gains, firms are also adoption more of the new technology into their physical capital stock. In other words, the growth in intangibles illustrated in the first panel of figure 1.1 parallels the time series growth in adoption rates in the industry. Importantly, while technology improvement does not decline given the continued adoption rate increases in panel three, it does slow down when investment levels in the second panel drops. If shared knowledge is important than both aggregate levels and rates contribute to the creation of intangible capital. To further motivate the idea that shared knowledge may be important, the fourth panel limits the analysis to the first well that a firm drills in a given county. For these wells, firms are likely to rely more on knowledge created by other firms as they have no experience in that geography. The plots show the relative performance of first wells in the strongest network quartiles as compared to the weakest. Industry-wide technology improvement trends parallel not only the aggregate industry level investment time series but

11. The plot shows results from regressions of monthly oil output on a dummy variable indicating that the output is from a horizontally drilled well.

the relative performance in higher network areas as well.

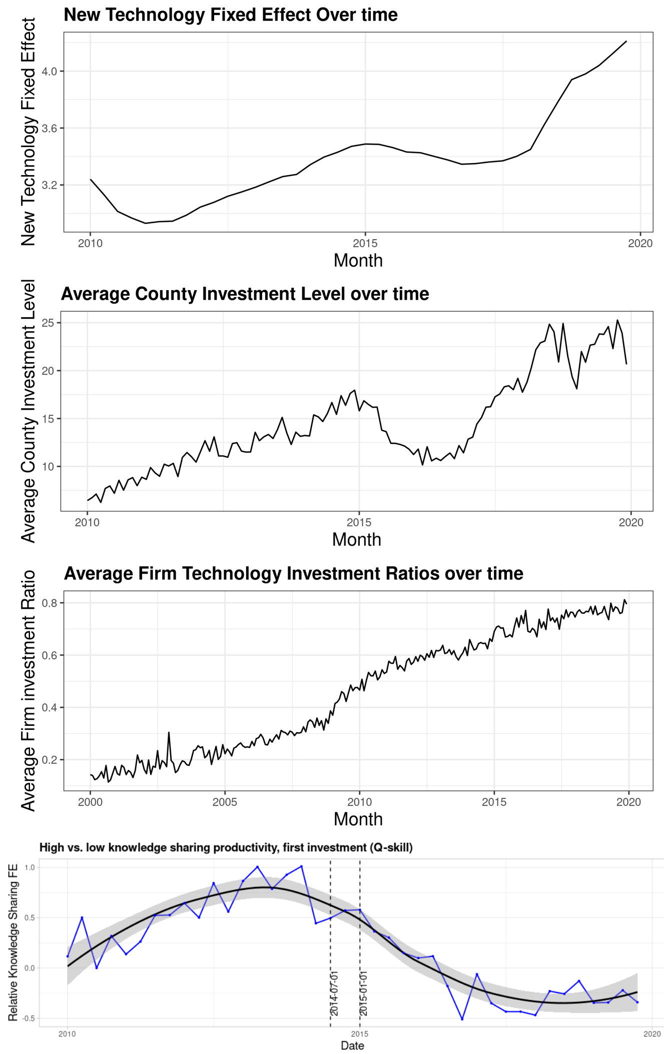

Figure 1.1: Empirical Motivation

The first panel plots coefficients from,

$$log(Y_{w,t}) = \beta \log(Age_{wt}) + \gamma^{q} \mathbf{1}_{\{Horizontal\ well\}} + \epsilon_{wt}$$

The model is a log version of a common production function used in the oil literature. The regressions are conducted at the well-month level with each specification re-done for each quarter of the sample. The second panel plots average monthly county level investments in units of wells drilled while the third shows average firm-level technology investment ratios in. The fourth panels show results from regressions,

$$\begin{gathered} log(O_{w,t}) = \beta log(Age_{w,t}) + \gamma_t \mathbf{1}_{\{ks^n, n=4\}} + \epsilon_{w,t} \quad |n = 1, 4 \\ new\ type\ only, \quad first\ wells \end{gathered}$$

It depicts the relative performance between the highest and lowest knowledge sharing counties of the first well a firm drills in a new county. Compared to all wells, the first well drilled is likely to rely more on existing knowledge created by other firms.

The institutional details section in the appendix discusses the oil and gas upstream process for the interested reader. Here I summarize the details which are relevant for this analysis. Unlike the global oil industry, the American oil and gas industry is large and dispersed. Small independent exploration and production firms drill alongside British Petroleum and Exxon Mobile. This makes shared learning particularly potent compared to industries like pharmaceuticals where large firms hold monopolies on resources. While there certainly exists "general purpose knowledge" in the drilling process, there is also geology-specific expertise. This allows for plausible variation in knowledge acquisition and usefulness. Figure A.1 shows the variance of drilled *depths* as well as the spacing of wells drilled in North Dakota and Texas respectively. The North Dakota example shows relatively well lined wells (most of these are horizontal so they line up accordingly). Also the depths are fairly homogeneous as depicted by the colors. Texas on the other hand shows wells drilled at a variety of depths. The example does not suggest one area is more productive or easier to drill. With the more homogeneous wells, it is expected that optimal well design will be similar making experience more easily transferred. On the other hand, there is more experience in Texas so information regarding what does not work is well developed. This heterogeneity will be a useful building block towards differences in knowledge sharing propensity.

The final good produced in the oil and gas space is a uniform good with limited variation in quality, a single barrel of oil. Additionally, fracking revolution in oil and gas was a recent technology revolution that had a meaningful impact on economic growth and data is available on both the older method of production and the new technology. Natural resource extraction requires firms to report technology details and production updates back to state regulators which results in a rich set of data. Finally, oil and gas extraction contains many similarities in regulation and market structure to other industries of particular interest such as wind and solar energy. In section 1.5, I take advantage of the leasing structure in American oil and gas. Unlike nationalized oil companies, American oil companies have to acquire leases which grant the firm the right to drill in a particular parcel of land. These leases are signed with individual land owners as well as the federal Bureau of Land Management. These leases give firms the right to drill during the primary term. If that primary term expires before a single well has been drilled, the firm loses the option to drill on that land. The expiration of this option presents a useful proxy for increased investment activity which will be used in the investment analysis of section 1.5.

1.2.3 Measuring Shared Knowledge Value

There are two ways to estimate the effect from sharing knowledge. The first is to change the availability of shared knowledge. This is considered in this paper through the data availability analysis. Institutionally, this study is possible because some firms require more data disclosure than others. The second method is to vary the usefulness of shared data. I propose a novel empirical approach which can be replicated in a number of circumstances. The idea is to find natural, exogenously constructed network structures which one can use to estimate cross-sectional network strengths. For example, in the oil & gas context, this network is spatial distance. Wells which are drilled close by are going to be more useful to you than wells drilled extremely far away. This does not eliminate the possibility of other types of knowledge transfers. For example, if an engineer moves from one state to another, their knowledge is likely to be useful as well. This spatial network extracts the impact of spatial distance which is particularly useful in the data sharing context and shoould be orthogonal to the human capital transfer.

Network Model: Spatial Panel I use a spatial autoregressive model commonly used in the household contagion literature to capture a localized network effect. The intuition behind the spatial lag model is to capture the impact of nearby, recently drilled wells ("influence") on a particular well of interest ("reference"). The model implements "nearby" using a physical distance measure and weights potential influence wells by the inverse distance to the reference well.[12]

Problematically, the direct effect of influence well k's outcome on the reference well l's outcome is the clearest estimate of both knowledge spillovers and other sources of contagion. While this would be the ideal measure to use, it is difficult to disentangle the two. Instead, I use an indirect measure of the relationship between influence and reference wells, firm skill. Using the nation-wide sample of production, I estimate a firm's quarterly skill level. The specification is described in section 1.2.4. I study the relationship between the skill of the firm that drilled the influence well and the outcome of the reference well. This skill measure introduces variation in ability beyond the influence of the area's inherent quality. The correlation between firm skill and approximations of the geology quality is around 0.19. The assumption underlying this indirect measure is that if no knowledge sharing existed, investing near a skilled firm should not explain variation in outcomes[13] If there is clustering of successful investments for other reasons[14], then much of that variation should be absorbed by the controls which include the lag outcomes. Further, in an inherently productive area drilling near a skilled firm should be no more informative than drilling near an un-skilled one. This knowledge sharing methodology captures a specific spillover effect based on the idea that as distance between wells increases, the knowledge will be less useful. Of course, it will not capture other sources of data sharing which could be present. For example, if firms

12. Note that for simplicity, I only use surface distances. With oil drilling, the depths represent different geological structures or "pay windows" as they are known. I do not use this dimension of distance.

13. It is expected that clustering around skilled firms is not uncommon. Thus, this will capture instances where that clustering is productive. Note that because oil production tends to decline when too many oils are drilled close to each other, it would not be unexpected for this effect to be negative.

14. ex: suppose a geological area is just extremely productive

share oilfield service providers as in [50], the inverse distance weighting is not the right way to capture that network effect. There's no reason why data-sharing between two unskilled firms cannot be useful.

The spatial lag model which estimates the network strength in the first stage of the empirical strategy is given by,

$$\mathbf{Y}_t^{g,H} = \gamma_1^{g,H}\mathbf{W}\mathbf{Y}_{t-1}^{g,H} + \gamma_2^{g,H}\mathbf{X}_t^{g,H} + \beta^{g,H}\mathbf{W}\mathbf{S}_{t-1}^{g,H} + \epsilon \tag{1.3}$$

Details regarding the econometrics of this specification is available in the appendix under the "spatial lag model" section. The outcomes of interest used in $\mathbf{Y}$ will be discussed in section 1.2.4 below. Here I point out some notable insights from this specification. The studies are conducted at the well level. t denotes the month when a well w which appears in row w pf vector $\mathbf{Y}$ occurs. $\mathbf{Y}_{t-1}$ contains a zero for well w but contains values for the outcomes of interest in rows where a well was drilled at time $t-1$.[15] γ_1 absorbs contagion effects. The inclusion of lags of the variable of interest should assuage concerns that the network model is only capturing clustering. In some co the empirical tests, I will utilize this coefficient, $_1$ to test this proposition. $\mathbf{X}$ includes geograpy level controls of interest such as the natural productivity of a particularly geography. The superscript H and g indicate that this specification is run separately for each county g and only during the initial high oil price regime. This method is similar to the "pre-sorting" technique often used in empirical asset pricing for portfolio sorts.[16] Finally $\mathbf{S}$, which is referred to as the indirect effect variable are chosen so that $\beta^{g,H}$ is unlikely to capture direct contagion effects. They are discussed in detail below. The model is a random effects model which is estimated using maximum like-

15. Again, this specification in matrix form is available in the appendix.

16. Note that the county is the grouping variable used to run the estimates, it does not define the set of wells that could possibly influence the reference well

lihood estimates. I assume the strictest error structure, the model does not allow for spatial correlation or correlation over time. The spatial effect is assumed to be in the mean.[17]

1.2.4 Indirect Effects & Outcomes of Interest

The outcomes of interest used in the spatial lag model B.4 as well as the skill variable relies on the commonly used production function for oil production. This is another reason why the oil and gas case study is particularly useful, the production function is well known and easy to estimate. The basic Arp's model is the most basic production function taught to geology students. It is given by,

$$O_{w,t} = Q_w Age_{w,t}^{\beta} \tag{1.4}$$

$$log(O_{w,t}) = \alpha_w + \beta log(Age_{w,t}) \tag{1.5}$$

where the second line contains the log form. The intercept, α_w captures baseline productivity rate $log(Q_w)$ which will be used as the outcome of interest $\mathbf{Y}$ in many of the network estimates.

This production function forms the foundation for a number of analyses throughout this paper. It is used to estimate firm skill but it also forms the basis for the main empirical specification. In particular I model $Q_{w,t}$ as a linear function of firm and geography level controls as well as the investment activity of other firms. This impact of investment activity on output $O_{w,t}$ through the linear model of $\hat{Q}_{w,t}$ forms the baseline productivity effect in section 1.3.

17. This assumption may be relaxed in robustness tests. However, it should be noted that correlation in the error structure is likely to be a completely different network. This would also be interesting but the empirical tests in this paper tests the network strength of the network as estimated.

The Firm skill vector **S** The vector **S** contains estimates of firm i's skill in the month t if a well is drilled that quarter and is zero otherwise. To estimate firm level skill, I modify the Arp's model in equation 1.4,

$$O_{w,t} = Q_w Age_{w,t}^{\beta}$$
$$log(O_{w,t}) = \alpha_w + \beta log(Age_{w,t}) + \epsilon_{w,t}$$

where $\hat{\alpha}_w$, the constant, essentially captures $log(Q_w)$. I amend the model to estimate a version of baseline production rate Q_w that is firm-specific as opposed to well-specific. Using the full dataset with every horizontal well drilled across all geographies I estimate,

$$log(O_{w,t}) = \beta log(Age_{w,t}) + \gamma_{i,q} \mathbf{1}_{\{Firm\ i\}} \times \mathbf{1}_{\{q\}} + \epsilon_{w,t}$$

Rather than α_w capturing the baseline production rate of well w, this model has $\hat{\gamma}_{i,q}$ which captures the firm-quarter baseline production rate. This coefficient is then used to index the firm-quarter skill level.[18] Vector S_t then contain this estimate, $\hat{\gamma}_{i,q}$ if firm i drilled a well in the quarter that contains month t. Importantly, if a firm drills multiple wells in the same quarter in the county, their skill estimate is removed from S_t so that a firm's own skill cannot impact its outcome.

Data Availability In some of the analyses, I'll substitute the indirect effect variable **S** for data availability matrix **A**. This variable measures when a supplementary data point is available for the influence well. This specification is detailed in the appendix section "Data

18. A similar exercise is done to obtain the geography-quarter level "skill" or production controls but with $\mathbf{1}_{\{Geo\ g\}} \times \mathbf{1}_{\{q\}}$ instead of the firm-quarter fixed effect

Availability Measure Values". The sets of data are divided into two, the first one contains general variables useful for improving the well production. The second set provides performance data related to the well such as flowing tubing pressure. For each set, a matrix with each well along the rows and each variable along the columns is defined with 1 in spaces where the data point is available for that well and 0 otherwise. The spatial lag specification then regresses Q, the baseline specification measure on the matrix of data availability. I then average over the coefficients for each data set to create a final "data availability" measure. There exists one for each data set.

1.2.5 Discussion of the Data

State regulators require operators to report their monthly production and these data are made publicly available. Monthly production data include total barrels of oil produced or oil equivalent amount of gas. The data used in this paper is compiled by Enverus[19]. The company's analytical data is widely used in the industry. Even the federal energy information administration (EIA) uses the data from Enervus in its reports and assessments of the American oil industry.

The data include the well-month production information along with their corresponding, static permit and lease information. Each well is associated with an operator who may or may not be the only firm actively operating the well. I use this variable to identify firms. Investments are measured by the number of wells drilled and they are recorded for this project in the month that the well is completed. The monthly production data is matched to the relevant permit by the API number which is used to identify all wells in North America.

19. Prevoiusly DrillingInfo

To be included in the data sample, wells must meet several requirements. First, the well has to be drilled after 2005 even though I only start the analysis in 2008. A well has to appear in the dataset for at least 30 months. This allows me to estimate lifetime production rates with more accuracy and removes wells that may have prematurely stopped production. While I don't explicitly remove counties on well count, there needs to be sufficient observations for the spatial lag model to be estimated so counties with very few wells are removed. The summary statistics data shows the average well count within the geographies that are included. The counties included are a random sample of all oil-producing counties in the continental United States.[20] No offshore or Alaskan wells are included.

1.2.6 Summary Statistics & Descriptive Facts

Table 1.1 shows summary statistics for the random sample of counties included in this paper. The full list of counties as well as the number of wells drilled in each one is located in the appendix under the "County list" section. Generally, the summary statistics show a wide distribution of wells and firms within each county. Further, investment activity between the two oil price regimes declines slightly but there is still significant activity. Table 1.2 describes the network effects which are estimated in the spatial lag model. In particular, it shows regressions of the estimated $\hat{\beta}$ from the spatial lag model in equation B.4 on various indicators of geological variance. In the institutional detail section, I used geological complexity as one motivation for why there may be variation in observed network effects. The results are broadly consistent with this idea. Estimated network strengths, $\hat{\beta}$ are significant and negatively correlated with proxies for well complexity. The coefficients on geology complexity and productivity are also negative but not significant.

20. This is due to computation limitations. The full sample of all counties will be included in the near future. The current county list is available in the appendix.

	N	mean	sd	min	25%	75%	max
	2008Q1 - 2014Q3						
Counties/well counts	90	359	399	6	67	512	1,486
Counties/firms	90	14	10	1	6	20	62
Firms/well counts	686	47	159	1	1	17	1,876
wells/months produced	32,377	66	79	1	13	147	222
	2014Q4 - 2017Q4						
Counties/well counts	87	235	313	1	31	277	1,276
Counties/firms	90	15	10	1	7	20	49
Firms/well counts	400	51	139	1	2	34.5	1,368
wells/months produced	51,705	49	36	1	31	46	184
	Investment: 2008Q1 - 2014Q3						
Counties/wells	90	333	371	3	50	525	1414
Counties/firms	90	124	119	3	35	175	583
	Investment: 2014Q4 - 2017Q4						
Counties/wells	90	240	330	0	30	259	1,433
Counties/firms	87	82	103	1	18	97	583

Table 1.1: Summary Statistics

The table shows summary statistics for the distribution of wells and firms across the different counties. The first column shows the counts for the first variable in the description, the rest of the rows then shows the distribution of the second variable in the description. For example, the first row indicates there are 90 counties which were randomly drawn to be included in the study. Across those 90 counties, there were an average of 359 wells drilled.

The statistics are also divided between the oil price regimes which will be used in the main empirical specification. Note that there may be overlaps. For example, there are 90 counties in the sample in the earlier period of the top panel and 87 counties in the later period. These are not necessarily new counties nor are they necessarily the exact same counties.

	Dependent Variable: $\hat{\beta}$, Q-skill network estimates					
	(1)	(2)	(3)	(4)	(5)	(6)
log(sd(Interval ft))	-0.064***					
	(0.0072)					
log(sd(True measured depth (ft)))		-0.0038				
		(0.0022)				
log(sd(IP (6 mon)))			-0.0016			
			(0.0044)			
log(sd(oil-gas ratio))					0.0024	
					(0.0017)	
Observations	1565	1565	1565	1565	1273	1565
R^2	0.139	0.139	0.139	0.139	0.188	0.147

Table 1.2: Standard Deviation Tests

The table shows results from regressing $\hat{\beta}$ from the spatial lag model denoted in equation B.4 on various measures which proxy for the exogenous complexity of different geologies. These regressions include geography level controls which are also used in the baseline specification and detailed in the appendix.

Interval (ft) is the interval length measured in feet. This acts as a proxy for well complexity. True measured vertical depth measures the depths of the oil wells that are drilled. This is the variable that is shown graphically in figure A.1. Where the underlying geology has several layers of useful resources, the variance of measured depths are likely to be higher. IP rates is also known as initial production rates which is used as an industry indicator for the potential productivity of the well. This is measured as the total barrels of oil produced in the first 6 months. The oil-gas ratio measures the county level oil vs natural gas production ratio.

1.3 Empirical Results: Baseline Technology Growth Effects

Figure A.6 shows results from the main empirical specification given in equation 1.2[21],

$$log(O_{w,t}) = log(Age_{w,t}) + \gamma^{ks_n}\hat{I}_t + \Gamma\left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage)$$
$$I_{g,t} = \eta + \beta^{ks_n}\mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma\left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage)$$

The figures plot the second stage coefficients, γ^{ks_n} which are separately estimated for each of the four estimated network strength quartiles. The first stage regresses $I_{g,t}$, investment by other firms in county g at time t, excluding the firm that drilled well w, on an indicator for the oil price regime. The second stage starts from the log version of the classic Arp's production function for oil wells

$$O_{w,t} = QAge^{\beta}$$
$$log(O_{w,t}) = log(Q) + \beta log(Age)$$

For the regression analysis, I add the instrumented investment level $\hat{I}_{g,t}$. I model baseline productivity Q using a vector of geography and firm level controls $[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]$, which are listed in the appendix. Baseline productivity $\hat{Q}$ is used as the outcome, $\mathbf{Y}$ in the spatial lag specification B.4 and firm skill is the indirect effect variable $\mathbf{S}$. The observations are at the well-month level.

The first panel of figure A.6 shows analyses in which the spatial lag model uses only horizontally drilled wells and the two stage least squares analyses contains all horizontal wells drilled in those counties. The full results are reported in table form in table A.1. The

21. In the network model of equation B.4, there is a strong assumption about cohort effects. Only influence wells drilled nearby *at the time the reference well is drilled* can become part of the influence set. For the productivity estimates, I relax this to include investment activity at any time in the well's life. The studies are robust to this difference. Because natural resources tend to deplete over time, we would expect more drilling activity to have a negative impact on output so this specification is actually conservative.

figure illustrates the effect. In the weakest network strength counties, increased investment activity by other firms actually leads to less productive wells as compared to times of lower investment activity. As you move up the knowledge sharing quartiles, the impact of ex-firm investment activity on productivity increases monotonically. In the highest shared knowledge bucket, the impact of one extra well drilled by other firms in the county is associated with a higher productivity coefficient of 0.13.

To interpret this result, recall that the regression is modeled after the log version of Arp's production model,

$$log(O_{w,t}) = log(Q) + \beta log(Age_{w,t}) + \gamma \hat{I}_{g,t} \tag{1.6}$$

$$O_{w,t} = Q \times Age_{w,t}^{\beta} \times exp(\gamma \hat{I}_{g,t}) \tag{1.7}$$

For $\hat{\gamma} \approx 0.13$ and $\hat{I}_{g,t} = 1$, this roughly translates to $exp(0.13) = 1.13$ or % 13 increase in productivity accounting for the baseline productivity rate $\hat{Q}$. Recall that this productivity rate is modeled as a linear function of firm and geography level controls in this analysis. Figure A.7 illustrates this effect using simulated wells which are drilled in each network quartile.

1.3.1 Old Technology Tests

This paper is motivated by technology revolutions and proposes that shared knowledge may be an important role in understanding the joint effect of technology improvement and adoption on episodes of meaningful technological improvement. However, the baseline results reported in table A.1 do not necessarily distinguish the estimated effect from the productivity spillovers which have been studied in the literature. Additionally, it is difficult to identify conditions under which these spillovers may be more or less effective using the existing strategies employed in the literature thus far. One advantage of not relying on plausibly exogenous knowledge shocks such as [41] is that I can make comparisons between different technology types and analyze the effect of shared knowledge.

The bottom panel of figure A.6 replicates the main specification in equation 1.2 using the old technology, vertically drilled wells. The network effects in equation B.4 uses baseline productivity $\hat{Q}$ as the outcome of interest, $\mathbf{Y}$ and firm skill as the indirect effect variable $\mathbf{S}$. However, the analysis only uses vertically drilled wells to measure the network strength. The two stage least squares implementation of the main thought experiment also uses vertically drilled wells only. Strikingly, the increasing effect seen in the top panel with horizontally drilled wells is not replicated in the vertical wells. The coefficients are *decreasing* as you move along the network quartiles on the x-axis.

