THE SOCIOECONOMIC CONSEQUENCES OF THE U.S. FRACKING BOOM

BY

MINHONG XU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Agricultural and Applied Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Assistant Professor Yilan Xu, Chair
Professor Madhu Khanna
Associate Professor Sandy Dall'Erba
Professor Daniel McMillen

ABSTRACT

The central theme of this dissertation is to understand the socioeconomic consequences associated with the technological breakthroughs in the U.S. energy landscape. The dissertation combines econometric methods with spatial techniques to identify the externalities in water and energy transport arising from the recent shale energy boom. The three essays focus on various aspects of influences on real estate market, wastewater management, and highway traffic safety, with each contributing to a broad picture of the overall costs and benefits of the fracking revolution. All the analyses indicate that despite the enormous economic benefits the negative externalities associated with pipeline and truck transport have resulted in additional social and environmental costs. Specifically, **in the first essay,** we focus on the extent to which the existence of oil and gas pipelines and the related incidents elevated mortgage lenders' risk perceptions and affected their lending and securitization decisions. We show a permanently lower origination rate by 1.9% in the pipeline-present areas compared to the pipeline-free areas, which was further enlarged by 1.8% whenever pipeline incidents happened. Moreover, lenders' risk management strategies differed by borrowers' income and evolved with the tightening of the securitization market. **In the second essay,** we focus on analyzing how unconventional drilling and production techniques could affect oil and gas wastewater production. We find that horizontal wells drilled in shale regions after 2010 produced an increasingly larger amount of wastewater at the initial stage of production compared to vertical wells with the difference decreasing over production age regardless of the cohort years. However, we find that unconventional wells have a lower lifespan cumulative wastewater and a higher production efficiency regarding the wastewater-to-energy ratio. **In the third essay,** we examine the effects of the fracking-related trucking on fatal crashes using evidence from North Dakota. A Poisson estimation shows that an additional post-fracking well within six miles of a road segment not only increased the count of fatal crashes involving large trucks by 7.5% but also intensified the crash severity. Moreover, wells' post-fracking transport increased the incidence of daytime crashes during both rush and non-rush hours but not nighttime crashes. Finally, evidence suggests that the increased accidents were due to a higher traffic volume rather than a higher accident rate or risky driving behaviors.

ACKNOWLEDGEMENTS

I would like to gratefully and sincerely thank my advisor Dr. Yilan Xu for her valuable guidance and support throughout the course of my doctoral work. Without her encouragement and instruction, I would not have been possible to complete this dissertation.

I appreciate the thoughtful and invaluable suggestions provided by Dr. Amy Ando, Dr. Katherine Baylis, Dr. Peter Christensen, Dr. Sandy Dall'Erba, Dr. Philip Garcia, Dr. Madhu Khanna, Dr. Daniel McMillen, and Dr. Erica Myers, and Dr. Charles Nelson. I am grateful for their guidance during my five years at the University of Illinois as a graduate student.

I would like to express my gratitude to the staffs and the peers in the ACE department. The valuable support provided by them have made my life much easier in the past five years.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The U.S. oil and gas production has undergone substantial increases since the late 2000s. From 2007 to 2017, crude oil production nearly doubled while natural gas production increased by 37%.[1] In particular, tight oil production made up 54% of the U.S total in 2017 (U.S. Energy Information Administration (EIA), 2018), while shale gas gross withdrawals by 2016 had been more than seven times higher than its baseline level in 2007 with its share in total natural gas withdrawals growing from 8% to over 50% over the period.[2] The successful extraction of oil and gas from shale rocks relies on two principle technologies - high-volume hydraulic fracturing ("fracking") and horizontal drilling. While the innovative combination of the two technologies has triggered the large-scale development of unconventional oil and gas resources nationwide, concerns about the associated environmental and social costs have caused states such as New York and Maryland and many local governments to ban fracking (Burnett and Weber, 2014). This dissertation aims to understand the effects of the technological innovations and the associated infrastructure boom on real estate market, water resources and road safety.

Although the United States has experienced multiple oil and gas booms in the past century, energy production has never been as controversial as today. Compared with conventional methods,

---

[1] Data source: U.S. Energy Information Administration monthly crude oil and natural gas production data. See more at https://www.eia.gov/petroleum/production/

[2] Data source: U.S. Energy Information Administration annual shale gas production data. See more at https://www.eia.gov/dnav/ng/ng_prod_shalegas_s1_a.htm

the fracking technology employed to extract shale oil and gas are more resource-intensive, as a well may require up to six million gallons of water to crack open shale rocks (U.S. Environmental Protection Agency, 2016). The intensive water requirement may result in regional water stress and competition with water users in the agriculture and municipal sectors (Jaffe, 2014; Freyman, 2014; Gallegos et al., 2015; Horner et al., 2016). Moreover, as a substantial portion of the injected water eventually returns to the surface, the chemicals contained in wastewater may lead to environmental pollution and health consequences if not appropriately treated (Veil et al., 2004).