This result is intuitive but important. Spillover effects on productivity can be distinct from spillover effects on learning. For example supply chains can be impacted by the construction of a similar plant nearby which has a positive effect on productivity. The ability to hire more experienced workers from a nearby office is also another form of productivity spillover which can be useful regardless of the maturity of the technology involved. However, this test shows that productivity spillovers which result from *learning* through shared knowledge does depend on the technology. To the my knowledge, this is one of the first empirical studies which uses the same industry and the same output but different production technology to analyze this issue. The results show that the shared intangible capital mechanism proposed in this paper is distinct from the productivity spillovers studied in the literature. Different types of spillovers also differ in the circumstances under which they are more or less effective. Data and experience which can be used by other firms is useful in this new technology environment while other forms of productivity spillovers such as human capital transfer may be useful to old type technologies as well. The result also highlights central point of this paper, that this specific type of spillover is particularly important for the development of *new* technology.

1.3.2 Mechanism Tests

This section reports a set of additional results which shed light on the mechanism through which this shared knowledge functions. They should be seen as complements to the main productivity results discussed above. The results are shown in figures A.8 and A.9 as well as table A.2.

Data Availability. Figure A.8 replicates the main specification in equation 1.2. However, in place of the indirect effect variable $\mathbf{S}$, the spatial network model uses the data availability matrix $\mathbf{A}$. The matrix is described in detail in section 1.2.4 above. The set of variables considered along the columns of $\mathbf{A}$ contain information regarding the fracking program and additional well production details which are not uniformly available[22]. The matrix contains one when the variable in the column is available for the influence well drilled in that row and a zero otherwise.

While the results are noisier than the baseline productivity results, there is still a much larger effect in the highest quartile and the value is statistically significantly different from the estimated effect in the other three buckets. Importantly, the matrix does not contain the value of the variable, only an indicator if the variable is *available*. For example, one variable included in this exercise is the proppant level in lbs. Proppants are used in fracking fluid to hold fractures open after treatment.[23] This measure does not say that larger proppant measures are associated with better outcomes. Rather, the fact that the data is

22. A full list is available in the appendix

23. See Schlumberger's oilfield glossary for a formal definition

available for nearby wells is useful. While this data sharing may not be the only mechanism by which this shared intangible effect works, this result sheds light on data availability as one clear channel. The appendix shows analogous results where the standard deviation of these variables is used instead of the data indicator. There is a clear negative effect from increased investment in that specification. Taken together, these two sets of results show the importance of publicly available data in contributing to the creation of valuable shared intangible capital as firms make physical investments.

Technology & Costs. Figure A.9 and table A.2 take two different approaches to better understanding the mechanism. Figure A.9 seeks to understand how differences in the network measurements affect productivity. On the other hand, table A.2 seeks to understand different ways in which the productivity based network effects manifests in outcomes.

In the top panel of figure A.9, the analysis is the main two stage least squares specification in equation 1.2. However, the spatial network model uses the horizontal interval length as the outcome of interest $\mathbf{Y}$ with firm skill as the indirect effect variable $\mathbf{S}$. The horizontal perforated interval length measures the portion of the well that is turned and perforated to allow for fracking. It is an important technological aspect of the new, horizontally fracked wells and a dimension along which one would expect significant learning and growth.[24] The spatial network captures areas where firms are more likely to drill more complex wells when drilling near skilled firms. The main two stage least squares analysis then considers whether productivity effects from investment activity is stronger due to this technology based network effect. This is distinct from the productivity based network effect studied above.

24. The appendix contains a figure entitled "Average horizontal interval lengths over time" which illustrates this industry wide growth over time.

While the result is not monotonic, the effect is largest in the strongest network quartile. When technological complexity is impacted by nearby firms, increased investment activity is also associated with more productive wells. The effect is under 0.05 so the magnitude is lower than the 13% estimated from productivity based networks but the effect is still strong. Further, the effect is statistically distinct from the lower magnitudes in the other three shared knowledge quartiles.

This analysis raises a related question which is important to the entire empirical framework. Despite using indirect measures such as firm skill and data availability, the reader may still be concerned that the results are simply capturing clustering effects which have nothing to do with the creation of shared knowledge from physical capital investment. The bottom panel of figure A.9 addresses this issue. Again, the results illustrated are from the baseline specification in equation 1.2. However, the network model here sorts on $\gamma_1^{g,H}$ in the spatial network model B.4 instead of $\beta^{g,H}$. In other words, the network estimate here is based on the coefficient on lag interval lengths. It captures areas where complex wells are more likely to be drilled near other complex wells instead of skilled firms. This network model can be interpreted as a form of technology clustering. The second stage results from this analysis shows that productivity does not respond more strongly to investment activity in areas with technology clustering.

Finally, table A.2 shows results from the baseline specification 1.2 with a different second stage. The first stage again regresses $I_{g,t}$, investment by other firms in county g at time t excluding the firm that drilled well w on an indicator for the oil price regime. The Second stage regresses drill time, the time it takes between spudding and completing a well as the dependent variable. The explanatory variables include the set of geography and firm control variables $[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]$ as well as the instrumented $\hat{I}_{g,t}$. The observations are limited to one observation per well w and t is the month in which the well is drilled (spud date).

Drill times represent one of the larger costs of drilling wells. Additionally, there is significant skill and thus learning along the drill time dimension. In addition to simple productivity effects, this acts as a useful test of additional learning parameters where shared knowledge may be useful. The results show a strong negative effect in the strongest network quartile. Firms actually drill faster, and thus less costly wells, when there is increased investment activity in the area. This effect is stronger than any supply chain issues which one should expect as more firms vie for access to the same pool of drilling servicers. There is also a potential positive effect from pooled servicing resources which could be a confounding factor. However, the network effect is estimated using $\hat{Q}$ as the $\mathbf{Y}$ and firm skill as the indirect effect $\mathbf{S}$. There is no reason to expect this positive pooled servicing effect to be correlated with the shared knowledge estimates.

This result suggests that the productivity network estimated in the spatial lag model captures elements of shared knowledge which are useful not simply for productivity purposes. It may also suggest that in addition to the pure data availability channel shown in figure A.8, there may be other forms of experience sharing that is not easily identified in the data. While there is precedent in the literature for considering drill times as a shared learning dimension[25], it is encouraging that the result holds in this specific shared intangible capital framework. Further, this presents an interesting opportunity for further research. When does shared learning along one dimension of firm activity lead to improvements in other dimensions as well?

25. See [50]

1.3.3 Robustness Tests

The appendix reports a number of robustness tests which address potential issues or alternative explanations to the results above. I briefly summarize the specifications and the results here but refer the reader to the appendix of the paper for detailed descriptions of the specifications and tables of the results.

Standard Deviation Tests "Standard Deviation Tests from extra data sets 1 and 2" studies the effect of investment activity on productivity based on a network that considers the informativeness of the data from nearby wells. I take the variables whose *availability* were used to construct data availability matrix **A** used in the data availability studies above.[26] I then index regions based on the average standard deviation of those variables. This measure proxies for the usefulness or informativeness of the data used in the data availability test. I then run the same main specification as in 1.2 but across these data informativeness buckets instead.[27] The regions with higher estimated standard deviations in the well production data of other firms do not exhibit a higher relationship between investment and output. The results are noisily estimated in each quartile but there is no evidence for a trend across the region types. The lowest standard deviation regions show the largest magnitude effect between investment and output though it is also the noisiest estimate. Thus, even though figure A.8 showed that data availability created when firms invest is important for output, the informativeness of that same dataset has the inverse effect.

Distribution Tests Another concern is that there is a version of a size effect in the results. Perhaps the productivity results are capturing effects from large firms with significant

26. Detailed descriptions are available in the appendix.

27. it should be noted that instead of dividing the quartiles with zero as the median, I divided the distribution into quartiles for everything that is nonzero.

presence in one area. Table "Own Firm Investments Effects" shows the baseline specification 1.2 but with the firm's own investment levels in the first stage instead of the ex-firm cnty investment variables that have been used thus far. The results show no trend in the coefficient magnitudes across network buckets. However, the first and third buckets are larger and significant in magnitude relative to the highest shared knowledge bucket. The effect of firms making larger investments in a single county is not stronger in high shared knowledge areas.

On a related note, table "Distribution tests: Dispersed Firm Sample" approaches the distribution question from a different angle. The ideal thought experiment for the spillover effect is one where firms are atomistic and changes to ex-firm investment closely track county level differences. A potential confound here is that there are counties where a single firm makes a large difference or indeed where there are only two firms so changes from any one firm have a large effect in the ex-firm result. An important theoretical insight gained from introducing the knowledge spillover mechanism in this paper is that those distributional effects in levels should not be ignored. Areas with a single large firm who can afford to drill a large number of wells still contributes to aggregate knowledge from each well. Thus, technology improvement is a function of the levels, even if that is disproportionately weighted. Still, I explore how the main spillover effect is impacted by these distributional effects. I replicate the main specification for a subset of firms in each county. To create the subset, I remove any firm that represents over 50% of the investments in a county during that month. I also remove their investments from the ex-firm investment calculations. This subset of firms is more dispersed by construction. The results are consistent with the baseline estimate. Even amongst small, dispersed firms, when other small firms in the region increase investment activity, the effect on output is strongest in high spillover areas.

1.4 Theoretical Results: Mechanism Discussion

Various versions of knowledge sharing, technology imitation, and productivity spillovers have appeared in the literature. Here, I formalize the mechanism behind the shared intangible capital framework. The full model is analyzed in a theoretical companion paper but additional details are available in the model appendix to this paper. In this section, I discuss the intuition behind the theoretical results and point out theoretical contributions from this mechanism to the broader knowledge and spillover literature. These results also present some simple empirical predictions which will be tested in the investment and adoption section 1.5 below.

Consider the following profit function,

$$\pi = P_t \left[A_{old} k_{it} (1 - \gamma_{it}) + A_{new,it} k_{it} \gamma_{it} \right]$$

where P_t denotes an aggregate demand price which is set exogenously. Firm i has capital stock $k_{i,t}$ at time t. The capital stock is allocated between new and old type capital using proportion $\gamma_{i,t}$. For example, a firm with $k_{it} = 10$ and $\gamma_{i,t} = 0.5$ has five units of the old technology used in production and five units of the new. A_old, the productivity of the old capital is assumed to have reached maturation and is constant. The productivity of new technology is given by,

$$A_{new,it} = [(\eta K_t \Gamma_t)^{\alpha_1}]^{\mu_1} (h_{it} \tilde{h})^{\mu_2} \tag{1.8}$$

$$\Gamma = \int \gamma g(k, \gamma, h) d\gamma \tag{1.9}$$

$$K_t = \int k g(k, \gamma, h) dk \tag{1.10}$$

$$\tilde{h} = (\eta_2 k_{it} \gamma_{it})^{\alpha 2} \tag{1.11}$$

h_{it} captures accumulated firm level skill while $\tilde{h}$ captures own-firm effects because firms potentially learn from their own contemporaneous investments. In other words, h_{it} is similar to being a good student over all while $\tilde{h}$ captures skill in a particular class which is a function of how much class-specific experience $(k_{it}\gamma_{it})$ a firm gathers. Γ_t, K_t denotes industry wide aggregate technology sophistication and capital stock.

The full HJB is denoted in the model appendix to this paper. Here I only discuss the problem solved by the firms. Individual state variables include $ki, t, \gamma_{i,t}, h_{i,t}$. The firms also have to take into account the aggregate state variables, K_t, Γ_t. There are two control variables or choices that the firm has to make. Firms choose aggregate investment level $x_{i,t}$ taking their technology sophistication level $\gamma_{i,t}$ as given. Firms also choose a stopping time τ during which it can change its technology sophistication level γ_{it} to γ'_{it} subject to an adjustment cost θ_s.[28] Firms can move more than once and they are only allowed to become more sophisticated, not less.

Studying equation 1.8 makes it clear how the mechanism functions. As $K_t\Gamma_t$, aggregate new technology stock, increases the productivity of the new technology improves for everyone regardless of their existing skill level. Second, as $K_t\Gamma_t$ increases and therefore $A_{new,it}$ increases, the marginal contribution of new technology capital stock to firm value increases.[29] As this contribution to firm value from new technology capital stock increases, the optimal investment level also increases. Thus, as investment in new technology increases, the technology improves. As the technology improves, more firms have an incentive to adopt the new technology. This mechanism will be illustrated in the results below. There are a few observa-

28. θ_s is also described in detail in the model appendix. It is increasing in the magnitude of the change (for example moving from 0 to 0.5 is more costly than moving from 0 to 0.1) and decreasing in aggregate new type capital installed. Intuitively, if more firms are using the new technology, it becomes less costly for other firms to become more sophisticated since they can learn from a larger pool.

29. In HJB form, $\frac{\partial V}{\partial k\gamma} \uparrow$.

tions which are less obvious. First, the only mechanism for technology change is investment x_{it} and changing technology specialization rates, γ_{it}. There are no technology productivity shocks. Further, any shocks that impact investments could also affect technology. This is particularly relevant to the literature on business cycles and technological growth such as [30]. Finally, any physical investment which grows the capital stock k_{it} has an impact technology growth. Thus, the distribution of firms matters for reasons which differ from the literature on firm dynamics and firm life-cycles. The importance is not due to the fact that smaller firms tend to be more willing to take risks on R&D but rather that firms along the size, specialization, and skill distribution make different optimal investments. Those physical capital investment decisions affect the level of shared knowledge which is created.

The results below show aggregate transition dynamics for variables of interest. The model appendix details how these solutions are found but show transitions of the economy from a stationary equilibrium with low aggregate prices to the stationary equilibrium of the economy following a "MIT" shock to the demand price P_t. Figure A.11 shows the transition dynamics of Γ_{it}, aggregate technological sophistication, $A_{new,it}$, average new technology productivity, $avg(h)$, average accumulated firm skill, and $avg(x_n)$, average new technology investment levels respectively. The results show comparative statics for economies with high knowledge sharing, $\alpha_1 = 0.7$, mid knowledge sharing, $\alpha_1 = 0.25$, high own-learning, $\alpha_2 = 0.0.7$ and mid own-learning, $\alpha_2 = 0.5$.

First, aggregate adoption rates are much more rapid when there are significant benefits to making large investments at one time. This economy is analyzed in the learning section of the appendix but does not appear in the main empirical analyses. The main empirical tests only compare $\alpha_1 = 0$ to $\alpha_1 > 0$. However, the theoretical result is still enlightening. Intuitively, when no one internalizes the benefits of knowledge but rather relies on aggregate experience, the diffusion is much slower. This intuition is also made clear in figure A.13. The optimal target new technology adjustment rates across the distribution of firms is much faster and the magnitudes are higher when firms can internalize the knowledge benefits of investing. Additionally, the initial buildup of new technology stock is slow (so slow that it is hard to see in the plots) but eventually, there exists a “tipping point” where the optimal new technology adjustment targets begin to increase sufficiently such that the aggregate new technology ratio begins to increase steadily. Importantly, figure A.14 shows that the optimal investment levels x_{it} do not change much over the transition. Rather it is the technology sophistication rates which adjust over time.

Second, the plots of $A_{new,it}$ show that productivity is actually lower in the high knowledge sharing areas relative to the mid knowledge sharing areas until the tipping point occurs. Comparing firm skills h_{it} in the bottom left panel, there is a rapid acquisition of skill when learning is internalized. This process is much slower with the shared knowledge economies. However, when the tipping point occurs, it is possible for the economy to eventually result in firms who are, on average, more skilled than the own-learning economies. Finally, the bottom right panel of figure A.11 shows the aggregate investment levels in new technology over time.

Figure A.12 shows the aggregate new type capital stock installed over time for economies that begin with different randomly drawn distributions across the (k, γ, h) dimensions. I reserve discussions of the different distributions for the theoretical companion. However, the results are an important contribution of the theoretical framework and provide a useful starting point for future empirical work. The results suggest that the distribution of firms operating in the economy at the time a new discovery is made may be important to understanding inter-industry or even inter-country differences in technological growth. For some starting distributions, there may never be sufficient capital stock buildup to reach the tipping point. Additionally, from a policy perspective, notice that old type capital eventually dies out in both types of economies. However, the rate at which it dies differs because the main mechanism here is shifting *adoption rates* not *investment levels*. For policy applications in energy and climate, this result has implications for the rate at which older, dirtier technologies are transitioned out of the economy. Policies which discourage shared knowledge may actually be optimal even if it has a positive effect on technological growth.

Finally, figure A.14 shows optimal γ'_{it} for firms along the (k, γ) distribution at different points in the transition. In the high knowledge sharing economies, adjustment rates are extremely slow. Additionally, the bulk of the large adjustments are made by smaller firms regardless of their current specialization rate, γ_{it}. This is one useful empirical prediction which will be considered in the section below and reported in A.4.

1.5 Empirical Results: Adoption and Investment Results

The shared intangible capital mechanism just discussed in the mechanism section above is distinct from productivity spillovers studied in the literature. In addition to taking many different forms such as human capital or efficiency gains, those spillover effects tend to manifest as shocks which die down over time. This mechanism takes a particular form: experience which can be accessed by other firms either through data availability or observation. More importantly, the theoretical discussion makes an additional point. If there is valuable shared intangible capital which is accessible to all firms, it should impact the future physical capital investment decisions as well since this knowledge potentially makes capital more productive. This second portion of the feedback mechanism, the investment that is attributable to network effects is an important instrument which allows shared intangibles to drive technology cycles. Note that this type of knock-on effect through additional investments can exist in other productivity spillovers as well. However, it is difficult to study empirically because future investments in an area which receives a plausibly exogenous knowledge shock may be endogenous. The novel empirical strategy of first sorting counties based on a network estimate, then analyzing the effect of plausibly exogenous knowledge shocks is useful for addressing this problem. I am able to compare differences in investment response to plausibly exogenous investment decisions by other firms between areas where shared knowledge is more or less useful.

Table A.3 shows results from an empirical experiment which is similar to the baseline. Recall that in the baseline productivity results, I use oil prices which impact investment levels overall to study the effect of varying capital investment levels on productivity. However, lower oil prices impact all investment activity which is important to the identification design of the main experiment. Here, I use a different variable to instrument for changes in investment level. Leases granting firms the right but not the obligation to drill tend to have a primary term. If no well is drilled before that expiration, the leases expire and the firms lose the option. In [50], the author shows significant clustering of drilling activity around lease expiration times. In the first stage, $Exp_{g,t}$ counts the total number of leases which are reaching their primary expiration at month t. I use $Exp_{g,t}$ to instrument for investment activity by other firms in the county. The specification is,

$$I^i_{g,t} = \gamma^{ks_n} \hat{I}_t + \Gamma \left[\mathbf{X}_{g,t} \; \mathbf{X}_{i,t} \right]' + \epsilon_{w,t} \quad (\textit{second stage})$$

$$I_{g,t-1} = \eta + \beta^{ks_n} Exp_{g,t-1} + \Gamma \left[\mathbf{X}_{g,t} \; \mathbf{X}_{i,t} \right]' + \nu_{g,t} \quad (\textit{first stage})$$

The knowledge sharing quartiles ks_n are sorted using the spatial lag model in B.4 with $\hat{Q}$, baseline productivity, as the outcome variable $\mathbf{Y}$ and firm skill as the indirect effect variable $\mathbf{S}$. In the second stage investment *levels* by firm i in county g at month t is regressed on the instrumented investment activity variable, $\hat{I}_{g,t}$. The first stage considers investment activity lagged by a quarter since it is likely that investment activity would not react immediately to data that is contemporaneously available.

The results are fairly noisy but the coefficients are consistent with intuition. When other firms increase their investment activity, they are also involuntarily creating knowledge which is potentially valuable to other firms and may impact those firms' investment decisions. In the lowest shared knowledge quartile, increased investment activity by other firms is associated with lower investment levels. Firm investment responses to investment by other firms is largest in magnitude in the highest shared knowledge quartile but the effect is not significant. The result in the third highest quartile is positive and significant.

Tables A.4 and A.5 show simple investment and adoption decisions made by firms of different characteristics. The specification is given by,

$$I^i_{g,t} = \alpha + \beta log(KS_g) + \beta_2 Char_{i,t} + \gamma log(Ks_g) \times Char_{i,t} + \Gamma[\mathbf{X}_{g,t} + \mathbf{X}_{i,t}] + \epsilon_{i,g,t}$$

Two sets of results are reported. The first in table A.4 sets $I^i_{g,t}$ as firm-county level investments in the new technology (number of horizontally fracked wells drilled). The second one in table A.5 uses $I^i_{g,t} = \frac{\frac{I^i_{g,t,new}}{I^i g,t,total} - \frac{k^i_{g,t,new}}{k^i_{g,t,total}}}{\frac{k^i_{g,t,new}}{k^i_{g,t,total}}}$. The first term in the numerator is the firm's investment ratio of new to total investments in county g at time t. The second term is it's existing new technology production ratio based on the firm's installed capital stock at time t. This is taken as a proportion of the firm's existing technology ratio.

The results are generally consistent with the predictions from the theoretical model discussed above and shed light on additional trends in firm types which present interesting opportunities for future research. Unsurprisingly, larger firms are more likely to make larger level investments in new technology. This effect is strongest during time periods when oil prices are increasing as shown in columns one and four of the top panel of table A.4. This because they are making larger investments in general and effect is accentuated in high knowledge sharing areas as shown by the interaction terms in that same panel. More specialized firms are less likely to take make larger investments in high knowledge sharing areas. Intuitively, specialized firms primarily make new technology investments in all counties so the high knowledge sharing areas are not distinct. On the other hand, less specialized firms may find the shared intangible capital value more useful. Leverage ratios does not explain variations in investment levels.

Adoption rates, which measures the investments needed to change a firm's technology profile, are consistent with the adoption predictions from the theoretical discussion. During periods of increasing oil prices as in column one of table A.5, larger firms are less likely to make new technology adoption decisions as compared to smaller firms. This effect is even more negative in the interaction with the knowledge sharing estimate. In the third panel of table A.5, it is notable that while leverage does not explain investment levels, it does have an impact on adoption rates. More importantly, the only time period when it is significant is during the oil price drop of 2015 and 2016. More levered firms were more likely to make technology adoption decisions, especially in strong network areas.

While the size and specialization effects are consistent with the predictions from the basic model discussed in the mechanism section 1.4, the leverage results suggest fruitful areas for future research. Recall that an one contribution of this paper is to make the point that *any* investment activity can be potentially useful, regardless of the firm's skill. Generally, better understanding the interaction between financial concerns such as debt overhang and external financing cost on investment in new technology will be important. Not only would it shed light on how technology revolutions occur but it establishes a previously unexplored link between the financial sector and real economic growth precisely because the shared intangible capital mechanism studied in this paper ties physical capital investment with technology growth.