Besides the water-intensive and disruptive nature of the technology itself, the associated energy and water transport activities also affect the overall quality of life as thousands of wells are drilled on people's private land onshore in the lower 48 states. Since the shale energy boom, both the volume and the mileage of energy pipeline transport has increased significantly, while truck use also grew in quantity and intensity in response to the increased transport demand for drilling equipment and fracking-related water. In the meantime, transport-related incidents such as pipeline spills, leaks, fire and explosion, and fatal traffic crashes soared subsequently.[3] These incidents not only resulted in life and property damage and loss but also generated substantial environmental and social cost due to inefficient risk management and resource allocation.

Understanding the net gains and losses of the U.S. shale energy boom is essential for both policymakers and the public to design and support appropriate policies. Unfortunately, due to the fast growth of drilling activities and the absence of experience, local governments usually lack effective methods to address the emerging issues, while the public perception is easily swayed by

[3] See the two reports on the website of Fractracker Alliance in 2014 and 2016 at https://www.fractracker.org/2014/09/truck-counts/ and https://www.fractracker.org/2016/11/updated-pipeline-incidents/

anecdotal and unrepresentative evidence. Failure to identify the pattern of unconventional energy production and the associated environmental risks would exacerbate the existing issues. However, simply banning fracking may lead to loss of royalties and tax revenue and higher energy prices borne by consumers (Burnett and Weber, 2014). As shale oil and gas production is projected to grow in both share and absolute volume (EIA, 2018), it is imperative to establish the causal inference underlying the controversies and shed light on risk management strategies and policy responses.

In the past decade, both industry-funded researchers and the academic community has been active in investigating the local consequences of shale energy production. The former applies the Input-Output model to estimate the economic impacts of oil and gas extraction on gross revenues, employment, and tax revenues. A non-exclusive list of examples includes the Center for Business and Economic Research (2008), Considine et al. (2009), Scott and Huang (2009), Weinstein and Clower (2009), Considine, Watson, and Blumsack (2010), Considine, Watson, and Considine (2011). Moreover, exploiting the technological shift in fracking and horizontal drilling interacted with local geology as an instrumental variable, researchers estimate the economic gains of the fracking revolution on income and employment (Weber, 2012; Feyrer, Mansur, and Sacerdote, 2017). Researchers also present evidence against the hypothesis of "Natural Resource Curse" from the recent shale energy boom (Allcott and Keniston, 2017; Weber, 2014; Fetzer, 2014; Maniloff and Mastromonaco, 2014; Brown, 2014).

Regarding the environmental consequences of the fracking boom, the focus of the current discussion lies in the water quality issue and the associated pollution and health concerns. While engineering studies directly analyze the chemicals of fracking-related wastewater (Osborn et al. 2011; Warner et al. 2013a, 2013b, 2014; Darrah et al. 2014; Harkness et al., 2015), economists estimate the housing prices depreciation to quantify people's perceived environmental risks of drilling and fracking activities (Gopalakrishnan and Klaiber, 2013; Weber, Burnett, and Xiarchos, 2014; Muehlenbachs et al., 2015). Researchers also show evidence of deterioration in the quality of life or total amenities by estimating people's willingness-to-pay for the fracking-induced loss in local amenities (Bartik, Currie, Greenstone and Knittel, 2016). A very recent study by Currie, Greenstone, and Meckel (2017) demonstrates the potential health impacts of fracking through a greater incidence of low-birth-weight babies identified near the fracking sites.

Despite the abundant literature investigating the economic benefits and the environmental costs directly related to drilling and fracking, little attention has been paid to the externalities stemming from the transport activities in oil and gas extraction. Although these activities occur upstream or downstream of energy production, the unique technology and production pattern of shale energy extraction determines that these activities could lead to severe local consequences as well. This dissertation consists of three stand-alone essays to test and understand the externalities arising from water and energy transport during the shale energy boom.

The first essay is entitled "**Environmental Hazards and Mortgage Credit Risk: Evidence from Texas Pipeline Incidents.**" This essay attempts to examine the impacts of pipeline hazards on home mortgage lenders' lending and securitization decisions in the real estate market. Specifically, we use a difference-in-difference model to test the permanent differences in mortgage credit access between the pipeline-present and the pipeline-free Census tracts and the additional differences when pipeline incidents occurred. Built upon the literature on the effects of natural and manmade disasters on people's risk perceptions of environmental hazards, this essay focuses on the perspective of home mortgage lenders who are more sensitive to information and own richer resources to evaluate and manage the potential risks. This essay further demonstrates how lenders' risk management strategies differed by borrowers' income and evolved with the tightening of the securitization market. The empirical evidence obtained confirms the expected concerns of lenders about pipeline risks, which specifically expands the scarce literature concerning mortgage lenders' risk perceptions related to environmental contamination.