1.6 Empirical Results: Industry-wide Technology Effects

The baseline results in section 1.3 and adoption results in section 1.5 show average effects relating investment levels and productivity. In this section, I take a broader, industry-wide view. I relate the estimated shared intangible effect with technological and adoption growth trends that were occurring in the industry over the full time series. In summary, can knowledge sharing be responsible for driving growth over time?

Table A.6 shows regression results from

$$\Delta Y^{j}_{w,t} = \alpha + \gamma^{ks_n} \hat{I}^{j}_{t} + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (\textit{second stage}) \tag{1.12}$$

$$I^{j}_{t} = \eta + \beta^{ks_n} \mathbf{1}_{\{PriceRegime\}} + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (\textit{first stage}) \tag{1.13}$$

The investment variable here is ex-firm investment which measures total investments in the geography by all other firms. ΔY measures a firm's average estimated skill with its geography specific one. The dependent variable here is $\frac{(geography-firmskill)-(firm\ skill)}{firm\ skill}$, the proportional difference between the firm's geography-specific skill measurement and its nation-wide average, calculated quarterly. Firm skill is estimated with the same methodology as the skill vector used in the spatial lag model. For this sample, the estimate is geography specific so it is estimated using,

$$log(O_{w,t} = \beta log(Age_{w,t}) + \gamma_{i,g,q} \mathbf{1}_{\{Firm_i\ ,geo\}} \times \mathbf{1}_{\{q\}} + \epsilon_{w,t}$$

and $\hat{\gamma}_{i,g,q}$ is the geography specific skill vector. Firm performance in high knowledge sharing areas outperform their nation-wide estimate when investment activity by other firms is high. When investment by other firms in these areas increase by one well, firms perform around 20% (0.13 in the highest knowledge sharing bucket and 0.21 in the third) better relative to their investments in other areas. No such effect exists in the low spillover areas. Not only is the treatment effect on the treated significant in high spillover areas, these results show their impact on the industry as a whole given the improved relative performance between areas.

Figure A.15 shows the growth effects of knowledge sharing from a different perspective. The y-axis plots fixed effects coefficients from

$$log(O_{w,t}) = \beta log(Age_{w,t}) + \gamma \mathbf{1}_{\{ks^n, n=4\}} + \epsilon_{w,t} \quad |n = 1, 4, new\ type\ only$$

The $\hat{\gamma}$ coefficient shows the relative productivity of horizontal wells drilled in the highest knowledge sharing regions compared to those in the lower knowledge sharing buckets. The regressions are run every quarter and the results show Q-skill based knowledge sharing as well as production efficiency-skill based spillovers. In both panels, the high knowledge sharing regions suffer a decline in output relative to their lower knowledge sharing counterparts. This specification differs from the baseline knowledge sharing results in directly comparing knowledge sharing regions. In 2015 when the oil price decline begins, the highest knowledge sharing regions begin to decline relative to their low knowledge sharing counterparts. The growth effect is clearly shown in the Q-skill panel. This is a cross-sectional breakdown of the productivity time series shown in figure 1.1 which first motivated the paper. At least in part, the growth in the technology shown in the industry as a whole prior to 2015 is attributable to the parallel growth from high knowledge spillover areas. When oil prices decline in 2015, the slow down in growth industry-wide correspond with declines in high knowledge sharing areas due to investment declines.

Figure A.15 shows the industry-wide trends from 2010 to 2020. The top panel shows the oil prices, the second panel shows average county level investments, the third panel compares productivity of wells in the highest knowledge sharing areas with the lowest, and the fourth shows the relative productivity of horizontally drilled wells as compared to vertical ones.

The third panel specification is given by,

$$log(O_{w,t}) = \beta log(Age_{w,t}) + \gamma_t \mathbf{1}_{\{ks^n, n=4\}} + \epsilon_{w,t} \quad |n = 1, 4, \quad \textit{new type only}$$

Two vertical lines mark the time frame of the rapid oil price decline in all four panels. The time trend in panel three corresponds with the effect estimated in the baseline specification. The productivity advantage in the high knowledge sharing areas declines when investment drops in response to oil price drops. Over that same time period, industry-wide technology growth for investments made using the horizontally drilled wells slows down significantly.

1.7 Additional Mechanism Comparisons

In the appendix, I discuss some alternative mechanisms which may be important for technology change. In particular, learning by doing within a firm may be of interest. I design a separate way of measuring learning effectiveness and explore the possibility that learning may be complementary to knowledge sharing. Because these mechanisms are supplementary and the work is more exploratory, I only discuss it in the appendix but the reader may be interested to know the effect of this additional factor in this unique, knowledge sharing setting.

1.8 Concluding Remarks

Knowledge gleamed from making physical investments is often not excludable between firms. As a new technology diffuses throughout the economy, this results in industry-wide as opposed to firm-specific knowledge growth which in turn induces more adoption. This paper showed evidence for such a mechanism in a recent technology revolution, the fracking boom in American oil & gas. Using a novel network stress-testing methodology, I showed that high knowledge sharing areas are disproportionately impacted by plausibly exogenous investment drops. Additionally, firm investment and adoption decisions are shown to be well explained by the cross-sectional variation in this network strength.

This paper shows that contrary to productivity spillovers, this effect is particularly relevant to new technology. When the same experiment is replicated in the same industry but using an older, more mature technology, the network effect disappears. Further, I provide evidence that the effect propagates through specific, technical channels. First, network estimates computed using design complexity also exhibit disproportionate productivity effects when aggregate investment levels decline. The same exercise where networks are estimated based on technology clustering (for example, complex wells tend to follow other complex wells) do not show the same effect. This indicates that technology imitation through shared experience as opposed to technology clustering is driving the effects. Finally, network estimates conducted using nearby data *availability* are shown to also be disproportionately affected by investment levels. This lends further support to the proposal that this public disclosure of shared data is important for new technology.

The paper also examines the joint effect on growth and adoption. Namely, the knowledge sharing networks are shown to explain significant variation in the adoption and investment decisions made by firms. Using an institutional feature of oil & gas leases to instrument for changes in lagged investment levels, I show that the impact on current investment decisions are much higher in high network areas. Further, I divide firms into size and specialization quartiles and show that their investment decisions in different network buckets are consistent with the theoretical model. In particular, larger firms make larger level investments in new technology as expected but they are less likely to make large *adoption* decisions.

1.8.1 Further Research

The results in this paper serves as a starting point for many interesting projects. First, the underlying theoretical framework makes a novel contribution to the literature. Because the level of investments is what leads to aggregate knowledge growth, the distribution of firms in the economy is critical. Even though large firms are not the most likely to make large proportional investments in the risky new technology, they make larger level investments. The distribution of firms in the economy and any constraints that restricts the evolution of that distribution over time is shown to explain variation in both adoption rates and technology growth levels in the theoretical model. Thus, the mechanism suggests a different predictable factor to consider when analyzing why technology takes off in some industries and countries as opposed to others.

In the companion theoretical paper, I explore important aspects of this mechanism that are undeveloped in this baseline paper. In particular, firms make dynamic, forward-looking decisions based on their expectations as to how aggregate knowledge will evolve. In that paper, I explore the implications of heterogeneous beliefs on the aggregate technology adoption path. Other considerations to be included include questions regarding financial constraints. Because investment activity is the primary mechanism for technological change, this mechanism creates a tight link between the financial sector and economic growth. A particularly important application of this model is in climate change and the energy transition. Regulation, adoption and growth of new technology, as well as existing costs from climate change can all be incorporated into the model and studied. This is an extremely important question for understanding the role that financial considerations outside of climate risk will play in the fight against climate change.

All of these theoretical considerations should also be taken up in empirical work. Better understanding whether or not financial constraints play a role in technology growth through this channel is important. The study should also be extended beyond oil & gas to other sectors. In particular, sectors of growing importance such as renewable energy should be considered. Consumer finance and real estate sectors are also important areas that have been increasingly considering new technology in how they deliver services to customers.

Broadly, the impact of spillovers as formulated in this paper intersects with a large literature that is suggests many avenues for future research. In particular, if overall firm values are impacted by the relative technology ratios of a firm, is that reflected in asset prices? What is the impact of firm valuations over this technology dimension on its ability to raise capital and make future investments in new technology?

REFERENCES

[1] Andrew Abel and Janice Eberly. Optimal investment with costly reversibility. *Review of Economic Studies*, 63:581–593, 1996.

[2] Simona Abis and Laura Veldkamp. The changing economics of knowledge production. *Working Paper*, 2020.

[3] Daron Acemoglu. Directed technological change. *Review of Economic Studies*, 69(4):781–809, 2002.

[4] Daron Acemoglu, Philippe Aghion, Leonardo Bursztyn, and David Hemous. The environment and directed technical change. *American Economic Review*, 102(1):131–166, 2012.

[5] Daron Acemoglu, Ufuk Akcigit, Douglas Hanley, and William Kerr. Transition to clean technology. *Journal of Political Economy*, 124(1), 2016.

[6] Daron Acemoglu and Veronica Guerrieri. Capital deepening and nonbalanced economic growth. *Journal of Political Economy*, 116(3), 2008.

[7] Yves Achdou, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Ben Moll. Income and wealth distribution in macroeconomics: a continuous time approach. *Review of Economic Studies*, forthcoming.

[8] Philippe Aghion and Peter Howitt. A model of growth through creative destruction. *Econometrica*, 60(2):323–351, 1992.

[9] Philippe Aghion and Peter Howitt. Endogenous growth theory. *MIT University Press*, 1997.

[10] Ufuk Akcigit and Sina Ates. Ten facts on declining business dynamism and lessons from endogenous growth theory. *American Economic Journal: Macroeconomics*, 13(1):257–298, 2021.

[11] Ufuk Akcigit, Emin Dinlersoz, Jeremy Greenwood, and Veronika Penciakova. Synergizing ventures. *NBER Working Paper*, (26196), 2020.

[12] Ufuk Akcigit, John Grisby, Tom Nicholas, and Stefanie Stantcheva. Taxation and innovation in the 20th century. *Quarterly Journal of Economics*, forthcoming, 2021.

[13] Ufuk Akcigit, Douglas Hanley, and Nicolas Serrano-Velarde. Back to basics: basic research spillovers, innovation policy and growth. *Review of Economic Studies*, 88(1):1–43, 2020.

[14] Ufuk Akcigit and William Kerr. Growth through heterogeneous innovations. *Journal of Political Economy*, 126(4):1374–1443, 2018.

[15] Luc Anselin. *Spatial Econometrics*, volume Chapter 14. 2001.

[16] Diego Anzoategui, Diego Comin, Mark Gertler, and Joseba Martinez. Endogenous technology adoption and r&d as the sources of business cycle persistence. *The American Economic Journal: Macroeconomics*, 11(3):67–110, 2019.

[17] David Argente, Douglas Hanley, and Sara Moreira. Patents to products: Product innovation and firm dynamics. *CEPR Discussion Paper NO. DP14692*, 2020(4), 2020.

[18] David Audretsch. Agglomeration and the location of innovative activity. *Oxford Review of Economic Policy*, 14(2), 1998.

[19] Tor-Erik Bakke and Tiantian Gu. Diversification and cash dynamics. *Journal of Financial Economics*, 123(3):580–601, 2017.

[20] Abhijit Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.

[21] Santiago Bazdresch. The role of non-convex costs in firms' investment and financial dynamics. *Journal of Economic Dynamics & Control*, 37:929–950, 2013.

[22] Nicholas Bloom. The impact of uncertainty shocks. *Econometrica*, 77(3):623–685, 2009.

[23] Nicholas Bloom, Tarek Hassan, Aakash Kalyani, Josh Lerner, and Ahmed Tahoun. The diffusion of disruptive technologies. *NBER working paper*, (28999), 2021.

[24] Nicholas Bloom, Charles Jones, John Van Reenen, and Michael Webb. Are ideas getting harder to find? *American Economic Review*, 110(4):1104–1144, 2020.

[25] Nick Bloom, Mark Shankerman, and John Van Reenen. Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393, 2013.

[26] Patrick Bolton, Hui Chen, and Neng Wang. A unified theory of tobin's q, corporate investment, financing, and risk management. *Journal of Finance*, 66(5):1545–1578, 2011.

[27] Patrick Bolton, Hui Chen, and Neng Wang. Market timing, investment, and risk management. *Journal of Financial Economics*, 109(1):40–62, 2013.

[28] Wesley Cohen and Steven Klepper. Firm size and the nature of innovation within industries: The case of process and product r&d. *The Review of Economics and Statistics*, 78(2):232–243, 1996.

[29] Diego Comin. The intensive margin of technology growth. *Handbook of Economic Growth*, 2, 2014.

[30] Diego Comin and Mark Gertler. Medium-term business cycles. *The American Economic Review*, 96(3):523–551, 2006.

[31] Thomas Covert. Experiential and social learning in firms: the case of hydraulic fracturing in the bakken shale. *Job Market Paper*, 2015.

[32] Paul Decaire, Erik Gilje, and Jerome Taillard. Real option exercise: empirical evidence. *Review of Financial Studies*, 33(7):3250–3306, 2019.

[33] Paul Decaire and Michael Wittry.

[34] Janice Eberly and Neng Wang. Capital reallocation and growth. *American Economic Review*, 99(2):560–566, 2009.

[35] Timothy Erickson and Toni Whited. Measurement error and the relationship between investment and q. *Journal of Political Economy*, 108(5):1027–1056, 2000.

[36] Maryam Farboodi, Roxana Mihet, and Thomas Philippon. Big data and firm dynamics. *CEPR Discussion Paper*, (DP13489), 2019.

[37] Maryam Farboodi and Laura Veldkamp. A growth model of the data economy. *NBER Working Papers*, (28427), 2021.

[38] Lucia Foster, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. Innovation, productivity dispersion, and productivity growth. *NBER Working paper*, (24420), 2018.

[39] Lucia Foster, John Haltiwanger, and CJ Krizan. Aggregate productivity growth: lessons from macroeconomic evidence. *New Developments in Productivity Analysis*, University of Chicago Press, 2001.

[40] Erik Gilje and Jerome Taillard. Do private firms invest differently than public firms? taking cues from teh natural gas industry. *Journal of finance*, 2016.

[41] Xavier Giroud, Simone Lenzu, = Quinn Maingi, and Holger M. Mueller. Propagation and amplification of local productivity spillovers. *NBER working paper*, (29084), 2021.

[42] Zvi Griliches. Hybrid corn: An exploration in the economics of technological change. *Econometrica*, 25(4):501–522, 1957.

[43] Bronwyn Hall and Beethika Khan. Adoption of new technology. *New Economy Handbook*, 2003.

[44] Christopher A. Hennessy and Amnon Levy ad Toni M. Whited. Testing q theory with financing frictions. *Journal of Financial Economics*, 83(3):691–717, 2007.

[45] Fumio Hyashi. Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica*, 50(1):213–224, 1982.

[46] Charles Jones. The past and future of economic growth: A semi-endogenous perspective. *NBER Working Paper*, 2021.

[47] Charles I. Jones. Growth and ideas. *Handbook of Economic Growth*, 1B, 2005.

[48] Boyan Jovanovic and Saul Lach. Entry, exit, and diffusion with learning by doing. *American Economic Review*, 79(4):690–699, 1989.

[49] Boyan Jovanovic and Yaw Nyarko. Learning by doing and the choice of technology. *Econometrica*, 64(6):1299–1310, 1996.

[50] Ryan Kellogg. Learning by drilling: Interfirm learning and relationship persistence in the texas oilpatch. *The Quarterly Journal of Economics*, 126(4):1961–2004, 2011.

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

[52] Steven Klepper. Entry, exit, growth, and innovation over the product life cycle. *American Economic Review*, 86(3):562–583, 1996.

[53] Patrick Kline and Enrico Moretti. Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. *Quarterly Journal of Economics*, 129(1):275–331, 2014.

[54] Josh Lerner and Ramana Nanda. Venture capital's role in financing innovation: What we know and how much we still need to learn. *Journal of Economic Perspectives*, 34(3):237–261, 2020.

[55] Josh Lerner, Morten Sorensen, and Per Stromberg. Private equity and long-run investment: The case of innovation. *The Journal of Finance*, 66(2):445–477, 2011.

[56] Brian Lucking, Nicholas Bloom, and John Van Reenen. Have rd spillovers changed? *Fiscal Studies*, 40(4):561–590, 2019.

[57] Adrien Matray. The local innovation spillovers of listed firms. *Journal of Financial Econometrics*, 141(2):395–412, 2021.

[58] Prachi Mishra, Nagpurnanand Prabhala, and Raghuram Rajan. The relationship dilemma: Why do banks differ in the pace at which they adopt new technology? *Working Paper*, 2021.

[59] William D. Nordhaus. The perils of the learning model for modeling endogenous technological change. *The Energy Journal*, 35(1):1–13, 2014.

[60] Jesse Perla and Christopher Tonetti. Equilibrium imitation and growth. *Journal of Political Economy*, 122(1), 2014.

[61] David Popp, Richard Newell, and Adam Jaffe. Energy, the environment, and technological change. *Handbook of the Economics of Innovation*, 2(chapter 21):873–937, 2010.

[62] Paul Romer. Endogenous technological change. *Journal of Political Economy*, 98(5):571–102, 1990.

[63] Nancy Stokey. Technology diffusion. *working paper*, 2020.

[64] Peter Thompson. How much did the liberty shipbuilders learn? new evience for an old case study. *The Journal of Political Economy*, 109(1):103–137, 2001.

[65] Barriers to Creative Destruction: Large Firms and non-productive strategies. Capital deepening and nonbalanced economic growth. *Federal Reserve bank of Atlanta, working paper series*, 2021.

[66] H. Peyton Young. Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American Economic Review*, 99(5):1899–1924, 2009.

APPENDIX A

FIGURES & TABLES

A.1 Preliminaries

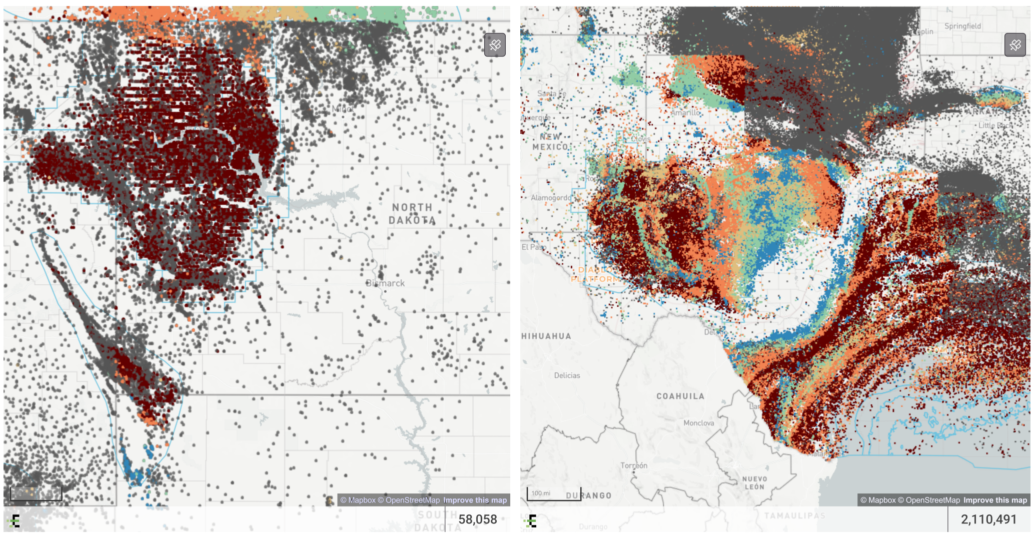

Figure A.1: Potential sources of geography heterogeneity in knowledge sharing propensity

The figure is pulled from Enverus analytics. It depicts the surface location of wells drilled in the Bakken shale of North Dakota and the Barnett and surrounding shale in Texas. The colors depict the depths of the wells drilled. There are significant geographic differences in both the variance of depths as well as the spacing of wells across the two regions. North Dakota in the top panel features well-lined wells with homogeneous depths. Texas in the bottom panel features more differences in depths as well as the surface locations of wells.

Note that this is not meant to suggest that one region is easier to drill or more productive. Rather, areas differ in the usefulness of other firms nearby. In North Dakota, it is likely more predictable on average. However, in Texas, there will be more information which can be used by other firms.

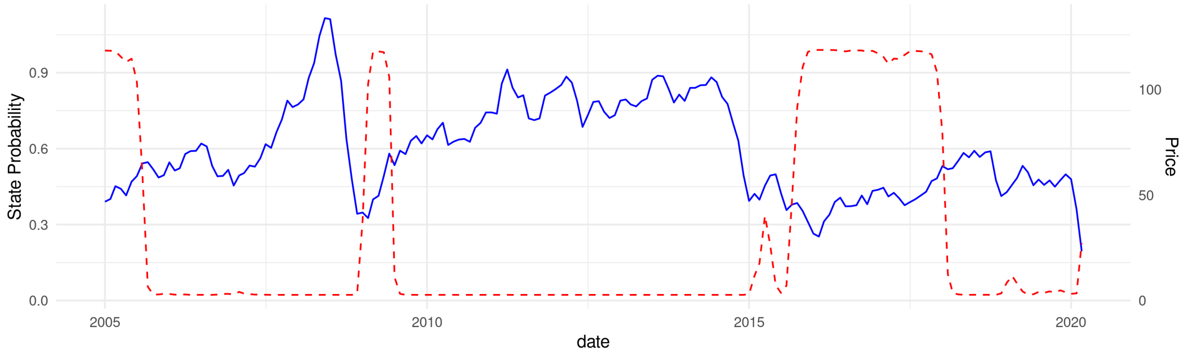

Figure A.2: Long term oil price trends with Markov Switching

The figure shows the WTI oil spot price in blue. Its corresponding y-axis is on the right. In red are the estimated probabilities from fitting a two stage Markov switching process to the data. Two large structural breaks can be seen. Between 2010-2015, the model estimates a low consistently low probability of being in the low state. The period from 2015 to 2018 shows consistently high probability of being in the low state. The two regimes depicted will be used to instrument for investment levels over those two, long time periods.

A.2 Empirical Framework

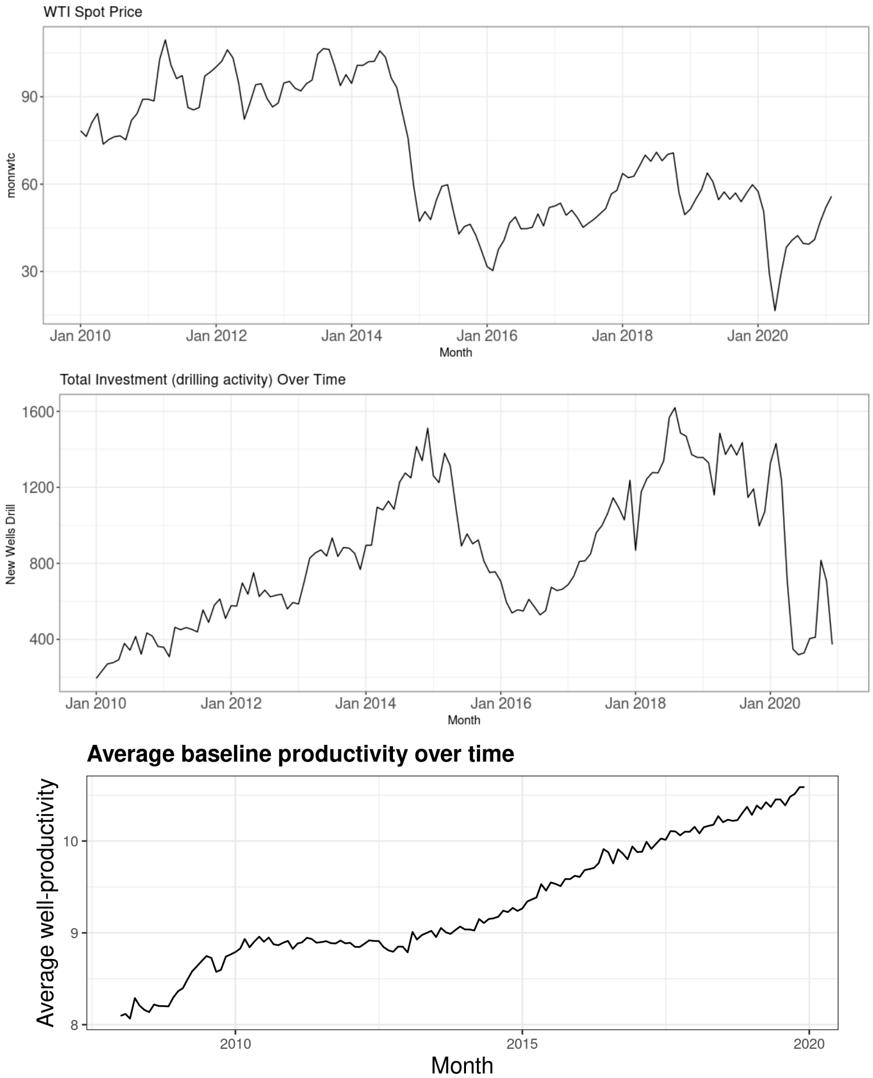

Figure A.3: Time series of oil prices, investment, and baseline production rates

The three panels show monthly WTI oil prices, total drilling activity, and the average estimated baseline production rate, $\hat{Q}$ over the time period in question. Drilling activity comprises all wells, vertical and horizontal. The data is estimated from the same dataset used in the empirical analysis of this paper.

The plots support the use of the oil price as an instrument. Over the period of oil price decline, the second panel shows a corresponding decline aggregate drilling activity. The empirical strategy treats this exogenous decline in investment as the natural experiment in a diff-diff framework. There are two threats to identification. The first is that this aggregate decline depicted in the second panel is drastically different between treatment and control groups. This will be discussed in the body of the paper. The second threat is that firms drill less productive wells when oil prices decline because they are less costly and not because of the impact on knowledge spillovers. The third panel shows increasing average production rates even through the oil price decline.

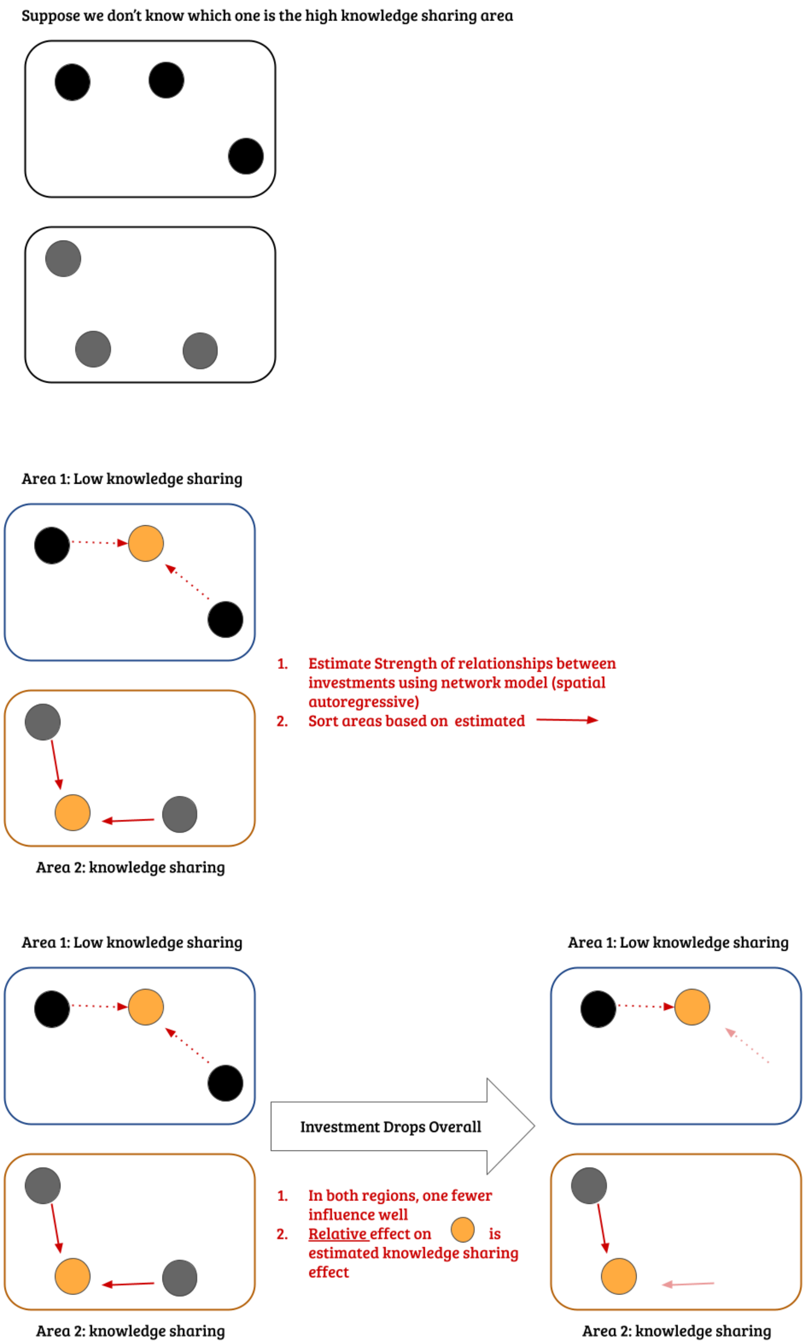

Figure A.4: Thought Experiment behind the primary empirical specification

The picture is a graphical representation of the thought experiment underpinning the main empirical specification. Despite the motivation for cross-sectional differences in network strengths, the econometrician does not know which is the higher knowledge sharing area. The first task is to estimate that network strength. Then, when investments decline, a node is essentially detracted from each network. The area with the stronger network should be disproportionately impacted by that decline.

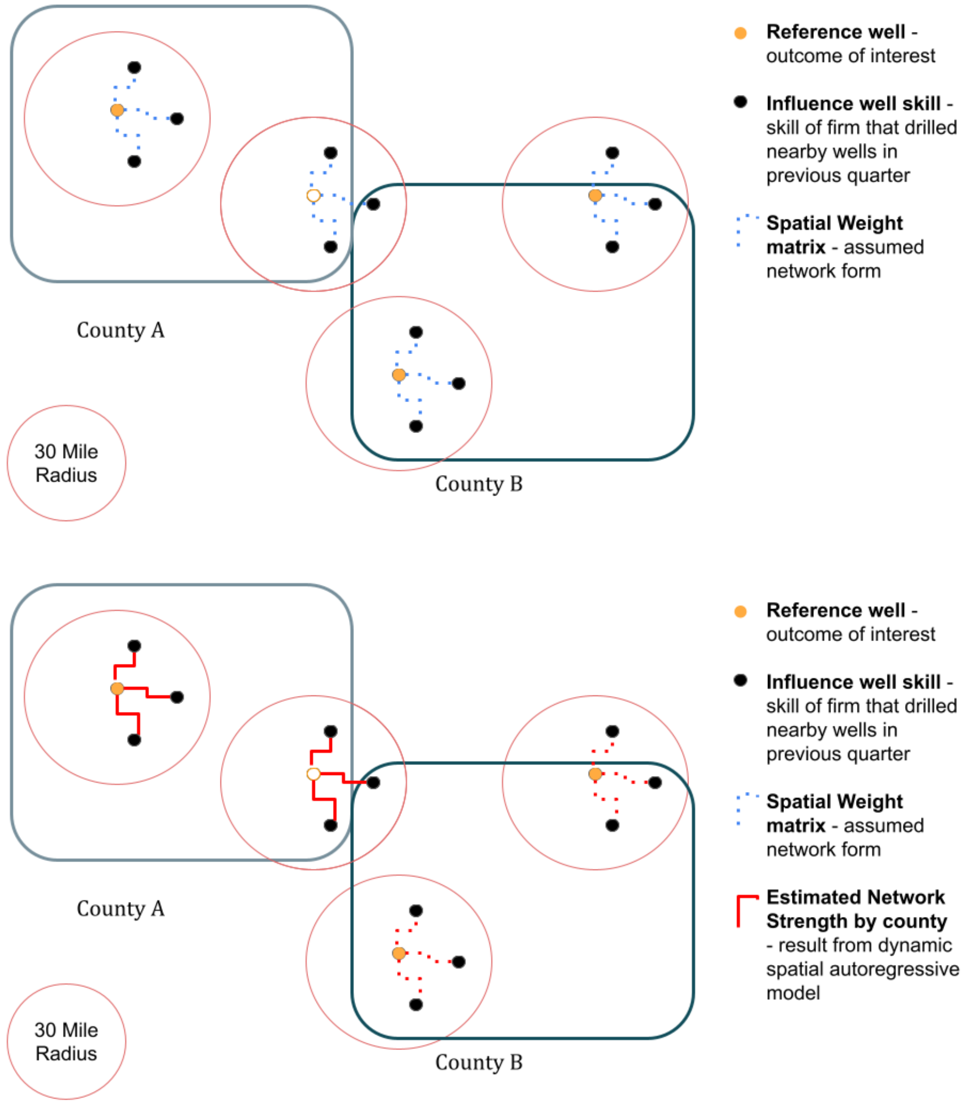

Figure A.5: Spatial lag model set-up with effect sizes

The picture is a graphical representation of how the spatial lag model is implemented to measure knowledge sharing. The details are available in the section on the spatial lag model.

A.3 Results

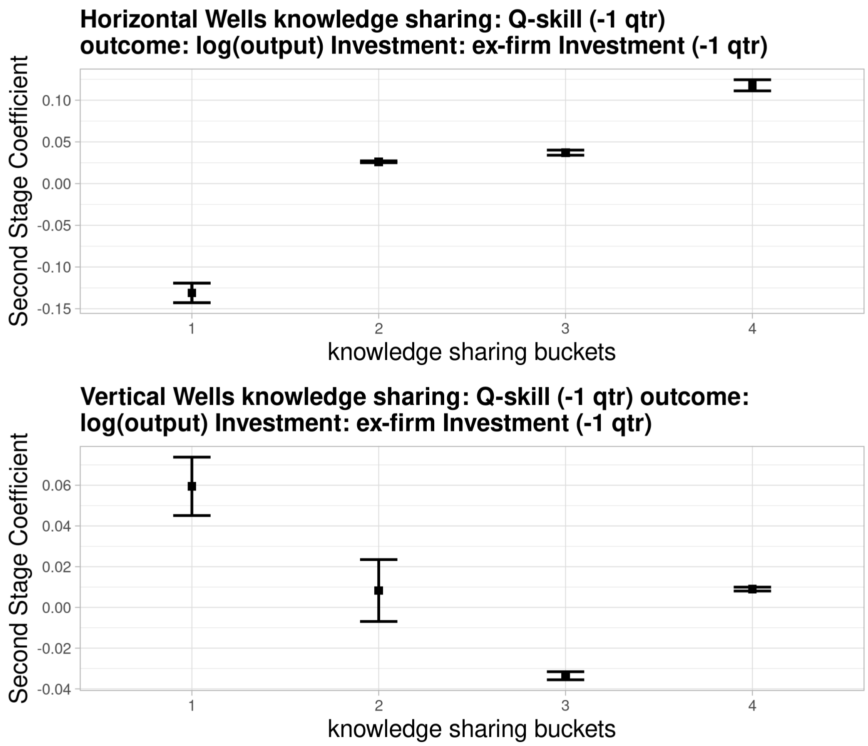

Figure A.6: Baseline Productivity Effect: New vs Old Type Technology

The figure plots second stage coefficients from the main empirical specification,

$$log(O_{w,t}) = log(Age_{w,t}) + \gamma^{ks_n} \hat{I}_t + \Gamma\,[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]' + \epsilon_{w,t} \quad (second\ stage)$$
$$I_{g,t} = \eta + \beta^{ks_n} \mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma\,[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]' + \nu_{g,t} \quad (first\ stage)$$

The first stage regresses I_t, investment by other firms in county g at time t excluding the firm that drilled well w on an indicator for the oil price regime. The Second stage starts from the log version of the classic Arp's production function for oil wells

$$O_{w,t} = QAge^{\beta}$$
$$log(O_{w,t}) = log(Q) + \beta log(Age)$$

For the regression analysis, I add the instrumented investment level $\hat{I}_{g,t}$. I model baseline productivity Q using a vector of geography and firm level controls $[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]$, which are listed in the appendix. The network strength from the spatial lag model is used to sort counties into quartiles which appear on the x-axis. Baseline productivity $\hat{Q}$ is used as the outcome ,$\mathbf{Y}$ in spatial lag specification B.4 and firm skill is the indirect effect variable $\mathbf{S}$.

The network effect is estimated once using all horizontal wells in a random selection of counties and once using the old vertical well types. The top panel shows results using horizontal wells in specification 1.2 with the horizontal network effect. The second panel shows the analysis using vertical wells and vertical network effects. The top panel shows a clear increasing effect from the creation of shared knowledge. By contrast, the old type technology shows a decreasing effect from investment. The impact of shared knowledge affects the productivity rate in new technology but the same effect does not hold in already developed technology.

	(1) Ex firm investments	(2) Knowledge Sharing 1	(3) Ex firm investments	(4) Knowledge Sharing 2	(5) Ex firm investments	(6) Knowledge Sharing 3	(7) Ex firm investments	(8) Knowledge Sharing 4
$\mathbf{1}_{\{High\ Price\ Regime\}}$	-1.16***		3.86***		1.37***		0.79***	
	(0.032)		(0.045)		(0.040)		(0.031)	
log(Age)	-0.080***	-0.36***	-0.61***	-0.13***	-0.74***	-0.33***	0.59***	-0.53***
	(0.015)	(0.0038)	(0.022)	(0.0034)	(0.024)	(0.0054)	(0.015)	(0.014)
Ex firm investments		-0.19***		0.072***		0.096***		0.46***
		(0.0067)		(0.0017)		(0.0054)		(0.020)
Constant	-18.4***	-12.7***	-50.9***	-2.32***	-38.3***	-9.88***	14.9***	-14.1***
	(0.50)	(0.18)	(0.93)	(0.14)	(0.84)	(0.22)	(0.52)	(0.42)
N	552279	552279	276275	276275	236027	236027	293589	293589
Underidentification		1331.0		6606.8		1148.5		635.4
Weak identification		1306.0		7238.8		1171.7		647.6

Marginal effects; Standard errors in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.1: Baseline Productivity Effect: Horizontal Wells

The table shows the full results for the analysis shown in figure A.6 above. The empirical specification is,

$$log(O_{w,t}) = log(Age_{w,t}) + \gamma^{ks_n} \hat{I}_t + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage)$$
$$I_{g,t} = \eta + \beta^{ks_n} \mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage)$$

The first stage regresses I_t, investment by other firms in county g at time t excluding the firm that drilled well w on an indicator for the oil price regime. The Second stage starts from the log version of the classic Arp's production function for oil wells

$$O_{w,t} = QAge^{\beta}$$
$$log(O_{w,t}) = log(Q) + \beta log(Age)$$

For the regression analysis, I add the instrumented investment level $\hat{I}_{g,t}$. I model baseline productivity Q using a vector of geography and firm level controls $[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]$, which are listed in the appendix. The network strength from the spatial lag model is used to sort counties into quartiles which appear on the x-axis. Baseline productivity $\hat{Q}$ is used as the outcome ,$\mathbf{Y}$ in spatial lag specification B.4 and firm skill is the indirect effect variable $\mathbf{S}$.

The first row shows the first stage results while the third role shows the second stage results. Notably, there is limited correlation between the rate of investment decline between oil price regimes and the estimated second stage effect. While investment levels increase in the low oil price period, the effect on productivity is negative. Amongst the other three network quartiles, Investment declines in response to the lower oil prices is the *lowest* in quartile 4 yet the second stage effect on the productivity of remaining wells is the largest in magnitude.

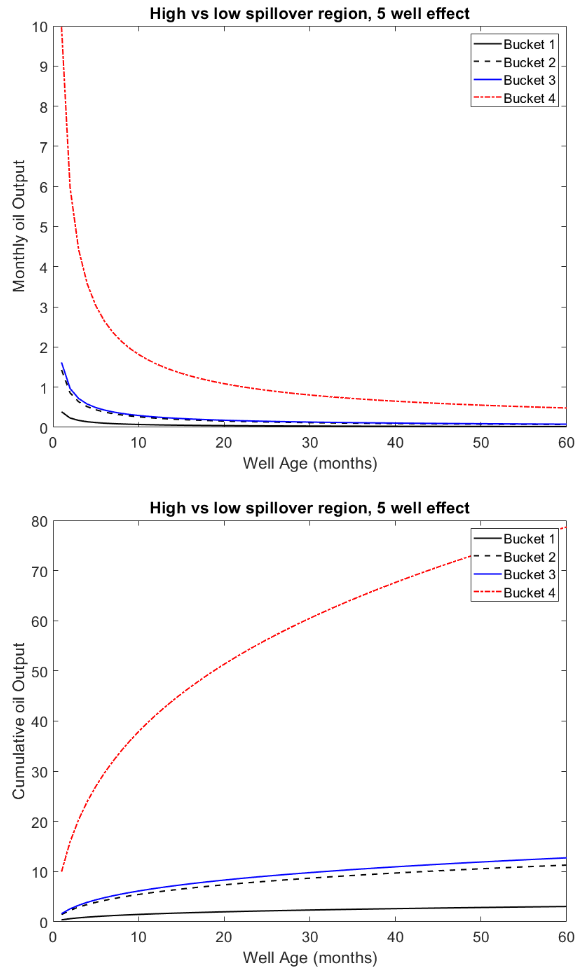

Figure A.7: Illustration of the Economic Magnitude from Baseline Regressions

The top panel illustrates the simulated economic impact from incorporating the knowledge spillover effect. Recall that the Arp's production model is,

$$O_{w,t} = Q_w Age_{w,t}^{\beta}$$
$$log(O_{w,t}) = \hat{\alpha} + \hat{\beta} log(Age_{w,t}) + \epsilon_{w,t}$$

Each line shows the same baseline productivity rate Q_w for a hypothetical well drilled in a county from each of the four network bucket. All wells share the same $Q = 0$[a] and $\beta = -0.6$. I then plot,

$$O_{w,t} = exp(\hat{Q} + \hat{\gamma}^{ks_n} \times I_t + \beta * log(Age_{w,t}))$$

with $\hat{\gamma}^{ks_n}$ corresponding to second stage results shown in table A.1 and $I_t = 5$.

The second panel shows the cumulative production values.

---

a. Note that this is for comparison only. The actual baseline Q would be much higher to match actual production numbers.

A.4 Mechanism Plots

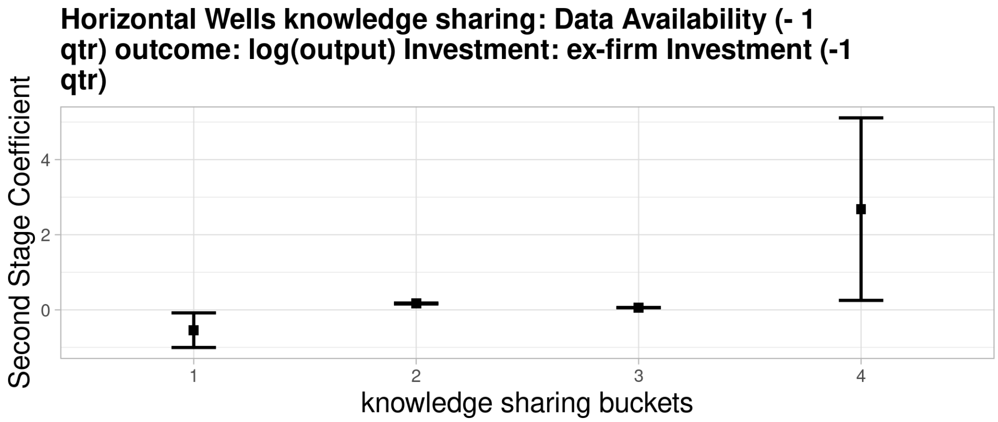

Figure A.8: Data Availability Productivity Effect

The figure plots second stage coefficients from the main empirical specification,

$$\begin{aligned} log(O_{w,t}) &= log(Age_{w,t}) + \gamma^{ks_n}\hat{I}_t + \Gamma\,[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]' + \epsilon_{w,t} \quad (second\ stage) \\ I_{g,t} &= \eta + \beta^{ks_n}\mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma\,[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]' + \nu_{g,t} \quad (first\ stage) \end{aligned}$$

The first stage regresses I_t, investment by other firms in county g at time t excluding the firm that drilled well w on an indicator for the oil price regime. The Second stage starts from the log version of the classic Arp's production function for oil wells

$$\begin{aligned} O_{w,t} &= QAge^{\beta} \\ log(O_{w,t}) &= log(Q) + \beta log(Age) \end{aligned}$$

For the regression analysis, I add the instrumented investment level $\hat{I}_{g,t}$. I model baseline productivity Q using a vector of geography and firm level controls $[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]$, which are listed in the appendix. The network strength from the spatial lag model is used to sort counties into quartiles which appear on the x-axis. Baseline productivity $\hat{Q}$ is used as the outcome ,$\mathbf{Y}$ in spatial lag specification B.4 and the indirect effect variable is data availability matrix $\mathbf{A}$ which is described in detail in section **??**. The set of variables considered along the columns of $\mathbf{A}$ contain information regarding the fracking program and additional well production details which are not uniformly available. The matrix contains one when the variable in the column is available for the well drilled in that row and a zero otherwise.

While the results are noisier, there is still a much larger effect in the highest quartile and the value is statistically significantly different from the estimated effect in the other three buckets. Importantly, the matrix does not contain the value of the variable, only an indicator if the variable is *available*. While this data sharing may not be the only mechanism by which this shared intangible effect works, this result sheds light on data availability as one clear channel. The appendix shows analogous results where the standard deviation of these variables is used instead of the data indicator. There is a clear negative effect from increased investment in that specification. Taken together, these two sets of results show the importance of publicly available data in contributing to the creation of valuable shared intangible capital as firms make physical investments.