The second essay is entitled "**Do Unconventional Oil and Gas Wells Generate More Wastewater? A Lifespan Perspective.**" While the question has been discussed in engineering studies, the available evidence is limited in two critical dimensions. The first limitation is that few researchers consider the effects of wells' age or cohort year, which are essential to understand how the production pattern varies over a well's lifespan and evolves across time. The second limitation is that most previous studies are regional case studies without considering the heterogeneity in geology and well configuration. This essay provides the first comprehensive and accurate quantification of wastewater generation using nationwide well-level energy and wastewater production data. The difference-in-difference-in-differences analysis shows that unconventional wells are more efficient concerning both lifetime volume and production efficiency despite the higher initial wastewater production. The findings challenge the perceptions that unconventional techniques lead to more wastewater while pointing out where the real concerns lie in. The cyclical pattern of wastewater production identified in this essay are informative for policymakers to design more effective regulations to manage wastewater in shale development.

The third essay is entitled "**Fraccidents: The Impact of Fracking on Road Traffic Deaths.**" Built upon the second chapter, this essay further examines how the unique wastewater production pattern related with fracking has affected truck transport and intensified the accident externality. Exploiting the monthly variations in a well's spud and completion dates, this essay estimates changes in fatal crash count due to the exogenous timing of well operations within a buffer area of a road segment. A fixed-effect Poisson estimation shows that an additional post-fracking well increased the crash count and the accident severity. Moreover, wells' post-fracking transport increased the incidence of daytime crashes during both rush and non-rush hours but not nighttime crashes. Evidence suggests that the potential mechanism that underlies the phenomenon was through a higher traffic volume rather than a higher accident rate or risky driving behaviors. This essay not only reveals the hidden costs of truck transport borne by the energy boomtowns but also sheds new light on risk management strategies for operators and local governments to manage the threats of fracking-induced trucking to road safety.

This dissertation contributes to the debate about shale energy development by providing evidence of its socioeconomic consequences in home mortgage lending, wastewater management, and traffic safety. The results will be of interest not only to policymakers interested in designing management practices to alleviate the socioeconomic costs of energy development, but also to economists interested in finding what drove the rising number of accidents related to fracking. The findings of this dissertation may provide new insight as to whether fracking is to blame for the growing health and safety concern and whether further regulatory efforts should be implemented to limit the technology in future energy development.

1.1 References

Allcott, H. and Keniston, D., 2017. Dutch disease or agglomeration? The local economic effects of natural resource booms in modern America. *The Review of Economic Studies*, *85*(2), pp.695-731.

Bartik, A.W., Currie, J., Greenstone, M. and Knittel, C.R., 2016. The local economic and welfare consequences of hydraulic fracturing.

Brown, J.P., 2014. Production of natural gas from shale in local economies: a resource blessing or curse? *Economic Review-Federal Reserve Bank of Kansas City*, p.5.

Burnett, J.W. and Weber, J.G., 2014. Is the Natural Gas Revolution all its Fracked Up to Be for Local Economies? *Choices*, *29*(4).

Center for Business and Economic Research, 2008. Projecting the Economic Impact of the Fayetteville Shale Play for 2008-2012. http://www.aogc.state.ar.us/sales/FayettevilleShaleEconomicImpactStudy2008.pdf

Considine, T.J., Watson, R. and Blumsack, S., 2010. The economic impacts of the Pennsylvania Marcellus shale natural gas play: an update. *The Pennsylvania State University, Department of Energy and Mineral Engineering*.

Considine, T.J., Watson, R.W. and Considine, N.B., 2011. The economic opportunities of shale energy development. *The Manhattan Institute, June 28*.

Considine, T., Watson, R., Entler, R. and Sparks, J., 2009. An emerging giant: Prospects and economic impacts of developing the Marcellus shale natural gas play. *University Park, PA: The Pennsylvania State University, Dept. of Energy and Mineral Engineering. August*.

Currie, J., Greenstone, M. and Meckel, K., 2017. Hydraulic fracturing and infant health: New evidence from Pennsylvania. *Science advances*, *3*(12), p.e1603021.

Darrah, T.H., Vengosh, A., Jackson, R.B., Warner, N.R. and Poreda, R.J., 2014. Noble gases identify the mechanisms of fugitive gas contamination in drinking-water wells overlying the Marcellus and Barnett Shales. *Proceedings of the National Academy of Sciences*, *111*(39), pp.14076-14081.

Feyrer, J., Mansur, E.T. and Sacerdote, B., 2017. Geographic dispersion of economic shocks: Evidence from the fracking revolution. *American Economic Review*, *107*(4), pp.1313-34.

Fetzer, T., 2014. Fracking growth.

Freyman, M., 2014. Hydraulic fracturing and water stress: water demand by the numbers. A Ceres Report.

Gallegos, T.J., Varela, B.A., Haines, S.S. and Engle, M.A., 2015. Hydraulic fracturing water use variability in the United States and potential environmental implications. *Water Resources Research*, *51*(7), pp.5839-5845.

Gopalakrishnan, S. and Klaiber, H.A., 2013. Is the shale energy boom a bust for nearby residents? Evidence from housing values in Pennsylvania. *American Journal of Agricultural Economics*, *96*(1), pp.43-66.