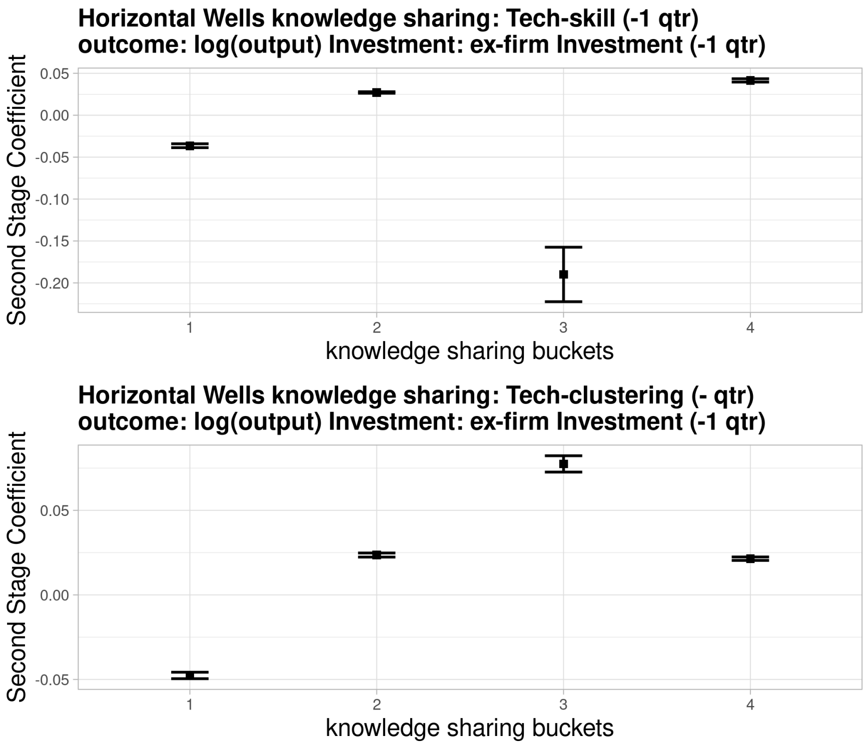

Figure A.9: Technology Indicator (horizontal interval length) Results

The two figures plot second stage coefficients from the main empirical specification,

$$log(O_{w,t}) = log(Age_{w,t}) + \gamma^{ks_n}\hat{I}_t + \Gamma\left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage)$$
$$I_{g,t} = \eta + \beta^{ks_n}\mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma\left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage)$$

The first stage regresses I_t, investment by other firms in county g at time t excluding the firm that drilled well w on an indicator for the oil price regime. The Second stage starts from the log version of the classic Arp's production function for oil wells

$$O_{w,t} = QAge^{\beta} \rightarrow log(O_{w,t}) = log(Q) + \beta log(Age)$$

For the regression analysis, I add the instrumented investment level $\hat{I}_{g,t}$. I model baseline productivity Q using a vector of geography and firm level controls $[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]$, which are listed in the appendix. The network strength from the spatial lag model is used to sort counties into quartiles which appear on the x-axis.

The two panels show different network effects. Horizontal perforated interval length measures the portion of the well that is turned and perforated to allow for fracking. It is an important technological aspect of the new, horizontally fracked wells. In both panels, interval length is used as the outcome ,$\mathbf{Y}$ in spatial lag specification B.4. In the top panel, firm skill is the indirect effect variable $\mathbf{S}$. In the second panel, counties are sorted based on $\gamma_1^{g,H}$, the coefficient on the interval lengths of the wells that were drilled in the preceding quarter. The second panel serves as a test of simple technology clustering as compared to a shared knowledge effect. The result in the top panel is not monotonic but the effect is still largest and significant in the fourth quartile of network strength. The same does not hold for technology clustering in the second panel.

	(1) Ex firm investments	(2) Knowledge Sharing 1	(3) Ex firm investments	(4) Knowledge Sharing 2	(5) Ex firm investments	(6) Knowledge Sharing 3	(7) Ex firm investments	(8) Knowledge Sharing 4
$\mathbf{1}_{\{High\ Price\ Regime\}}$	-8.68***		11.3***		2.82***		1.92***	
	(0.53)		(0.67)		(0.53)		(0.58)	
Ex firm investments (-1 qtr)		0.72		-0.41		-2.66		-11.0**
		(0.41)		(0.44)		(1.76)		(3.85)
Constant	-98.5***	82.5	-144.5***	208.5	-98.2***	-380.2*	65.2***	1033.3***
	(7.50)	(168.6)	(10.2)	(114.0)	(10.2)	(164.2)	(8.81)	(276.4)
N	14891	14891	9992	9992	8683	8683	6807	6807
Underidentification		275.1		267.8		27.9		10.8
Weak identification		268.5		284.0		28.2		10.9

Table A.2: Cost Indicator (Drill time) Results

The table reports from the main empirical specification,

$$log(O_{w,t}) = log(Age_{w,t}) + \gamma^{ks_n} \hat{I}_t + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage)$$

$$I_{g,t} = \eta + \beta^{ks_n} \mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage)$$

The first stage regresses $I_{g,t}$, investment by other firms in county g at time t excluding the firm that drilled well w on an indicator for the oil price regime. The Second stage regresses drill time, the time it takes between spudding and completing a well as the dependent variable. The explanatory variables include the set of geography and firm control variables $[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]$ as well as the instrumented $\hat{I}_{g,t}$. The observations are limited to one observation per well w and t is the month in which the well is drilled (spud date).

Drill times represent one of the larger costs of drilling wells. Additionally, there is significant skill and thus learning along the drill time dimension. In addition to simple productivity effects, this acts as a useful test of additional learning parameters where shared knowledge may be useful.

The results show a strong negative effect in the strongest network quartile. Firms actually drill faster, and thus less costly wells, when there is increased investment activity in the area. This effect is stronger than any supply chain issues which one should expect as more firms vie for access to the same pool of drilling servicers. There is also a potential positive effect from pooled servicing resources which could be a confounding factor. However, the network effect is estimated using $\hat{Q}$ as the $\mathbf{Y}$ and firm skill as the indirect effect $\mathbf{S}$. There is no reason to expect this positive pooled servicing effect to be correlated with the shared knowledge estimates.

A.5 Theoretical Results

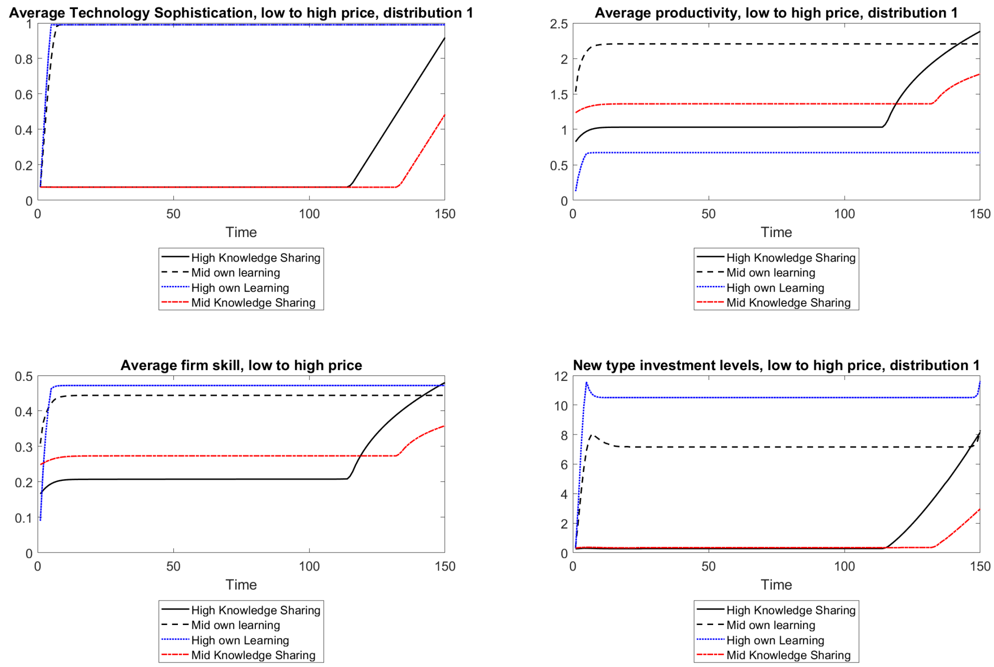

Figure A.10: Theoretical Solutions: Aggregate Dynamics

The figure depicts aggregate transition dynamics from economies with different, immutable knowledge sharing attributes. For production function,

$$
\begin{aligned}
\pi &= P_t \left[A_{old} k_{it}(1-\gamma_{it}) + A_{new,it} k_{it} \gamma_{it}\right] \\
A_{new,it} &= [(\eta K_t \Gamma_t)^{\alpha_1}]^{\mu_1} (h_{it}\tilde{h})^{\mu_2} \\
\Gamma &= \int \gamma g(k,\gamma,h) d\gamma \\
K_t &= \int k g(k,\gamma,h) dk \\
\tilde{h} &= (\eta_2 k_{it} \gamma_{it})^{\alpha 2}
\end{aligned}
$$

High knowledge sharing is given by $\alpha_1 = 0.5, \alpha_2 = 0$, mid knowledge sharing is $\alpha_1 = 0.25, \alpha_2 = 0$, High own learning is $\alpha_1 = 0, \alpha_2 = 0.7$ and mid own learning is $\alpha_1 = 0, \alpha_2 = 0.5$.

The two panels show $\Gamma_t, A_{new,it}, h_{it}, x_n$ respectively. The transitions begin with low demand price P_t and is subject to an "MIT" shock that sets P_t to a higher level. The distribution used to solve for the stationary distribution prior to the shock is randomly drawn with the limitation that there is only mass in areas with low γ. In other words, some firms attempt the new technology but there is limited adoption at first. This distribution is allowed to evolve accord to the kolmogorov forward equation throughout the transition.

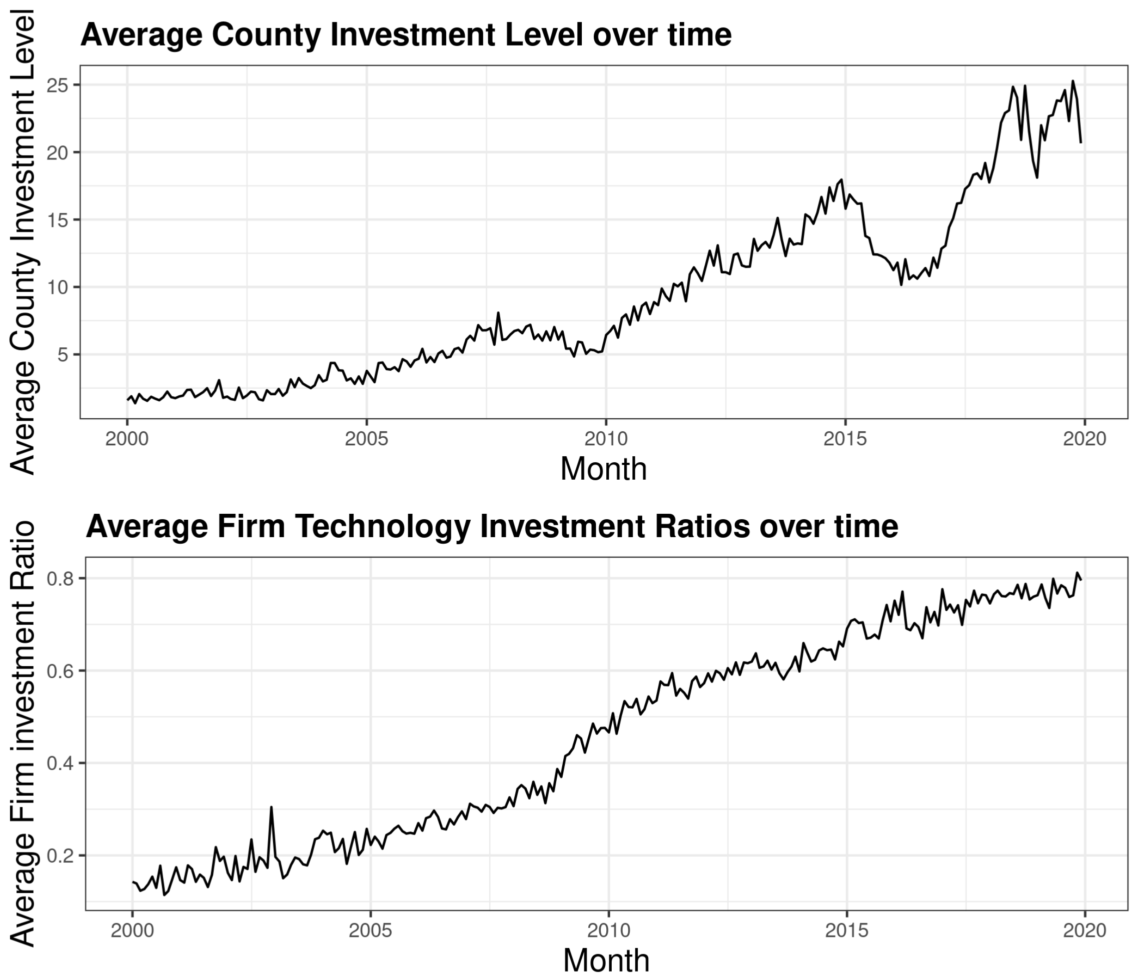

Figure A.11: Theoretical Solutions: Empirical Counterpart to Aggregate Dynamics

The figure plots empirical counterparts of the theoretical transition dynamics shown in A.11. The first panel shows average county-level investment measured in total wells drilled in the county that month. The bottom panel takes each firm's production profile and calculates the new technology to total capital stock ratio. The figure then plots averages over all firms over time.

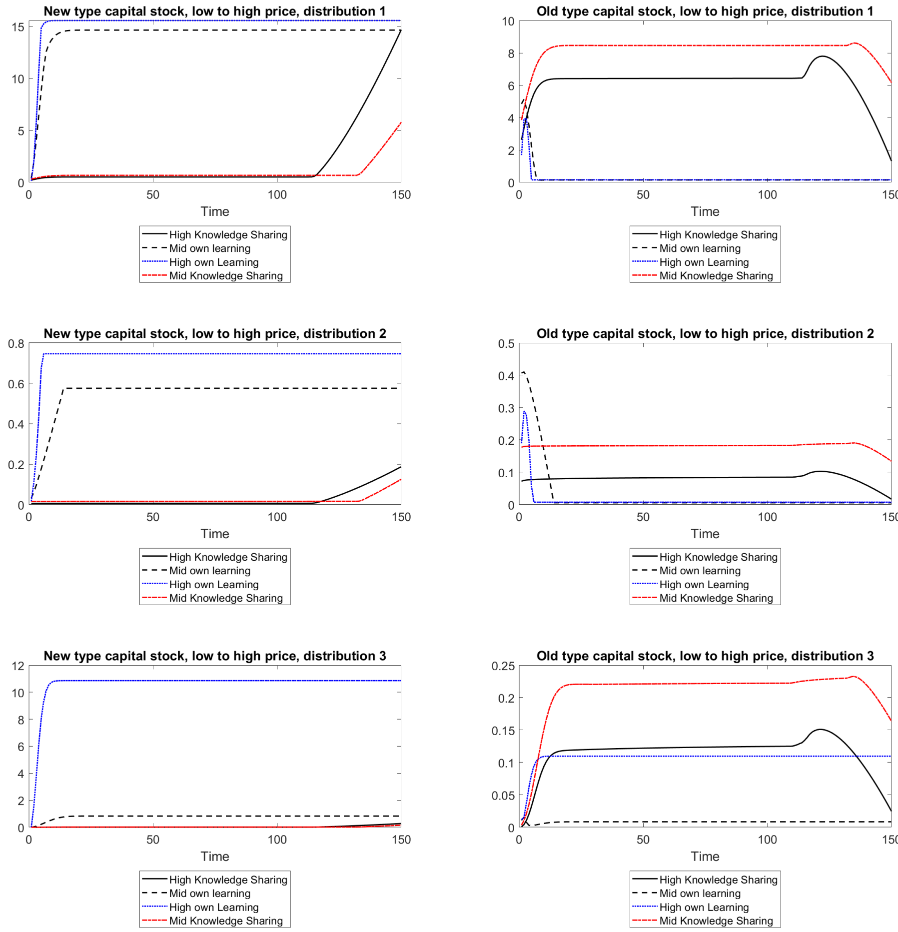

Figure A.12: Aggregate Dynamics: Distribution

The figure depicts aggregate capital stock transition dynamics from economies with different, immutable knowledge sharing attributes. For production function,

$$\pi = P_t \left[A_{old} k_{it}(1-\gamma_{it}) + A_{new,it} k_{it} \gamma_{it}\right], \quad A_{new,it} = [(\eta K_t \Gamma_t)^{\alpha_1}]^{\mu_1} (h_{it}\tilde{h})^{\mu_2},$$
$$\Gamma = \int \gamma g(k,\gamma,h) d\gamma, \quad K_t = \int k g(k,\gamma,h) dk \quad \tilde{h} = (\eta_2 k_{it} \gamma_{it})^{\alpha 2}$$

High knowledge sharing is given by $\alpha_1 = 0.5, \alpha_2 = 0$, mid knowledge sharing is $\alpha_1 = 0.25, \alpha_2 = 0$, High own learning is $\alpha_1 = 0, \alpha_2 = 0.7$ and mid own learning is $\alpha_1 = 0, \alpha_2 = 0.5$.

The panels along the rows show economies with different initial distributions while the columns show new vs. old capital stock. All three distributions begin with mass only in the lower parts of the γ dimension. In initial distribution 1, all firms attempt the new technology but it is a small part of their capital portfolio. In distribution 2, only large firms attempt the new technology and in distribution 3, only small firms attempt it. The transitions begin with low demand price P_t and is subject to an "MIT" shock that sets P_t to a higher level. This distribution is allowed to evolve accord to the Kolmogorov forward equation throughout the transition.

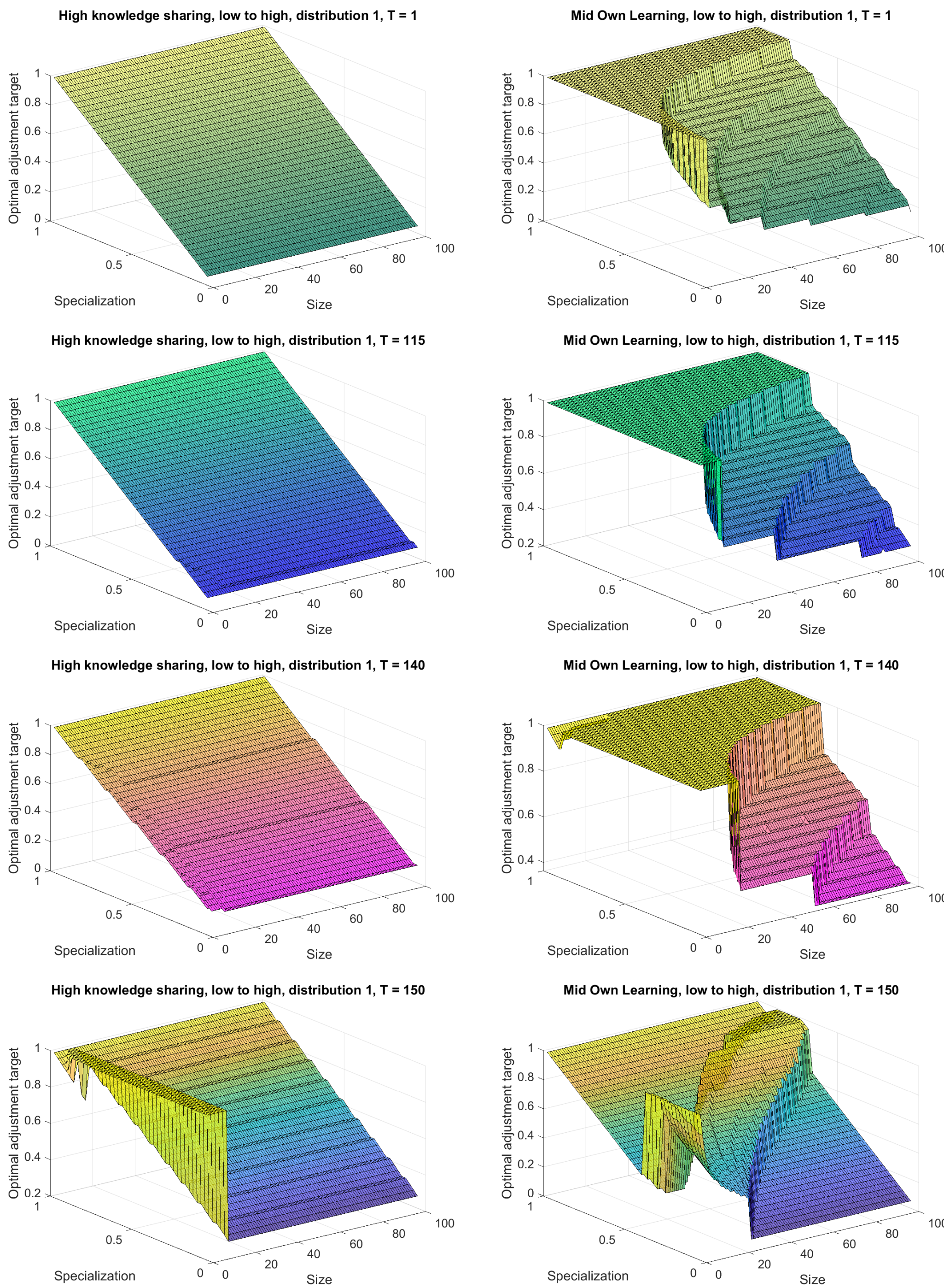

Figure A.13: Optimal Adjustments

The figure shows optimal adjustment targets γ' for firms along the size and specialization distribution. The left column shows the high knowledge sharing economies and the right column shows learning economies. The rows trace out this target throughout the transition period from an "MIT" shock. The transitions begin with low demand price P_t and is subject to an "MIT" shock that sets P_t to a higher level.

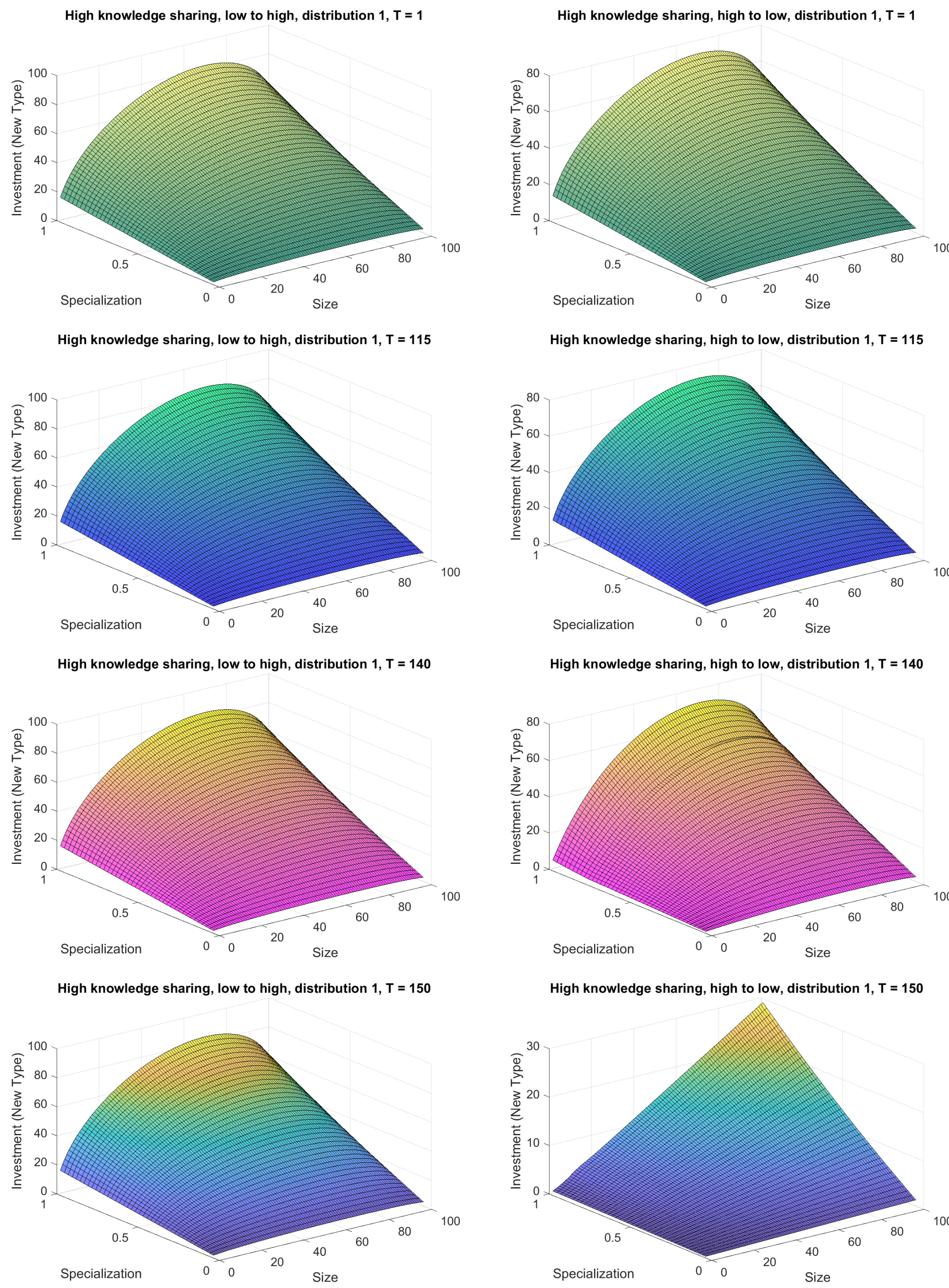

Figure A.14: Optimal Investment

The figure shows optimal investment levels, x_n for firms along the size and specialization distribution. The left column shows the high knowledge sharing economies and the right column shows learning economies. The rows trace out this target throughout the transition period from an "MIT" shock. The transitions begin with low demand price P_t and is subject to an "MIT" shock that sets P_t to a higher level.

A.6 Adoption And Investment

	(1) log(Ex firm Inv) -1 qtr	(2) Knowledge Sharing 1	(3) log(Ex firm Inv) -1 qtr	(4) Knowledge Sharing 2	(5) log(Ex firm Inv) -1 qtr	(6) Knowledge Sharing 3	(7) log(Ex firm Inv) -1 qtr	(8) Knowledge Sharing 4
log(Expiring Leases #), -1 qtr	0.10***		-0.016		0.11***		0.0042	
	(0.0082)		(0.013)		(0.0099)		(0.017)	
log(Ex firm Inv), -1 qtr		-1.61***		2.36		0.94**		25.3
		(0.36)		(3.15)		(0.32)		(106.5)
Constant	-2.33***	-5.31***	-1.02	1.32	-3.64***	7.87***	-2.42***	49.9
	(0.43)	(1.56)	(0.61)	(4.59)	(0.48)	(1.87)	(0.65)	(253.4)
N	6555	6555	3199	3199	3726	3726	2075	2075
Underidentification		157.5		1.43		113.4		0.058
Weak identification		162.3		1.41		124.4		0.057

Table A.3: The impact of shared intangibles on investment levels

The table shows results from

$$I^i_{g,t} = \gamma^{ks_n} \hat{I}_t + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage)$$
$$I_{g,t-1} = \eta + \beta^{ks_n} Exp_{g,t} + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage)$$

The knowledge sharing quartiles ks_n are sorted using the spatial lag model in B.4 with $\hat{Q}$, baseline productivity, as the outcome variable $\mathbf{Y}$ and firm skill as the indirect effect variable $\mathbf{S}$.

Leases granting firms the right but not the obligation to drill tend to have a primary term. If no well is drilled before that expiration, the leases expire and the firms lose the option. In [50], the author shows significant clustering of drilling activity around lease expiration times. In the first stage, $Exp_{g,t}$ counts the total number of leases which are reaching their primary expiration at month t. I use $Exp_{g,t}$ to instrument for investment activity by other firms in the county.

In the second stage investment *levels* by firm i in county g at month t is regressed on the instrumented investment activity variable, $\hat{I}_{g,t}$. The first stage considers investment activity lagged by a quarter since it is likely that investment activity would not react immediately to data that is contemporaneously available.

The results are fairly noisy but the coefficients are consistent with intuition. When other firms increase their investment activity, they are also involuntarily creating knowledge which is potentially valuable to other firms and may impact those firms' investment decisions. In the lowest shared knowledge quartile, increased investment activity by other firms is associated with lower investment levels. Firm investment responses to investment by other firms is largest in magnitude in the highest shared knowledge quartile but the effect is not significant. The result in the third highest quartile is positive and significant.

	(1) 2010Q1-2014Q2	(2) 2014Q3-2015Q4	(3) 2016Q1-2018Q4	(4) 2018Q4-2022Q1
	Firm Size (output in BOE)			
Log(knowledge sharing), Q-skill	-1.96*	-0.44	-0.031	-0.11
	(0.96)	(1.42)	(0.17)	(0.21)
Firm Size Bucket=2	5.57	2.77	-2.45**	7.45**
	(2.93)	(4.98)	(0.75)	(2.73)
Firm Size Bucket=3	7.23**	4.21	-0.028	6.91***
	(2.71)	(5.11)	(0.97)	(1.94)
Firm Size Bucket=4	10.7***	6.74	0.80	6.30***
	(2.70)	(5.05)	(0.58)	(1.57)
Firm Size Bucket=2 × Log(knowledge sharing), Q-skill	1.77	0.30	-0.55*	0.70
	(1.02)	(1.41)	(0.27)	(0.40)
Firm Size Bucket=3 × Log(knowledge sharing), Q-skill	2.32*	0.86	0.057	0.78**
	(0.94)	(1.44)	(0.22)	(0.26)
Firm Size Bucket=4 × Log(knowledge sharing), Q-skill	3.05**	1.04		
	(0.95)	(1.42)		
	Firm Specialization $\frac{New\ Technology\ Capital}{TotalCapital}$			
Log(knowledge sharing), Q-skill	1.62***	-0.0041	-0.71	1.16
	(0.26)	(0.51)	(0.53)	(1.48)
Firm Specialization	-0.99	2.95	3.10	-5.26
	(1.18)	(2.07)	(2.09)	(5.54)
Log(knowledge sharing), Q-skill × Firm Specialization	-0.73*	0.77	0.71	-1.19
	(0.37)	(0.59)	(0.61)	(1.51)
	Leverage Ratio: Compustat $\frac{dlttq}{atq}$			
Log(knowledge sharing), Q-skill	-0.42	0.10	0.43	-0.37
	(0.68)	(0.62)	(0.56)	(0.98)
Leverage Ratio	4.15	-3.19	-2.65	11.0
	(4.64)	(4.51)	(2.87)	(7.11)
Log(knowledge sharing), Q-skill × Leverage Ratio	1.50	-0.87	-1.17	3.25
	(1.69)	(1.51)	(1.21)	(2.18)

Table A.4: Investments by Firm type

The table shows results from

$$I^i_{g,t} = \alpha + \beta log(Ks_g) \times Char_{i,t} + \Gamma[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]' + \epsilon_{i,g,t} \quad |t \in period_j$$

The regressions are conducted at the firm-county-month level. Each regression is conducted separately for time periods j listed across the columns. The dependent variable here is new technology investment levels given by units of horizontal wells drilled. The three panels separate firms based on their size, specialization and leverage ratio. Note that in the first panel, firms are divided into quartiles rather than using their continuous firm value given the large dispersion in firm sizes.

	(1) 2010Q1-2014Q2	(2) 2014Q3-2015Q4	(3) 2016Q1-2018Q4	(4) 2018Q4-2022Q1
		Firm Size (output in BOE)		
log(Knowledge Sharing)	18.2*	0.37	-0.048*	0.18
	(7.46)	(0.28)	(0.022)	(0.15)
Firm Size Bucket=2	-60.5**	3.18	-1.21	-0.42
	(23.4)	(3.02)	(1.29)	(1.45)
Firm Size Bucket=3	-49.8*	-0.77	-0.23	-5.98
	(20.0)	(1.08)	(0.82)	(5.15)
Firm Size Bucket=4	-61.1**	-0.40	-0.15	-0.29
	(21.0)	(1.04)	(0.71)	(0.83)
Firm Size Bucket=2 × log(Knowledge Sharing)	-19.6*	0.53	-0.48	0.075
	(8.84)	(0.78)	(0.37)	(0.15)
Firm Size Bucket=3 × log(Knowledge Sharing)	-16.3*	-0.53	-0.027	-2.14
	(7.32)	(0.32)	(0.10)	(1.80)
Firm Size Bucket=4 × log(Knowledge Sharing)	-19.4*	-0.36		
	(7.55)	(0.28)		
		Firm Specialization $\frac{New\ Technology\ Capital}{TotalCapital}$		
log(Knowledge Sharing)	-3.25***	0.46**	-0.26***	-8.24
	(0.72)	(0.16)	(0.076)	(6.29)
Firm Specialization	6.43*	-4.68***	-1.43***	21.2
	(2.51)	(0.61)	(0.27)	(17.4)
log(Knowledge Sharing)× firmboerat	3.16***	-0.47*	0.28***	8.64
	(0.79)	(0.18)	(0.082)	(6.58)
		Leverage Ratio: Compustat $\frac{dlttq}{atq}$		
log(Knowledge Sharing)	0.53	-0.22*	-0.0053	0.051
	(0.28)	(0.099)	(0.066)	(0.039)
Leverage Ratio	0.079	1.20*	-0.49	0.20
	(1.88)	(0.54)	(0.44)	(0.37)
log(Knowledge Sharing)× levratio	0.13	0.44*	-0.16	0.087
	(0.69)	(0.22)	(0.19)	(0.11)

Table A.5: Adoption Rates by firm type

The table shows results from

$$I^i_{g,t} = \alpha + \beta log(Ks_g) \times Char_{i,t} + \Gamma[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}]' + \epsilon_{i,g,t} \quad |t \in period_j$$

The regressions are conducted at the firm-county-month level. Each regression is conducted separately for time periods j listed across the columns. The dependent variable here is new technology adoption rates given by $\frac{\frac{I^i_{g,t,new}}{I^i{g,t,total}} - \frac{k^i_{g,t,new}}{k^i_{g,t,total}}}{\frac{k^i_{g,t,new}}{k^i_{g,t,total}}}$. The first term in the numerator is the firm's investment ratio of new to total investments in county g at time t. The second term is it's existing new technology production ratio based on the firm's installed capital stock at time t. This is taken as a proportion of the firm's existing technology ratio.

The three panels separate firms based on their size, specialization and leverage ratio. Note that in the first panel, firms are divided into quartiles rather than using their continuous firm value given the large dispersion in firm sizes.

A.7 Industry Level Effects

Dependent Variable: Geography Comparisons of performance $\frac{Firm^{geo}_{skill,t} - Firm_{skill,t-1}}{Firm_{skill,t-1}}$								
Q-skill knowledge sharing								
	(1) Ex firm investments	(2) Knowledge Sharing 1	(3) Ex firm investments	(4) Knowledge Sharing 2	(5) Ex firm investments	(6) Knowledge Sharing 3	(7) Ex firm investments	(8) Knowledge Sharing 4
$\mathbf{1}_{\{High\ Oil\ Price\}}$	-0.0026		-0.0035*		0.012***		0.027***	
	(0.0029)		(0.0015)		(0.00065)		(0.00095)	
Ex firm Investments		5.06		0.33		0.21**		0.13***
		(6.47)		(0.99)		(0.070)		(0.038)
Constant	-0.031**	0.18	-0.17***	0.036	-0.059***	0.0097	-0.029***	-0.0093
	(0.010)	(0.25)	(0.012)	(0.17)	(0.0035)	(0.0054)	(0.0059)	(0.0082)
N	2147	2147	4248	4248	19983	19983	13134	13134
Underidentification		0.77		5.56		331.6		733.6
Weak identification		0.76		5.61		344.3		818.0
Model Diff		2.96		0.72		0.13		2.31
Diff pvalue		0.085		0.39		0.72		0.13

Marginal effects; Standard errors in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.6: Average Growth effects: Within geography vs firm average

The Table shows results from:

$$\begin{aligned} Y^j_{w,t} &= \alpha + \gamma^{ks_n} \hat{I}^j_t + \Gamma\left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage) \\ I^j_t &= \eta + \beta^{ks_n} \mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma\left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage) \end{aligned}$$

in table format. The first stage investment results are shown in the odd-numbered columns while the second stage results for each knowledge sharing quartile is shown in increasing order in the even-numbered columns. The knowledge sharing value, ks, is Q-skill. These are estimated in the spatial lag model by regressing $\mathbf{Y} = Q$, on a weighted vector of nearby firm skill.

Y is given by the proportional difference between the firm's within-geography estimated skill level and the firm's average skill level nation-wide. The first stage investment measure is ex-firm investment which is given by $county - investments - firminvestments$. The Firm performance variable is estimated using the same methodology in estimating the skill vector for the Spatial lag model. The Arp's model is applied with a firm-level fixed effect interacted with the geography and quarter where relevant.

$$Y_{w,t} = \beta log(Age) + \gamma \mathbf{1}_{\{Firm\}} \times \mathbf{1}_{\{q\}} \times \mathbf{1}_{\{geo\}} + \epsilon_{w,t} \tag{A.1}$$

Wald tests for the statistical difference between estimates is shown in "Model Diff" and "Diff pvalue".

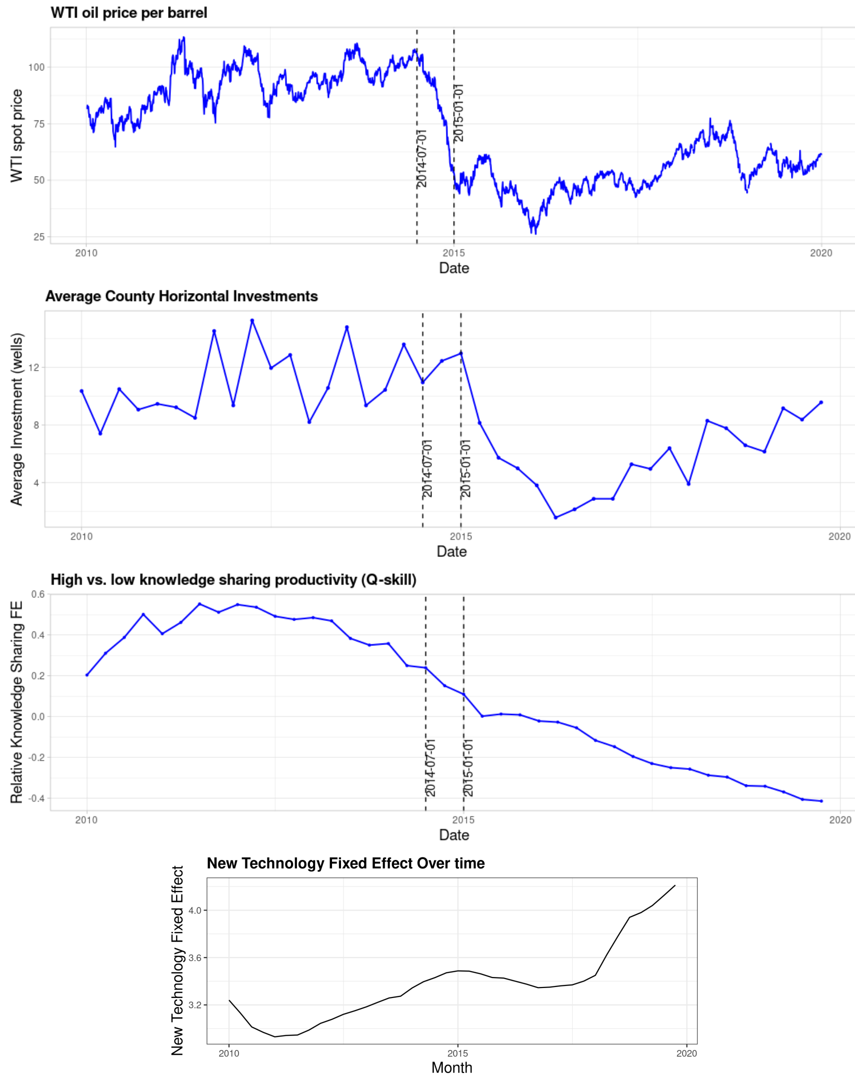

Figure A.15: Relative productivity between knowledge sharing regions

The top panel shows the WTI oil price over the same time period. The two lines show the corresponding time period when oil prices were undergoing substantial declines. The second panel shows average county level investment rates including the large decline that occurs when oil prices drop.

The third panels show results from regressions,

$$log(O_{w,t}) = \beta log(Age_{w,t}) + \gamma_t \mathbf{1}_{\{ks^n, n=4\}} + \epsilon_{w,t} \quad |n = 1, 4, \quad new\ type\ only$$

These are estimated in the spatial lag model by regressing $\mathbf{Y} = Q, production\ efficiency$ on a weighted vector of nearby firm skill.

Regressions are run at a quarterly level and the dummy variable is 1 for high knowledge sharing region wells ($n4$) and 0 for low knowledge sharing region wells ($n1$).

The fourth panel repeats the relative productivity of horizontally drilled wells as compared to vertical wells.

APPENDIX B

SUPPLEMENTARY MATERIALS

B.1 Firm and Geography controls

Two sets of controls are used throughout the main empirical studies. One set captures geography -level characteristics and the other is a series of firm controls. They are listed and described here:

Geography Controls

Variables are estimated at the county -month level. For variables that include geological formation data, the effective formations within a county are aggregated.

- formation quality (interval) - variance adjusted measure of the formation quality. This measurement is conducted at the geological formation level which may include many counties. An individual county may also be located in multiple formations. This measurement is given by,

  $$\frac{avg(IP)}{sd(interval\ (ft))}$$

  The initial production rate, "IP" is the cumulative production over the first 6 months. This is often used by industry practitioners as a first guess at how productive a well will be. Interval length measures the length of the horizontally perforated interval in feet. This interval is the portion of the pipe where with the perforations so that fracking fluid can be released into shale rock. I use this as a proxy for well complexity.

- formation quality (drill time) - variance adjusted measure of formation quality. This measurement is conducted at the geological formation level which may include many counties. An individual county may also be located in multiple formations. This measurement is given by,

$$\frac{avg(IP)}{sd(drill\ time)}$$

The IP rate is as defined for the interval based formation quality above. Drill time is measured by the number of days between spud and completion time. Spudding describes the process of beginning to drill a well, at completion it can begin to produce. Drill time is another proxy for how complex a well is.

- average output - Geography level average IP rate.
- county well # - Total number of wells in operation in the county at month t.
- interval variance - interval length variance (ft) of the operating wells in the county at month t.
- county technology specialization - proportion of the county's total output that is from horizontal wells.

Firm Controls: Firm level controls are measured using the full sample at month t.

- sd(interval length) - standard deviation of interval length (ft) across all wells operated by the firm at month t.
- sd(production efficiency) - standard deviation of production efficiency across all wells operated by the firm at month t; production efficiency measures $\frac{avg(IP)}{Interval\ (ft)}$
- well count - total number of wells operated by the firm at month t
- drilling efficiency - average drilling efficiency over all wells operated by the firm; drilling efficiency measures $\frac{drill\ time}{interval\ (ft)}$

- average IP - average initial production rate for all wells operated by the firm at month t.
- average production efficiency - average production efficiency for all wells operated by the firm at month t
- interval efficiency - to estimate a firm's interval efficiency at quarter q of month t, I use the log version of Arp's model for monthly oil production,

$$log(O_{w,t}) = \alpha_w + \beta log(Age_{w,t}) + \gamma \mathbf{1}_{\{i\}} \times \text{interval length}_w$$

  γ is the interval efficiency. It measures the effectiveness of interval lengths on output interacted with a firm dummy. α_w is the proxy for baseline production rate Q.
- skill - Estimated skill vector used in the spatial lag model of the main analysis.

B.2 County List

The spatial lag model to measure knowledge sharing was applied to a random sample of counties. In this section, I list the counties that are included in the analysis along with the total number of horizontal wells included from each county. The lists are split by horizontal and vertical wells since the study is done once using all horizontal and all vertical wells. The list shows the state followed by the county name.