Harkness, J.S., Dwyer, G.S., Warner, N.R., Parker, K.M., Mitch, W.A. and Vengosh, A., 2015. Iodide, bromide, and ammonium in hydraulic fracturing and oil and gas wastewaters: environmental implications. *Environmental science & technology*, *49*(3), pp.1955-1963.

Horner, R.M., Harto, C.B., Jackson, R.B., Lowry, E.R., Brandt, A.R., Yeskoo, T.W., Murphy, D.J. and Clark, C.E., 2016. Water use and management in the Bakken shale oil play in North Dakota. *Environmental science & technology*, *50*(6), pp.3275-3282.

Jaffe, M, 2014. When drought occurs, fracking and farming collide. The Denver Post. February 9, 2014. http://www.denverpost.com/business/ci_25089583/when-droughtoccurs-fracking-and-farming-collide .

Maniloff, P. and Mastromonaco, R., 2014. The local economic impacts of hydraulic fracturing and determinants of dutch disease. *Colorado School of Mines, Division of Economics and Business Working Paper*, *8*.

Muehlenbachs, L., Spiller, E. and Timmins, C., 2015. The housing market impacts of shale gas development. *American Economic Review*, *105*(12), pp.3633-59.

Osborn, S.G., Vengosh, A., Warner, N.R. and Jackson, R.B., 2011. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *proceedings of the National Academy of Sciences*, *108*(20), pp.8172-8176.

Scott, L.C. and Huang, J., 2009. The economic impact of the Haynesville shale on the Louisiana economy in 2008. *Prepared for the Louisiana Department of Natural Resources*.

U.S. Energy Information Administration, 2018. Annual energy outlook 2018. Washington, DC: U.S. Office Energy Analysis, Department of Energy. https://www.eia.gov/outlooks/aeo/pdf/AEO2018_FINAL_PDF.pdf

U.S. Environmental Protection Agency. 2016. Hydraulic fracturing for oil and gas: Impacts from the hydraulic fracturing water cycle on drinking water resources in the United States. Washington, DC: Office of Research and Development. https://cfpub.epa.gov/ncea/hfstudy/recordisplay.cfm?deid=332990

Veil, J.A., Puder, M.G., Elcock, D. and Redweik Jr, R.J., 2004. *A white paper describing produced water from production of crude oil, natural gas, and coal bed methane* (No. ANL/EA/RP-112631). Argonne National Lab., IL (US).

Warner, N.R., Christie, C.A., Jackson, R.B. and Vengosh, A., 2013a. Impacts of shale gas wastewater disposal on water quality in western Pennsylvania. *Environmental science & technology*, *47*(20), pp.11849-11857.

Warner, N.R., Kresse, T.M., Hays, P.D., Down, A., Karr, J.D., Jackson, R.B. and Vengosh, A., 2013b. Geochemical and isotopic variations in shallow groundwater in areas of the Fayetteville Shale development, north-central Arkansas. *Applied Geochemistry*, *35*, pp.207-220.

Warner, N.R., Darrah, T.H., Jackson, R.B., Millot, R., Kloppmann, W. and Vengosh, A., 2014. New tracers identify hydraulic fracturing fluids and accidental releases from oil and gas operations. *Environmental science & technology*, *48*(21), pp.12552-12560.

Weber, J.G., 2012. The effects of a natural gas boom on employment and income in Colorado, Texas, and Wyoming. *Energy Economics*, *34*(5), pp.1580-1588.

Weber, J.G., 2014. A decade of natural gas development: The makings of a resource curse?. *Resource and Energy Economics*, *37*, pp.168-183.

Weber, J., Burnett, J. and Xiarchos, I., 2014. Shale gas development and housing values over a decade: evidence from the Barnett Shale.

Weinstein, B.L. and Clower, T.L., 2009. Potential economic and fiscal impacts from natural gas production in Broome County, New York. *Retrieved October*, *15*, p.2009.

CHAPTER 2

ENVIRONMENTAL HAZARDS AND MORTGAGE CREDIT RISK: EVIDENCE FROM TEXAS PIPELINE INCIDENTS

2.1 Introduction

The rapidly rising number of pipeline incidents in recent years has called public attention to pipeline hazards. In the past decade, there have been 6,313 pipeline leaks, spills, and other incidents, leading to 145 fatalities, 614 injuries and over 4,398 million dollars in property damage.[4] Although most pipelines are constructed in rural areas, a growing population is exposed to the hazards due to urban sprawl over the years. In addition to immediate harm to life and property, pipeline incidents generate a negative and persistent externality on housing prices (Simons 1999; Simons, Winson-Geideman and Mikelbank 2001; Hansen, Benson and Hagen 2006), and often result in environmental contamination in the neighborhoods involved (Islam *et al*. 2016; Matheny 2016; Medina 2016). Both housing prices drop and the potential environmental hazards pose a direct threat to the value of properties nearby, which increases the risks of mortgage lending to these neighborhoods. The purpose of this study is to examine how the existing pipeline infrastructure and the associated incidents impact mortgage lenders' perceptions of environmental risks and thus affect their lending and securitization decisions.

---

[4] Data Source: Authors' calculation using the data from the United States Department of Transportation Pipeline and Hazardous Materials Safety Administration. The original data are available at https://www.phmsa.dot.gov/pipeline/library/data-stats

Mortgage lenders view environmental hazards as top concerns among a myriad of external risk factors when evaluating potential lending profitability. Their reservations first stem from the potential impairment of the collateral, which may occur right after contaminants are discovered and persist in the long run due to environmental stigma damages (McCluskey and Rausser 2001 and 2003). Moreover, borrowers' repayment ability could also be weakened by the costs associated with required investigation and remediation (Davis and Levy 2012). The presence of environmental hazards thus increases the credit risk of the borrowers with limited financial viability. More importantly, under the current environmental laws, lenders themselves may be held liable for the entire cleanup of the contaminated sites in many circumstances. For example, the federal Superfund statute protects a foreclosing lender from the direct environmental liabilities if the lender makes commercially reasonable efforts to sell the property at the earliest practicable time. Despite the broad statutory protection, bright-line rules such as a fixed period for subsequent sale are absent for this safe harbor, which creates uncertainties for lenders seeking both the exemption protection and the optimal time to sell the foreclosed property in a challenging market (Ahrens and Langer 2008; Gracer and Leas 2008). Moreover, as petroleum-related contamination is not covered by the Superfund exemption protection, it poses a particular concern to lenders dealing with oil pipeline hazards, in which case lenders could be subject to the liability imposed by other state and federal laws.

Perceiving the potential environmental risks and uncertainties, lenders may avoid or at least limit the number of loans involving contaminated properties (Davis and Levy 2012). However, when the expansion of mortgage securitization enabled lenders to transfer the credit risk easily to investors in the secondary market, lenders may choose to originate the risky loans and package them for securitization (Jimenez and Saurina 2006; Dell'Ariccia, Igan and Laeven 2012; Keys, Seru and Vig 2012; Simkovic 2013). Such originate-to-distribute model prevailed before the Great Recession when the secondary market was full of private financial institutions (Berndt and Gupta 2009; Purnanandam 2011). Compared with private securitizers, the Government-Sponsored Enterprises (GSEs) have always been maintaining prudent guidelines towards the hazards from oil and gas storage and pipeline transportation that affect the properties securing the loans.[5] As private mortgage securitization virtually disappeared after the Great Recession, the secondary market was dominated by the GSEs, which was accompanied with a sharp tightening of underwriting standards during the same time (Simkovic 2013). Considering the marketability of the loans after the crisis, lenders could adjust their risk management strategies accordingly in response to the evolution of securitizers' guidelines.

In this study, we empirically investigate mortgage lenders' risk perceptions of pipelines using evidence from the 2005-2011 Home Mortgage Disclosure Act (HMDA) loans in Texas. Texas has the largest pipeline infrastructure in the United States with more than 439,771 miles of pipelines representing roughly one-sixth of the total pipeline mileage of the entire country (Railroad Commission of Texas 2017). We focus on the nonmetropolitan ("nonmetro") Census tracts of Texas where pipeline infrastructure is more concentrated and pipeline failures occur more frequently compared to the metropolitan ("metro") areas. A considerable number of these loans are from lenders that do not have any physical branches in the nonmetro counties where the collateral securing the loan is located, because lenders with branches only in the nonmetro areas are exempt from the HMDA data reporting requirements (Igan, Mishra and Tressel 2012).

[5] See more at https://www.fanniemae.com/content/guide/servicing/d1/1/01.html

Focusing on these loans enables us to observe lenders' risk-taking in a context of severe information asymmetry between lenders and borrowers (Xu and Zhang 2012).

We employ a difference-in-difference (DID) method to estimate the effects of pipeline infrastructure and the related incidents on mortgage lending. We define all the Census tracts where pipelines exist as the treatment group ("pipeline-present areas") and the Census tracts free of any pipeline infrastructure as the control group ("pipeline-free areas"). We exploit the exogenous variations in the timing, location, and the associated property damage to estimate the treatment effects of pipeline incidents. The DID model measures the effects of pipeline infrastructure by the permanent difference in lenders' credit decisions between the treatment and the control groups. It further captures the treatment effects of pipeline incidents by the relative change in the credit outcomes between the two groups across the years with versus without any incidents.

We find that an average mortgage loan in the pipeline-present areas was 1.9% less likely to be originated compared to that in the pipeline-free areas after controlling for borrowers' creditworthiness and demographic characteristics. We interpret the permanent group difference in credit availability from two aspects. First, we find that the denied loans in the treatment group were 3.2% more likely to be rejected due to insufficient collateral, which suggests that lenders perceived the properties in the pipeline-present areas as having lower collateral value. Second, we identify that the permanent difference in the origination rates emerged only among the low- to middle- income borrowers but not among the upper-income borrowers. The results indicate that the differential lending behaviors also resulted from lenders' concerns about borrowers' ability to repay a loan in case of remediation costs required for environmental contamination.