Horizontal county list

AR CLEBURNE (1,095),CA KERN (1,861),CO GARFIELD (36),LA BIENVILLE (231),LA BOSSIER (427),LA CADDO (725),LA PLAQUEMINES (87),MT HILL (6),MT RICHLAND (931),ND BOWMAN (216),ND BURKE (352),ND DIVIDE (759),ND DUNN (2,800),ND

WILLIAMS (3,107),NM EDDY (4,326),NM LEA (3,900),OH MONROE (432),OK ALFALFA (1,237),OK BEAVER (244),OK BECKHAM (192),OK CANADIAN (1,376),OK CARTER (280),OK CUSTER (195),OK DEWEY (551),OK GRANT (470),OK HUGHES (852),OK KAY (117), OK KINGFISHER (1,658),OK LATIMER (26),OK LOGAN (416),OK PAYNE (414),OK PITTSBURG (1,113),OK SEMINOLE (203),OK WASHITA (309),OK WOODWARD (63),PA ARMSTRONG (138),PA BRADFORD (1,437), PA CLEARFIELD (98),PA FAYETTE (221),PA GREENE (1,296),PA LYCOMING (919),PA TIOGA (746),TX ATASCOSA (1,216),TX BEE (66),TX BRAZOS (481),TX CLAY (27),TX DENTON (1,393),TX DEWITT (2,325),TX FREESTONE (82),TX GAINES (129),TX GLASSCOCK (1,279),TX GONZALES (1,939),TX GREGG (31),TX HARRISON (502),TX HOCKLEY (34),TX HOOD (818),TX HOWARD (1,943), TX HUTCHINSON (27),TX JACK (141),TX LAVACA (398),TX LEON (142),TX LIPSCOMB (1,172),TX LOVING (2,526),TX MADISON (263),TX MAVERICK (196),TX MCMULLEN (2,139),TX MONTAGUE (994),TX MOORE (10),TX NACOGDOCHES (247),TX OCHILTREE (776),TX PARKER (1,397),TX POTTER (48),TX REAGAN (1,901),TX REEVES (3,902),TX ROBERTS (384),TX ROBERTSON (164),TX SMITH (49),TX WEBB (3,013),TX WHEELER (889),TX WINKLER (420),TX WISE (1,596),TX YOAKUM (396),TX ZAPATA (16),TX ZAVALA (459),UT DUCHESNE (245),WV JACKSON (24),WV MARSHALL (553),WV MINGO (129),WY CONVERSE (1,084), WY SWEETWATER (107)

Vertical county list

AR LOGAN (530), AR SEBASTIAN (365), LA BIENVILLE (444),LA BOSSIER (905),LA CADDO (1,610),LA GRANT (82),NM EDDY (3,936),NM LEA (2,520),OH MONROE (696),OH STARK (385),OK HARPER (202),OK KAY (438),OK NOBLE (337),OK NOWATA (402), OK WOODWARD (783), PA ARMSTRONG (2,107), PA CLARION (1,097),PA CLEARFIELD (719),PA ELK (782), PA FAYETTE (1,444),PA FOREST (2,298),PA GREENE (981), PA INDIANA (1,927), PA VENANGO (1,296),PA WASHINGTON (701),TX BEE (405),TX

CROSBY (636),TX DENTON (285),TX DUVAL (469), TX FREESTONE (1,255),TX GAINES (1,616),TX GARZA (360), TX GLASSCOCK (3,093),TX GREGG (340), TX HOCKLEY (428),TX HOWARD (2,167),TX IRION (1,205),TX JACK (1,101),TX LAVACA (328),TX LEON (410),TX LOVING (357), TX MCMULLEN (360),TX MEDINA (487), TX MITCHELL (943), TX MONTAGUE (446),TX MOORE (528),X NACOGDOCHES (958), TX NUECES (265), TX REAGAN (2,347), TX REEVES (914),TX SMITH (392), TX UPSHUR (199),TX WEBB (1,503), TX WHEELER (953), TX WICHITA (1,511), TX WINKLER (683),TX WISE (359), TX YOAKUM (1,020),TX YOUNG (645), TX ZAPATA (1,146), UT DUCHESNE (3,062), WV HARRISON (673), WV JACKSON (315),WV LEWIS (689), WY CARBON (1,242), WY FREMONT (431),WY SWEETWATER (2,163)

B.3 Institutional Details

The American oil and gas industry is large and dispersed unlike its more concentrated international integrated oil giants such as BP and Total. The actual process of oil production is conducted by "upstream" companies who specialize in exploration and production. These companies do not operate gas stations or refine oil. They explore, drill, and deliver the crude oil to what are known as "midstream" companies. Despite that, all of the major integrated oil companies do departments involved in exploration and production. This market structure makes for an advantageous empirical setting for this case study. There is a disperse set of firms from who vary significantly in characteristics like size and skill. The industry is not dominated by a handful of firms even though large firms are active.[1] Finally, the end product which they produce is uniform so there is a clear demand price faced by all firms with limited product differentiation.

1. Additionally, because oil production is such a capital-intensive endeavor, joint ventures and profit sharing agreements are common. While I don't measure these directly, it is another reason to believe that knowledge spillovers are likely in this industry.

The exploration and production process Traditional oil wells consist of vertical wells which are drilled straight down to large oil reservoirs. Directional and horizontal wells are designed so that the drill can be angled. While this technology has existed for decades, it was not truly transformative until this technique was combined with hydraulic fracturing. By drilling horizontally into shale rock and fracturing the rock with chemicals through holes in the horizontal perforated interval, these new wells are able to access the oil and natural gas trapped within shale rock. In this paper, I'll use horizontal wells as the indicator for the new technology instead of fracking data. In the time period studied, these will be virtually indistinguishable.

After a firm has secured a lease giving it the right to drill[2] a well and they've conducted significant research including seismic tests, they submit permits to state regulators for the permission to drill. These permits include general descriptions of the well including information on the location of the well, whether the well will be vertical or horizontal, how far down the vertical portion of the well will be, how long the horizontal interval of the well will be and basic information on the operators. Additional information will vary from state by state. For example, some states maintain strict data on fracking fluids used because of environmental concerns while others do not. I use the availability of these additional data variables as an alternative measure of knowledge availability in the tests below.

The time between scouting, permitting, and drilling can differ significantly. Most leases give firms a three year primary term. If they have not drilled at least one well in that period, the lease expires. In the studies, I'll use overall investment activity within a quarter or two to approximate shared knowledge because it is difficult to measure precisely when firms knew what given this variance in time between preparation and actual drilling. Further, using

2. These can be from private actors or the bureau of land management.

permitting dates to estimate knowledge sharing is likely to introduce a lot of noise as well and risks losing large portions of the observations due to data paucity.

There is significant heterogeneity across geographies due to geological differences. Figure A.1 shows the variance of drilled *depths* as well as the spacing of wells drilled in North Dakota and Texas respectively. The North Dakota example shows relatively well lined wells (most of these are horizontal so they line up accordingly). Also the depths are fairly homogeneous as depicted by the colors. Texas on the other hand shows wells drilled at a variety of depths. The example does not suggest one area is more productive or easier to drill. With the more homogeneous wells, it is expected that optimal well design will be similar making experience more easily transferred. On the other hand, there is more experience in Texas so information regarding what does not work is well developed. This heterogeneity will be a useful building block towards differences in knowledge sharing propensity.

Oil prices & Investment. The most relevant oil price benchmark used by American oil companies is the West Texas Intermediate. While there is often a spread between the WTI spot price and other global oil prices like the Brent, the WTI generally tracks world oil prices over a long time horizon. As a result, it is difficult to argue that any individual firm can influence the price. Figure A.2 depicts the WTI over the time period of interest. It has been well documented that oil prices can be volatile and there's been a large literature studying this important time-series. I do not take advantage of the useful short-term variance in oil prices but focus on a large structural break that occurred during this time period. Plotted in red in figure A.2 is the estimated probabilities from fitting a two-stage Markov switching model to the price. There are two significant regimes between 2010 and 2018. In the sections below, I'll argue that these two regimes provide a useful source of exogenous variation in long-term firm investment trends.

B.4 Spatial Lag Model

Let Y be an outcome of interest described in section "Indirect Effects & Outcomes of Interest" of the main paper. The general format of the spatial autoregressive model is run for each county, g separately. g indexes all oil producing counties in the United States. T is the total number of months where drilling activity has occurred in county g. Importantly, for time $t < T$ where there is no drilling activity, the month still appears in the data. Let N be the total number of geo-spatial coordinates where any well has ever been drilled in county g.

$$\mathbf{Y}_t^{g,H} = \gamma_1^{g,H}\mathbf{W}\mathbf{Y}_{t-1}^{g,H} + \gamma_2^{g,H}\mathbf{X}_t^{g,H} + \beta^{g,H}\mathbf{W}\mathbf{S}_{t-1}^{g,H} + \epsilon \tag{B.1}$$

$\mathbf{Y}_t$ is a $N \times 1$ vector which is nonzero in any $n \in N$ if that well was drilled in month t and zero otherwise. $\mathbf{W}$ is a $N \times N$ matrix of spatial weights, $\mathbf{X}_t$ is a $N \times K$ matrix of K different controls corresponding to the reference well n and time-varying geography controls, and $\mathbf{S}_t$ is a $N \times 1$ vector of estimated firm skill. H is an index for the oil price regime and is 1 during the pre-period when oil prices are high and 0 otherwise. I maintain the g and H superscript in the matrix format to emphasize that these regressions are conducted by county-year individually. In the subsequent discussion, I will suppress the g, H for notational purposes but the reader is asked to remember that these analyses are conducted separately for each price regime and county. The coefficient $\beta^{g,H}$ is not used for statistical inference, rather it measures the impact of other firms on a particular investment outcome. To reduce noise, I use the point estimate but only for geographies where the coefficient is statistically significant. Other geographies are left out of the analysis. I repeat this exercise in the appendix using the lower end of the 95% confidence interval but for all counties.

A more direct estimate of knowledge sharing would be to use $\hat{\gamma}_1^{g,H}$ as the index of knowledge sharing instead of $\hat{\beta}^{g,H}$. This coefficient captures persistence and contagion over time and between firms which also contain important evidence of knowledge sharing. However, that estimate suffers from contamination by confounding factors outside of knowledge sharing. There could be a number of reasonable explanations for why $\hat{\gamma}_1^{g,H}$ is significant. There are fewer, compelling stories for why the skill level of nearby firms, estimated using data both within and outside of the county, should have a meaningful effect. [3] Finally, I do not use the full time-varying $\hat{\beta}^{g,H}$ as the relevant knowledge sharing estimate. I pre-sort counties using their average $\hat{\beta}^{g,H}$ measures in the years prior to 2015 then form the knowledge sharing quartiles using that estimate.

B.4.1 The spatial weight operator

The terms, $\mathbf{WY}_{t-1}$ and $\mathbf{WS}_{t-1}$ are analogous to the lag operator $\mathbf{L}.y$ commonly seen in autoregressive models. Rather than regressing outcome y on its lag, y_{t-1} observation, the specification regresses y on a number of nearby investments in vector Y_{t-1}. $\mathbf{W}$ acts as an operator weighting the relative impact of those wells by its inverse distance. To see this more clearly, consider a simple example. Suppose the outcome Y has realizations denoted by I. We are interested in a reference well $w = 3$ drilled in month t for this sample set with

3. Of course, $\hat{\beta}^{g,H}$ is likely an underestimate of actual knowledge sharing. The estimated effects from knowledge sharing is also likely to be conservative because the measurement of knowledge sharing is itself fairly conservative.

$N = 5$.

$$Y_t = \mathbf{W}Y_{t-1} + \mathbf{W}S_{t-1} \tag{B.2}$$

$$\begin{bmatrix} 0 \\ 0 \\ I \\ 0 \\ 0 \end{bmatrix}_t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{d(2,3)} & 0 & 0 \\ 0 & \frac{1}{d(3,2)} & 0 & \frac{1}{d(3,4)} & \frac{1}{d(3,5)} \\ 0 & 0 & \frac{1}{d(4,3)} & 0 & 0 \\ 0 & 0 & \frac{1}{d(5,3)} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ I' \\ 0 \\ 0 \\ I'' \end{bmatrix}_{t-1} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{d(2,3)} & 0 & 0 \\ 0 & \frac{1}{d(3,2)} & 0 & \frac{1}{d(3,4)} & \frac{1}{d(3,5)} \\ 0 & 0 & \frac{1}{d(4,3)} & 0 & 0 \\ 0 & 0 & \frac{1}{d(5,3)} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ S' \\ 0 \\ 0 \\ S'' \end{bmatrix}_{t-1} \tag{B.3}$$

In this example, there are 5 wells drilled in the county but only $w = 2, 4, 5$ are in the "effect set" of well $w = 3$[4] because $w = 1$ is too far away. However, well $w = 4$ was not drilled in time $t - 1$ so it does not impact the reference well $w = 3$. I', I'' are the outcomes for wells 2 and 5 while S', S'' are the estimated time $t-1$ skill levels of the firms that drilled wells 2 and 5. $d(w, w'')$ denotes the distance between wells w, w'' measured using their coordinates. For time t the corresponding data row for well $w = 3$ can be re-cast in traditional OLS notation for a specific data observation w,

$$y_{w,t} = \gamma_1 y_{w,t-1} + \beta s_{w,t-1} + \epsilon_{w,t}$$

$$I_{w,t} = \gamma_1 \underbrace{\left[\frac{1}{d(w,w')} I'_{w',t-1} + \frac{1}{d(w,w'')} I''_{w'',t-1}\right]}_{y_{w,t-1}} + \beta \underbrace{\left[\frac{1}{d(w,w')} S'_{w',t-1} + \frac{1}{d(w,w'')} S''_{w'',t-1}\right]}_{s_{w,t-1}} + \epsilon_{w,t}$$

For well $w = 3$ then, the coefficients γ_1 and β capture a version of a linear regression of outcome $y_{w,t}$ on a weighted sum of influence well outcomes $y_{w,t-1}$ and firm skills $s_{w',t-1}$. The explanatory variables on the right side also include a vector of controls X_t^g. The controls correspond to firm and geography level time-varying controls at time t for firm i who drilled

4. I'll define this shortly

reference well w as well as geography level variables[5].

To construct the dataset, I define an 'effect set", wells which are close enough that they should be included as a potential influence data point. This effect set is implemented through the weight matrix $\mathbf{W}$. Consider well w contained in vector Y. Let $\bar{w} < N$ be another well in the county. If $\bar{w}$ is in the effect set, then element $(w, \bar{w})$ contains the inverse distance between $w, \bar{w}$. If $\bar{w}$ is not in the effect set for well w then $(w, \bar{w})$ will contain a zero. This is a symmetric matrix so the same weighting will be applied to well $\bar{w}$ from element w. In the example above (eq B.2), well 1 is not included in the reference set of well 3 so the matrix $\mathbf{W}$ will always assign zero weight between those wells. Also note that $w = 3$ is included in the effect set for wells 2,4,5. If, for example, well 4 is drilled at time $t + 1$, then $\mathbf{W}$ will assign weight to well 3 outcome I when 4 becomes the reference well.

For every well in the sample, I conduct a geospatial match with every other well in the sample and keep pairs which are within thirty miles of the reference well. Thirty is arbitrarily chosen to maximize the use of the dataset. For smaller effect sizes, large amounts of data from important oil producing regions such as Colorado often do not enter the dataset. For larger effect sizes, the model becomes intractable to estimate in dense oil producing states such as Texas. Thirty miles is empirically the best compromise to maximize use of the data while maintaining tractability. Importantly, these spatial matches are not limited by municipal distinctions so that wells which are likely in the same *geological* formation but across municipal county lines are included if they are close enough. Then, for every municipal county g, the dataset includes every reference well physically located in that county and every well in its "effect set" which appears because its corresponding coordinate in $\mathbf{W}$ will

5. These include estimates of firm skill, geography side production propensity, the number of producing wells in firm i's portfolio nation-wide at time t, geography-wide technology efficiency, the firm's new technology investment ratio, and the geography's new technology production ratio.

be nonzero.

B.5 Data Availability Measure Values

The data availability measures used in the spatial lag model comprises a set of variables which are not uniformly available across all geographies. To construct the measure, I first use the geo-matching done in the spatial lag model. For those matched wells that were recently drilled, I create a variable that is 1 if the variable is available for the influence wells and 0 otherwise. Then, I define data availability for variable u to be $\frac{\sum \mathbf{1}_{\{u\ available\}}}{Total\ Influence\ Wells}$ for each reference well. For example, if a given well has two influence wells but the data availability measure is only available for one of those wells, the measure will be $\frac{1}{2}$.

For the standard deviation measure which is used to test the usefulness of information also starts from the geo-matched reference and influence wells. Here, the actual value of the dataset is weighted by the inverse distance just as in the main spatial lag model. Then, I take the standard deviation of the distance weighted u variable.

This exercise is repeated for the two sets of variables described below. Then, for the spatial lag specification I run,

$$\mathbf{Y}_t^{g,H} = \gamma_1^{g,H}\mathbf{W}\mathbf{Y}_{t-1}^{g,H} + \gamma_2^{g,H}\mathbf{X}_t^{g,H} + \beta_u^{g,H}\mathbf{E}\mathbf{A}_{t-1}^{g,H} + \epsilon \tag{B.4}$$

where $\mathbf{A}$ is a $N \times k$ matrix of the data availability measure, which contains $\mathbf{1}_{\{u\ available\}}$. Note that unlike the skill vector, this variable is not weighted by distance. Instead, I include an indicator matrix $\mathbf{E}$ whose values are $\frac{1}{Total\ Influence\ Wells}$ for each reference well in row $n \in N$. The $\mathbf{E}$ matrix is zero for influence wells that are outside of the effect set for the reference well so it still contains a spatial element. Then, for each county, I average over $\beta_u^{g,H}$ during the pre-period $H = 1$ and over all u variables in the set to construct a general data availablity index. It should be interpreted as regions where the availability of the data is positively associated with outcomes $\mathbf{Y}$.

For the standard deviation test, a similar exercise is conducted. Let $\mathbf{U}$ denote an analogous matrix to $\mathbf{A}$ except that each element contains the actual realization of the extra data variables instead of an indicator equal to 1 if that variable is available. To be concrete, consider u_1, the variable denoted in column 1 of $\mathbf{U}$. I first create a weighted vector for this variable. let i denote influence wells in the effect set for reference well n. Then, the weighted vector is given by $(w_1 u_{1,1}, \ldots, w_i u_{1,i}, \ldots)$. I then take the standard deviation over this vector. This exercise is repeated for each of the variables described in B.5. Finally, I take the mean of the standard deviations calculations across all of the variables in data set 1 or 2. This variable is used directly to sort wells (as opposed to counties) in the baseline analysis. Note that it does not get applied to the spatial lag model like the other variables. This tests directly whether or not informativeness of the shared data has an effect.

List of Data availability Variables

The data availability variables are described here. The are broadly categorized into two sets based on the nature of the variables. The first captures more closely associated with inputs to drilling, especially the fracking process. The second set has to do with well design problems that are associated with the physical nature of the geology. These variables are not uniformly better for production which is why I use availability measures instead of the variable realizations themselves.

Data Availability Set 1 (measures in lbs)[6][7]

- proppant - sand or ceramic used in the fracking process to hold fractures open. This is usually combined with chemicals and water
- biocide - chemical used to to maintain production flow, inhibits corrosion and allows for smooth flow.
- breaker - chemical used to reduce the viscosity of fluid
- buffer - chemical comprised of water, acide, and salt to control the pH level of stimulation fluids.
- clay control - additive in fracking to control the migration of clay particles when it interacts with water-based fluid.
- cross linker - compound used to create a viscous gel that stimulates the productivity of a well.
- friction reducer - additive used to reduce the friction when pumping sand and other fracking fluids into the well.
- gelling agent - reacts with oil to form solids which can be used in sealing or limiting sand production
- iron control - reduces iron build-up in fluids to maintain flow of the fracturing chemicals.
- scale inhibitor - prevent or slow down scaling (solid deposits) that prevent fluid flow through the pipeline.
- surfactant - chemical used to add emulsion to fluid to reduce viscosity.

6. Most of these measures are recorded in units of lbs per 1000 ft of perforated interval or square feet where appropriate.

7. Many of these definitions are available form the Schlumberger oilfield glossary, https://glossary.oilfield.slb.com/en/Terms/

B.6 Learning

The knowledge sharing estimates described above do not exclude the possibility of learning or experimentation by individual firms. To facilitate analysis of learning as either a complementary or alternative explanation, I include a set of learning propensity estimates described here. Broadly, the philosophy behind these estimates is to measure the effectiveness of either experimenting or acquiring experience *within a geography.* There are a number of ways that one can measure experimentation. I use the most clearly measured variable in my data, horizontal fracked interval lengths and drill times. Of course, there could be other ways that firms are experimenting which will not be captured in this measure. For the experience measure, the reader is reminded that oil is a natural resources which suffers from decline over time. Thus, the experience outcomes should be carefully interpreted. The experimentation measure is specified in B.5 for production outcome $Y_{w,t}$,

$$
\begin{aligned}
Y_{w,t} \quad &= \alpha + \beta_{geo,1}\mathbf{1}_{\{geo\}} \times Exp + \beta_{geo,2}\mathbf{1}_{\{geo\}} \times Imit \\
&+\Gamma[\mathbf{X}_{i,t}\ \mathbf{X}_{geo,t}]' + FE_f + FE_q + \epsilon_{w,t} \\
Exp_I \quad &= \left|Interval - Avg(Interval)_{f,q}\right| \\
Imit_I \quad &= -1 \times \left|Interval - Avg(Interval)_g = g, t\right|
\end{aligned}
\tag{B.5}
$$

Exp variables capture the deviation of a particular well from the firm average nationwide. $Imit$, on the other hand, captures how close a well's parameters are to the average in the geography, or "imitation". This variable captures a version of the knowledge sharing estimate without the weighting by firm skills. Naturally, optimal well design varies across regions so imitation is expected. The difference in returns to experimentation as compared to imitation, $\beta_1 - \beta_2$, capture regions where deviating from what other firms are doing leads to higher Y outcomes.

In addition, I also include experience measures which are estimated in

$$Y_{w,t} = \alpha + \beta^{exp}_{geo}\mathbf{1}_{\{geo\}} \times Exp_{f,t} + +\Gamma[\mathbf{X}_{f,t}\ \mathbf{X}_{geo,t}]' + FE_f + FE_q + \epsilon_{w,t} \tag{B.6}$$

experience is how many months a firm has drilled horizontal wells in a specific county. I use months instead of well numbers so that firms which drill multiple wells in the same monthly only receive credit once. This is a possible under-estimate as firms who drill more in a given month likely learn more. However, I measure the impact of experience on well-level outcomes and it would be problematic to classify a better outcome from well number ten drilled in a given month as being due to the fact that it is the tenth well even though it was drilled at the same time as well number one from that month.

For each estimated learning (experimentation or experience) effect, I sort regions by quartiles as I did the knowledge sharing measures. Here, I show results for knowledge sharing $\times$ learning quartiles and repeat the baseline specification.

$$\begin{aligned} Y^j_{w,t} &= \alpha + \gamma^{ks_n}\hat{I}^j_t + \Gamma\left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage) \\ I^j_t &= \eta + \beta^{ks_n}\mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma\left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage) \end{aligned}$$

The first stage I^j_t here is the firm-county level investment activity. In addition to sorting on knowledge sharing regions, this specification also sorts within each knowledge sharing bucket on learning effectiveness quartiles. Tables B.3 and B.4 show sample results from Q - skill knowledge sharing. Note that I only use firm-county investments in the first stage since the purpose is to study the possible impact from learning propensity. The results are shown for the effect of experimentation over interval lengths on Q, baseline production efficiency. I also only show results for the highest knowledge sharing regions. Table B.4 contains the second stage result for the highest knowledge sharing bucket only. In those regions, the lowest learning quartile has the strongest effect from investment. Within high knowledge sharing regions, the highest learning regions are not significantly impacted by a firm's investment level. This suggests that learning is not likely to be complementary here and that own firm learning is not driving the baseline knowledge sharing result shown in the paper. Table B.3 shows the same results but for regions where learning propensity is measured by experience. Here, there is some suggestion that experience is positively associated with the knowledge spillover effect. The only region where the second stage coefficient is positive is in the third learning bucket.

As described, the learning variable is also estimated in much the way that the spatial lag model empirically estimates knowledge sharing propensity. Implicitly, this makes the same assumption that there are regions where experimentation or experience is more likely to be useful. Much of the motivation behind knowledge sharing propensity also apply here. Geological complexity likely makes it difficult for a firm to carry through knowledge from one well to the next. However, learning is more likely to be the result of firm action than the sharing of no-excludable knowledge.