We find that the difference in the origination rates between the pipeline-present and the pipeline-free areas was further enlarged by 1.8% whenever an incident happened. We find no evidence showing that the severity of property damage of the incidents affected lenders' credit decisions, which implies that it could be the information shock rather than the actual damages of the incidents that increased lenders' risk perceptions. Another possibility could be that lenders were not informed of the real loss promptly because the investigation and public disclosure usually takes time. Meanwhile, we find that the average loan amount in the pipeline-present areas did not change after the incidents, which suggests that immediate property depreciation in the affected neighborhoods could be minimum. We interpret the declined origination rates as lenders' response to the uncertain future costs of bearing the direct environmental liabilities. We find that such temporary change in lending behaviors disappeared one year after, which is consistent with the findings that lenders' uncertainties would diminish as cleanup was completed and regulatory compliance was achieved (Jackson 2001). Again, the treatment effects of incidents emerged only among the low- to middle-income borrowers. We run a series of robustness checks to rule out the possibility that risky borrowers sorted to the pipeline-present areas after the incidents, as we find no simultaneous changes in borrowers' creditworthiness, demographic composition, or any confounding risk factors observed by lenders in the incidents-affected areas. Nor do we find any evidence of cumulative sorting based on the density of pipeline infrastructure and the frequency of pipeline incidents.

To investigate how the ease of securitization has affected lenders' risk management for different segments of the population, we split the baseline sample into subsamples by the years before, during, and after the late 2000s subprime mortgage crisis as well as by the income of borrowers. We find that before the Great Recession lenders managed pipeline hazards in a passive way by taking actions only after observing real incidents. Specifically, they denied the low- to middle-income borrowers and sold the originated loans from the upper-income borrowers to the

GSEs when incidents happened. Our findings suggest that the GSEs were able to screen the loans by the observable signals such as borrowers' income when they could not discern the localized environmental risks. We also find that after the crisis lenders started to lower the origination rates in the pipeline-present areas especially among the low- to middle-income borrowers. The findings suggest that the tightened securitization market during this period forced lenders to manage pipeline risks more aggressively by avoiding the properties exposed to pipeline hazards systematically.

This study complements and extends the studies on how natural or man-made disasters have affected people's risk perceptions of environmental hazards (Nelson 1981; Gamble and Downing 1982; Michaels and Smith 1990; Kolhase 1991; Kiel 1995; Carroll *et al*. 1996; Pope 2008; Naoi, Seko and Sumita 2009; McCoy and Walsh 2014). Existing studies focus on the information asymmetry in environmental risks between sellers and buyers in the housing market, covering a broad range of topics such as earthquakes, toxic waste sites, nuclear power plants, chemical plants, incinerators, landfills, flood zones, and wildfire hazards. Highly publicized incidents or public information disclosure, serving as exogenous information shock, could significantly increase buyers' marginal willingness-to-pay for the distance from the hazardous sites. Our study adds a novel perspective of the mortgage market where lenders are more sensitive to information and own richer resources to evaluate environmental conditions and manage the potential risks. The empirical evidence obtained in the study confirms the expected concerns of lenders about pipeline risks, which expands the scarce literature concerning mortgage lenders' risk perceptions related to environmental contamination (Healy and Healy 1992; Wilson and Alarcon 1997; Worzala and Kinnard 1997; Bond *et al*. 1998; Jackson 2001).

2.2 Empirical Method

We investigate mortgage lenders' potential reactions in both lending and securitization activities to measure their perceived risks of pipeline hazards. First, we expect lenders to decrease the probability of originating loans exposed to pipeline risks. Instead, they are more likely to deny the loans or conditionally approve the loans. In the latter case, the conditionally approved loans are subject to final verification such as initial inspection, property appraisal, and other stipulations before they move on to final approval.[6] Alternatively, lenders may transfer the credit risk in their portfolio to the secondary market by securitizing the mortgages subject to pipeline hazards. They could sell the originated loans to either the GSEs or private institutions that purchase and package loans for sale to investors (Simkovic 2013). In summary, we investigate four potential outcomes for each loan, including the likelihoods of being originated and denied, the probability of being sold to any purchasers in the secondary market, and the particular prospect of being sold to the GSEs. In the last two circumstances, we focus on the loans that have already been originated by lenders.

We treat the existence of pipeline infrastructure as a potential hazard and thus define all the Census tracts with the presence of pipeline infrastructure as the treatment group. Then we define the remaining Census tracts where no pipelines have been built as the control group. We start the analysis from estimating the following linear probability model:

(1) $$y_{ijt} = \beta_0 + \beta_1 pipe_j + \beta_2 pipe_j * incid_{jt} + \gamma \mathbf{x}_{ijt} + \varphi \mathbf{z}_{jt} + L_i + C_j + Y_t + \varepsilon_{ijt}$$

where i indexes individual loan, j indexes Census tract, and t indexes year. The dependent variable y_{ijt} is the discrete credit outcome for loan i, the treatment indicator $pipe_j$ denotes

[6] See more at https://www.ffiec.gov/hmda/faqreg.htm#action

whether any pipeline goes through Census tract j, $incid_{jt}$ indicates whether any pipeline incident happens in Census tract j of year t, and ε_{ijt} is the individual-specific error term. This model controls for a vector of individual-specific covariates, $\mathbf{x}_{ijt}$, a vector of Census tract-level characteristics, $\mathbf{z}_{jt}$, lender fixed effects, L_i, county fixed effects, C_j, and year fixed effects, Y_t. Thus, our findings are not driven by different business models of lenders, any time-invariant county-level characteristics, or any common economic shocks to the whole market in a given year. The coefficient β_1 captures the average difference in the credit outcomes of individual loans between the treatment and the control groups, which is referred to as the permanent effects of pipeline infrastructure. The coefficient β_2 measures any additional difference across the two groups whenever pipeline incidents occur, which is referred to as the incident effects.

2.3 Data

2.3.1 Home Mortgage Data

We get access to the information regarding home mortgage lending activities from the Home Mortgage Disclosure Act (HMDA) data, which are maintained by the Federal Financial Institutions Examination Council (FFIEC). The HMDA data cover both depository institutions (banks, savings associations, and credit unions) and non-depository institutions (for-profit mortgage lending institutions), as long as they meet the reporting criteria, such as whether the total assets of the institution exceeded the coverage threshold and whether the institution had a branch or office in a metropolitan statistical area (MSA) on the preceding December 31.[7] Berkovec and Zorn (1996) estimate that the lenders covered by the HMDA constitute roughly 80% of the total

[7] For more information see https://www.ffiec.gov/hmda/reporter.htm

U.S. mortgage originations. These covered lenders are required to report their credit decisions on every mortgage loan they receive, the loan amount, the location of the property tied to the loan up to the Census tract level, the borrowers' characteristics, and the type of purchasers if the loan is sold in the same calendar year.

In this study, we focus on all the depository institutions covered by the HMDA, which differ systematically from the non-depository institutions in that the former obtain funds mainly through accepting deposits from the public and are usually subject to more rigorous regulatory scrutiny. We limit the sample to the conventional, one to four-family, owner-occupied, and first-lien home purchase loans. With these restrictions, all the loans in our sample fall below the Federal Housing Finance Agency's (FHFA) conforming loan limits for Texas of each year. We also exclude the loans from the applicants who are not natural persons, such as a business and a corporation or partnership, which can be identified by the absence of the applicants' demographic characteristics. Further, we drop the loan applications withdrawn by the applicants or closed due to incompleteness, and the loans whose preapproval request were denied by the financial institutions or approved but not accepted by the applicants. Finally, we exclude the loans purchased by an institution to avoid double-counting, since these loans were reported by both the originating institution and the purchasing institution (Dell'Ariccia, Igan and Laeven 2012).

Each loan in our final sample falls into one of the following three statuses: (1) originated, (2) denied by the financial institutions in the current calendar year, (3) approved but not accepted. The HMDA codes a loan as status (2) if the borrower has supplied all the necessary information but fails lenders' creditworthiness conditions. Instead, if the borrower has met the creditworthiness conditions but fails any other requirements, such as clear-title requirements and acceptable property survey, which leads the loan not to be consummated, then the loan is coded as status (3).[8] We examine the first two statuses to identify the change in the likelihoods of the loans being originated and denied. Thus, the difference between the treatment effects on these two outcome variables reflects the effect on the probability of the loans being approved by lenders but not accepted by borrowers.

The HMDA requires lenders to document up to three reasons for each denied loan voluntarily. These reasons could be borrowers' ineligibility in the debt-to-income ratio, employment history or credit history, insufficient collateral or cash, unverifiable information, incomplete credit application, and denied mortgage insurance. In particular, the debt-to-income ratio would be indicated as a denial reason if the applicant had insufficient income for the amount of credit requested or had excessive obligations in relation to income. Employment history would be indicated if the applicant had temporary or irregular employment. Insufficient collateral value would be indicated if the applicant did not have sufficient type or value of collateral determined by the appraisal of a qualified expert.[9] In our sample, lenders reported at least one denial reason for 79% of the denied loans. We define a set of denial reason dummies for each specific reason according to whether the reason was reported among one of the three denial reasons provided by the lender. Thus, we can compare the reasons for the denied loans from the pipeline-present areas with those from the pipeline-free areas across the years with versus without pipeline incidents.