B.7 Standard Error in the spatial lag model

To sort firms using the spatial lag model, I only keep counties where the $\beta^{g,H}$ coefficient on skill is statistically significant. Note that by sorting regions using the results from the spatial lag model as an ordinal ranking, I avoid problems with generated regressions. The sampling variance from the spatial lag model does not directly enter the main specification. However, the spirit of the problem may still be of concern: there is sampling variance in the spatial lag model. In table B.5, I show results from the baseline specification for Q and production efficiency skill based knowledge sharing. The regression is:

$$Y^j_{w,t} = \alpha + \gamma^{ks_n} \hat{I}^j_t + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage) \tag{B.7}$$

$$I^j_t = \eta + \beta^{ks_n} \mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage) \tag{B.8}$$

However, I form the knowledge sharing quartiles using $\beta^{g,H} - 2 * se$, the lower end of the confidence interval. I include all counties in these results, even those where the skill coefficient was not significant. The result for Q-skill based knowledge spillover illustrates the problem. The results are actually decreasing in knowledge spillover buckets. Future work may consider different ways of using the first stage network estimates and incorporate the sampling variance. Importantly, the empirical specification is consistent within the model. The empirical test conducts a stress test of the network *as it is estimated*. This problem reflects work that is beyond the scope of this paper. Namely, how should the netowrk be estimated in the first place?

B.8 Model Solution Method

I provide a brief overview of how the model is solved. Interested readers should refer to the theoretical companion cited in the main paper. There are four main problems that need to be solved. The main solution methodology is based on work done by [7].

1. The investment problem without technology adjustment
2. The technology adjustment problem
3. The equilibrium variables
4. The transition dynamics

The investment problem without adjustment. To begin, the solution to the HJB without adjustment needs to be solved as it serves as the starting point for the stopping problem. I discretize the value function over an equidistant grid and solve the problem using finite difference methods. I use an implicit, upwind scheme.

The technology adjustment problem. With the optimal V from the problem without adjustment in hand, it serves as the initial guess for the stopping time problem. I re-cast the problem as a linear complementarity problem which has standard solution algorithms. The formulation is given by,

$$min\left\{\beta V(w) - HJB_x(V), V - V^*(V)\right\} = 0 \tag{B.9}$$

The iteration methodology here is similar to that of the investment problem.I find the value of V^{n+1} that minimizes the problem in B.9 until it converges. The problem of choosing the right value for $V^*(V)$ is not standard. For example, in the canonical stopping time problem, a firm can choose to exit and recover the scrap value. In that case, $V^*(V)$ would just be the scrap value which can be constant and exogenously defined or some proportion of the firm's value at the time of exit. In the first case, the more appropriate notation would be V^* whereas the second case would be $V^*(V)$ to account for the fact that the scrap value varies with firm value V. I point out these notation differences to provide intuition for the problem. In this case, the exit value is varying with firm value.

Recall that to use finite-difference methods, the entire state space is discretized so that the value function solution can be approximated. In the case of γ, I do not treat the grid as approximations to the continous variable γ. Rather, I impose that firms can only be at distinct values of γ. Practically, this does not pose large problems for the investment solution. I do not need to account for $\frac{\partial V}{\partial \gamma}$ since it does not change. In the LCP problem, the technology change is set up as discrete jumps anyway so it is intuitive to interpret firms as being divided into discrete buckets of technology specialization.

Let V^n be the continuation value at iteration n and let g denote different realizations of values in the γ-grid. For each value of w, I interpolate $V^n(k, A^{new}, \gamma')$ at each potential value of γ'[8]. Note that the value of γ' depends also on where the firm is in the $(k, A^{new}(h))$ state space. In other words, firms of different size and skill will have different potential values for new technology rate γ'. For each point in the state space (k, A^{new}, γ) I pick the γ' that corresponds to the maximum $V^*(k, A^{new}, \gamma')$. Then, that value is $V^*(V)$ which is used in the larger LCP iteration solving in equation B.9.

Equilibrium To find the equilibrium solution, start with a guess of the equilibrium objects, $K\Gamma$ and p_o, p_n. These are assumed to be the stationary values. I then solve for the stationary distribution using the discretized version of the kolmogorov-forward equation at the stationary solution so that $\frac{\partial g}{\partial t} = 0$ for density function g. With the distribution in hand, I update aggregate capital which also gives me the aggregate productivity. I use a bisection method to solve for p_o, p_n. Note that in equilibrium, all three of these values have to clear the market simultaneously.

To solve for the dynamic transitions, I start the economy at a random distribution and an entire time series guess for the eqilibrium objects. The distribution has the characteristic that only a small handful of firms are involved in using the new technology, everyone else uses the old. Then, I use the time-dependent counterpart to the HJB and KFE to solve the problem forward for the whole time series. I then update the entire vector of aggregate capital, p_o, p_n using the same methods as in the stationary equilibrium solution. This is

8. For the version without debt or exits, this is fairly straightforward. With debt which can take on any value within a range, I

repeated until entire vector of dynamic equilibrium objects converges.

B.9 Appendix Figures & Tables

Preliminaries

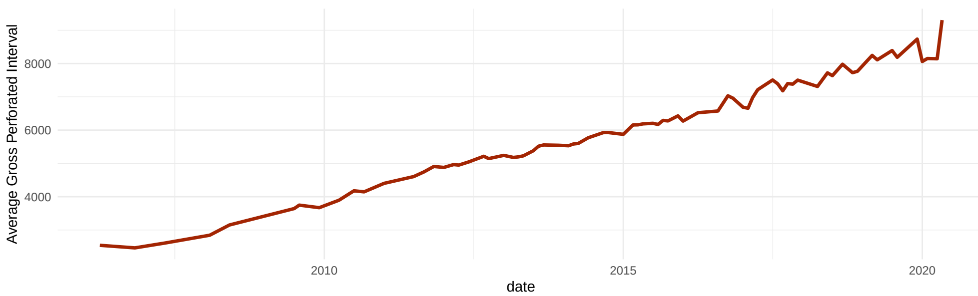

Figure B.1: Average horizontal interval lengths over time

The figure shows the average perforated interval lengths in feet over time. In the horizontal well figure from B.2, this value corresponds to the part of the well that is horizontal and perforated. These are for horizontally drilled wells only. The data is pulled from the permits firms submit when beginning the drilling process. Optimal interval lengths will vary across geographies. On average, horizontal interval lengths have been growing and represent an important aspect of the technology growth in fracking.

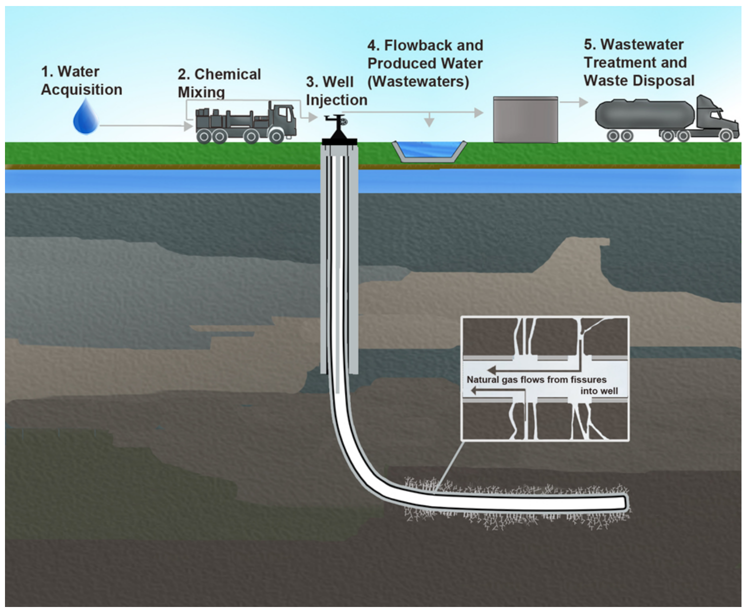

Figure B.2: Horizontally drilled wells example

The figure is pulled from a report by the Environmental Protection Agency. It illustrates how the horizontally fracked well technology functions. The horizontal intervals are perforated and fracking fluid is discharged into shale rock at high temperatures and the resource is released.

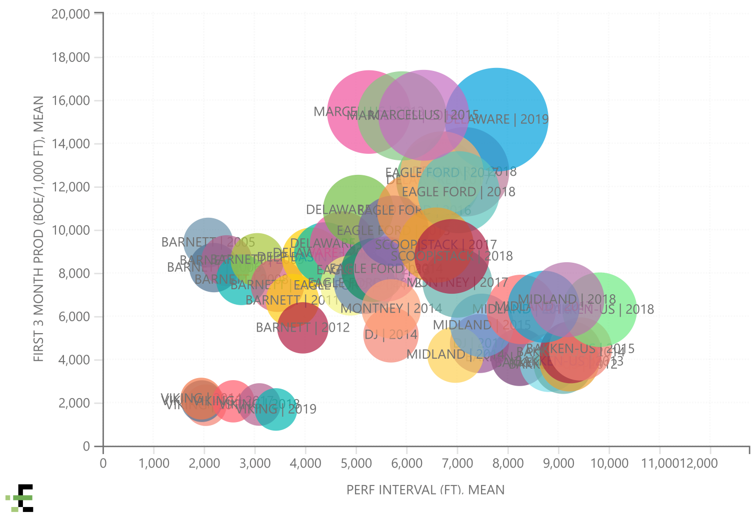

Figure B.3: Efficacy of longer horizontal wells

The figure is pulled from Enverus analytics. The bubbles reflect oil producing geological formations by production size. On the x-axis is the average perforated interval for wells in those formations while the right axis contains the corresponding average IP rates for those wells. The figure illustrates the heterogeneity in the efficacy of more complex (longer interval length) wells by geography. The optimal well length is geography-specific which implies that learning and information is useful across counties.

B.9.1 Results

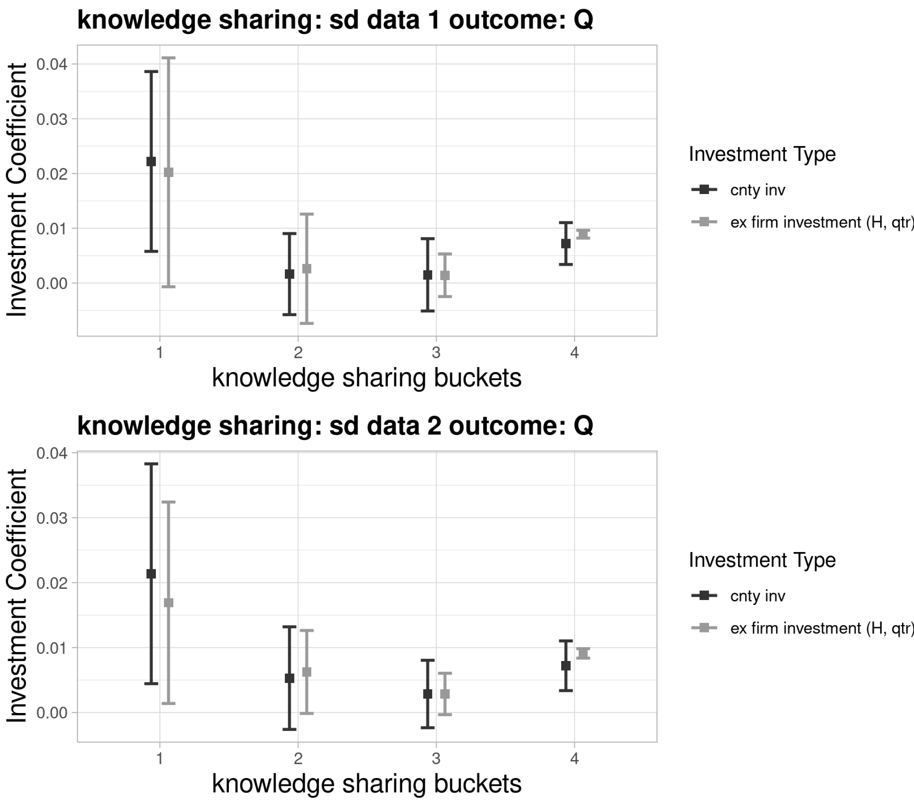

Figure B.4: Standard Deviation Tests from extra data sets 1 and 2

The figure plots coefficients γ^{ks_n} from well-level regression:

$$Y^j_{w,t} = \alpha + \gamma^{ks_n} \hat{I}^j_t + \Gamma \left[\mathbf{X}_{g,t} \ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage)$$
$$I^j_t = \eta + \beta^{ks_n} \mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma \left[\mathbf{X}_{g,t} \ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage)$$

The network mode used to sort these buckets is estimated using $\hat{Q}$ as the outcome measure but the standard deviation of the additional variables from data availability matrix A is used as the indirect effect variable. Two different sets of variables are used, the first one relates to the fracking program of horizontally fracked wells while the second is a general set of well design variables such as casing pressure. Both of these are detailed in the descriptions above. These values were not placed through the spatial lag model. I simply sort counties based on their average standard deviation measure. The outcome variable of interest is Q. The measures are divided into $n = 1, 2, 3, 4$ buckets. t denotes the month that the well was drilled.

The values in gray show, $j = ex - firm$, investments by all firms in a county excluding the firm that drilled well w. The black shows, $j = county$, county level investments. These are separate regressions conducted at each j, ks_n level. The county level specification clusters standard errors at the county level.

Dependent Variable: Q

Q-skill Knowledge sharing								
	(1) Firm-county investments (-1 qtr)	(2) Knowledge Sharing 1	(3) Firm-county investments (-1 qtr)	(4) Knowledge Sharing 2	(5) Firm-county investments (-1 qtr)	(6) Knowledge Sharing 3	(7) Firm-county investments (-1 qtr)	(8) Knowledge Sharing 4
$\mathbf{1}_{\{High\ Price\ Regime\}}$	3.28***		0.21		7.28***		23.7***	
	(0.64)		(0.51)		(0.70)		(0.83)	
Firm -county investments (-1 qtr)		0.053***		1.87		0.061***		0.010***
		(0.016)		(4.60)		(0.0090)		(0.0017)
Constant	0.98	-1.17	9.44	-21.9	-15.7*	-5.30***	89.4***	-6.98***
	(5.51)	(0.62)	(5.63)	(48.2)	(7.07)	(0.93)	(9.78)	(0.62)
N	9830	9830	7924	7924	6584	6584	8799	8799
Underidentification		26.9		0.17		99.2		534.4
Weak identification		26.7		0.17		108.7		814.0
Model Diff		79.3		8.13		313.3		0.98
Diff pvalue		5.2e-19		0.0044		4.1e-70		0.32

Table B.1: Own Firm Investments Effects

The Table shows results from:

$$\begin{aligned} Y_{w,t}^{j} &= \alpha + \gamma^{ks_n} \hat{I}_t^j + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage) \\ I_t^j &= \eta + \beta^{ks_n} \mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage) \end{aligned}$$

in table format. The first stage investment results are shown in the odd-numbered columns while the second stage results for each knowledge sharing quartile is shown in increasing order in the even-numbered columns. The investment measure is $Firm - countyinvestment$. The knowledge sharing value, ks, is Q-skill. These are estimated in the spatial lag model by regressing $\mathbf{Y} = Q$, on a weighted vector of nearby firm skill. The outcome variable is baseline production Q. The measures are divided into $n = 1, 2, 3, 4$ buckets.

Model Diff in the last row of each panel shows F-test values for Wald tests of statistical differences between each model while the last row shows the corresponding p-values. The tests are increasing so the first stage shows the results for $\gamma^{ks_1} = \gamma^{ks_2}$ and the last column shows $\gamma^{ks4} = \gamma^{ks_1}$.

Dependent Variable: Q

Production Efficiency-skill Knowledge sharing	(1) Ex-firm investments (ex-large)	(2) Knowledge Sharing 1	(3) Ex-firm investments (ex-large)	(4) Knowledge Sharing 2	(5) Ex-firm investments (ex-large)	(6) Knowledge Sharing 3	(7) Ex-firm investments (ex-large)	(8) Knowledge Sharing 4
$\mathbf{1}_{\{High\ Price\ Regime\}}$	47.2***		21.4***		13.7***		11.4***	
	(2.48)		(1.07)		(0.49)		(0.52)	
Ex-firm Investments (ex-large)		0.0053**		0.0099***		0.030***		0.037***
		(0.0019)		(0.0029)		(0.0024)		(0.0046)
Constant	-286.4***	-1.26	-138.4***	-0.79	64.3***	-11.0***	-38.4***	-4.90***
	(20.9)	(0.73)	(7.81)	(0.64)	(6.61)	(0.61)	(4.12)	(0.55)
N	1692	1692	5649	5649	14259	14259	5120	5120
Underidentification		256.0		325.0		746.2		422.9
Weak identification		361.7		399.1		775.5		483.8
Model Diff		2.91		64.0		0.53		16.2
Diff pvalue		0.088		1.2e-15		0.47		0.000058

Table B.2: Distribution tests: Dispersed Firm Sample

The Table shows results from:

$$Y_{w,t}^{j} = \alpha + \gamma^{ks_n}\hat{I}_t^j + \Gamma\left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage)$$
$$I_t^j = \eta + \beta^{ks_n}\mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma\left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage)$$

in table format. The first stage investment results are shown in the odd-numbered columns while the second stage results for each knowledge sharing quartile is shown in increasing order in the even-numbered columns. The investment measure is ex-firm investment which takes total county level investments minus firm-county level investments. The knowledge sharing value, ks, is production efficiency-skill. These are estimated in the spatial lag model by regressing $\mathbf{Y} = production\ efficiency$ on a weighted vector of nearby firm skill. Production efficiency measures oil output weighted by the well complexity input. The outcome variable is baseline production Q. The measures are divided into $n = 1, 2, 3, 4$ buckets.

The regression is for a subset of the sample in each county. Any firm that represents more than 50% of the horizontal production in a county is removed. They are not included in the set of firms who could potentially be impacted by knowledge spillovers and their investment levels are not included when measuring the first stage investment variable.

Model Diff in the last row of each panel shows F-test values for Wald tests of statistical differences between each model while the last row shows the corresponding p-values. The tests are increasing so the first stage shows the results for $\gamma^{ks_1} = \gamma^{ks_2}$ and the last column shows $\gamma^{ks4} = \gamma^{ks_1}$.

Second Stage Results: Q dependent variable

Interval Experimentation Learning on Q outcome

Q -Skill Knowledge Sharing

	(1) Learning 1	(2) Learning 2	(3) Learning 3	(4) Learning 4
Firm-County Investment	0.15***	-0.025***	0.0068**	-0.016
	(0.032)	(0.0066)	(0.0021)	(0.0099)
Constant	-4.78	4.12	-1.80	-10.6***
	(2.89)	(3.03)	(1.45)	(0.91)
N	1586	3018	2379	7020
Underidentification	27.4	41.1	193.6	36.7
Weak identification	28.4	46.6	193.0	38.1

Table B.3: Learning through experimentation Effects

The tables shows second stage results from,

$$Y_{w,t}^{j} = \alpha + \gamma^{ks_n} \hat{I}_t^j + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage)$$
$$I_t^j = \eta + \beta^{ks_n} \mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage)$$

The investment level which is instrumented here is the firm-county level investment. The specification captures the effect from making larger investments within a county on baseline productivity. The reported results first sorts on the same knowledge sharing quartiles as in the baseline but then further sorts *within* each knowledge sharing quartile by estimates of how useful experimentation is in each county. The results only show the highest knowledge sharing quartiles. Within these high network area counties, learning through experimentation is not effective.

Second Stage Results: Q dependent variable

Interval Experience on Q outcome				
Q-Skill Knowledge Sharing				
	(1) Learning 1	(2) Learning 2	(3) Learning 3	(4) Learning 4
Firm- County Investment	-0.0023	0.0067	0.014***	-0.011
	(0.023)	(0.0062)	(0.0022)	(0.0060)
Constant	26.5**	-7.81*	-10.2***	-9.03***
	(8.99)	(3.44)	(1.08)	(0.78)
N	192	2821	3046	7944
Underidentification	35.1	35.3	309.2	83.6
Weak identification	42.3	37.8	273.2	91.5

Table B.4: Learning through Experience Effects

The tables shows second stage results from,

$$\begin{aligned} Y^j_{w,t} &= \alpha + \gamma^{ks_n} \hat{I}^j_t + \Gamma \left[\mathbf{X}_{g,t} \ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage) \\ I^j_t &= \eta + \beta^{ks_n} \mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma \left[\mathbf{X}_{g,t} \ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage) \end{aligned}$$

The investment level which is instrumented here is the firm-county level investment. The specification captures the effect from making larger investments within a county on baseline productivity. The reported results first sorts on the same knowledge sharing quartiles as in the baseline but then further sorts *within* each knowledge sharing quartile by estimates of how useful experience is in each county. The results only show the highest knowledge sharing quartiles. The results are mixed. Within these high network area counties, learning through growing experience is not effective in the highest learning buckets but the results are positive in the third bucket.

Dependent Variable: Q

Q-skill based knowledge sharing, lower Confidence Interval Sort

	(1) Ex-firm Investments	(2) Knowledge sharing 1	(3) Ex-firm Investments	(4) Knowledge sharing 2	(5) Ex-firm Investments	(6) Knowledge sharing 3	(7) Ex-firm Investments	(8) Knowledge sharing 4
$\mathbf{1}_{\{High\ Oil\ Price\}}$	20.6***		22.1***		85.8***		46.7***	
	(2.10)		(1.44)		(1.17)		(0.98)	
Ex-firm Investments		0.021***		0.014***		0.0044***		0.0064***
		(0.0041)		(0.0019)		(0.00027)		(0.00062)
Constant	-64.5***	-3.09***	-104.5***	-3.00***	-102.7***	-3.94***	-76.5***	-6.85***
	(12.0)	(0.53)	(11.3)	(0.57)	(11.1)	(0.27)	(8.61)	(0.35)
N	6115	6115	6864	6864	26539	26539	18433	18433
Underidentification		95.8		231.1		3869.9		1754.6
Weak identification		96.1		237.1		5396.0		2268.4
Model Diff		2.92		28.2		235.9		0.41
Diff pvalue		0.087		0.00000011		3.0e-53		0.52

Table B.5: Lower $\hat{\beta}$ Confidence Interval Tests

The tables shows results from,

$$Y_{w,t}^{j} = \alpha + \gamma^{ks_n} \hat{I}_t^j + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \epsilon_{w,t} \quad (second\ stage)$$
$$I_t^j = \eta + \beta^{ks_n} \mathbf{1}_{\{High\ Oil\ Price\}} + \Gamma \left[\mathbf{X}_{g,t}\ \mathbf{X}_{i,t}\right]' + \nu_{g,t} \quad (first\ stage)$$

Y is the Q baseline productivity rate for well w. This only shows up once for each well at time t when the well was drilled. This experiment replicates a version of the baseline specification. However, rather than sorting counties based on knowledge sharing estimates from the spatial panel model where $\hat{\beta}$ is statistically significant, I sort based on $\hat{\beta} - 2 * se$, the lower end of the confidence interval.