The HMDA further indicates whether a loan is sold in the same calendar year if the loan has been originated. Although the precise identity of the purchasing institution is not provided, the

---

[8] See more at https://www.ffiec.gov/hmda/faqreg.htm#action

[9] See more explanations at https://www.ffiec.gov/hmda/pdf/2013guide.pdf

HMDA data report whether the purchaser is private or government-owned or sponsored. Therefore, we consider an originated loan to be sold to the secondary market if a purchaser type can be identified in the HMDA data. Moreover, according to whether the purchaser is Fannie Mae (FNMA), Ginnie Mae (GNMA), Freddie Mac (FHLMC) or Farmer Mac (FAMC), we can identify whether the loan was sold to the GSEs.

We collect additional individual- and Census tract-level information from the HMDA data. Specifically, we control for the loan amount and the applicant's annual gross income in $10,000's, gender, race, ethnicity, and the presence of co-applicants. Also, we take into account Census tract-level time-variant characteristics based on the borrowers' geography, including population density and median family income in $10,000's, and time-invariant characteristics including minority population percentage, the number of owner-occupied units, and the number of 1- to 4-family units.

2.3.2 Pipeline Infrastructure and Pipeline Incidents

Our measure of pipeline infrastructure is based on the shapefiles of pipelines provided by the U.S. Energy Information Administration. Figure 1 illustrates the geographic locations of major pipelines that transport crude oil, hydrocarbon gas liquids, natural gas, and petroleum products in Texas. In the figure, we use the 2003 Rural-Urban Continuum Code to distinguish the nonmetro areas from the metro areas. By comparison, 86% of the nonmetro Census tracts are covered by pipelines, while this percentage in the metro areas is only 52%. Moreover, the average length of pipelines in the nonmetro Census tracts is more than eight times as long as that in the metro Census tracts. We also identify the geographic locations of the petroleum refineries in Texas using the shapefiles provided by the U.S. Energy Information Administration.[10] We calculate the distance from the centroid of each Census tract to the nearest operable petroleum refineries in Texas to control for the impact of refinery facilities, which could otherwise confound the permanent effects of pipeline infrastructure.

Our measure of pipeline incidents is obtained from the Pipeline and Hazardous Materials Safety Administration (PHMSA) of the U.S. Department of Transportation. The data contain records for the full universe of incidents reported by the operators of federally-regulated and state-regulated natural gas and hazardous liquid pipelines. For each reported incident, the PHMSA records the date, location, causes, and consequences regarding fatality, injury, and total property damage. These incidents take the form of leak, rupture, spill, ignition, or explosion and could be caused by multiple reasons. Among all the pipeline incidents that occurred in Texas from 2005 to 2011, 26% were caused by corrosion failure, 29% by equipment failure, 12% by excavation damage, 7% by incorrect operation, 10% by material or weld failure, 3% by natural force damage, 5% by other outside force damage, and the remaining 7% were caused by other unrecorded reasons.

Since the HMDA use Census tract identifiers from the Census 2000 series for the mortgage data from 2005 to 2011, we project the locations of all the incidents with non-missing geographic coordinates over this period onto the 2000 Texas Census tract map. Then we calculate the number and the associated property damage of the incidents at the Census tract level, which is the lowest geographic level that can be identified in the HMDA data. Figure 2 illustrates each Census tract's total property damage over the seven years. By comparison, the average financial loss due to

[10] The data are available at https://www.eia.gov/maps/layer_info-m.php

pipeline incidents in the nonmetro Texas is nearly eleven times as much as that in the metro areas, which is consistent with the distribution pattern of the pipeline infrastructure across the two regions.

2.3.3 Sample

In this study, we limit the sample to the nonmetro Texas to obtain more precise estimates of the effects of pipeline hazards. We construct a pooled cross-sectional dataset for the loans in 746 Census tracts of 174 nonmetro counties from 2005 to 2011. Each observation represents an individual loan application in one year, containing information on lender's credit decisions, the existence of pipeline infrastructure, the history of pipeline incidents in the Census tract where the mortgaged property is located, and the other individual- and Census tract-level characteristics.

In total, we identify 315 pipeline incidents taking place in 160 nonmetro Census tracts during the sampling period. Figure 3 plots the annual counts of the incidents and the inflation-adjusted total property damage for each year over the sampling period. The annual frequency of pipeline incidents ranges from the minimum of 33 in 2008 to the peak of 58 in 2009. Meanwhile, the total financial loss varies from 3.25 million dollars in 2006 to 116.5 million dollars in 2005, which is highly correlated with the severity of the incidents in each year.

Table 1 reports the summary statistics by the treatment and control groups. On average, the treatment group had a lower loan origination rate by 1.3 percentage points, a higher loan denial rate by 1.8 percentage points, a lower share of loans sold to the secondary market by 6.1 percentage points, and a lower share of loans sold to the GSEs by 4.4 percentage points. By comparison, while 5.9% of the mortgaged properties in the treatment group were affected by pipeline incidents, the proportion in the control group was only 0.1%. Meanwhile, compared to the average property damage of $101,036 in the treatment group, the financial loss in the control group was almost trivial. The extremely rare incidence and the minor property damage reflect the spillover effects from the adjacent Census tracts, which could happen when oil or gas spills or other contaminants migrate from the pipeline-present neighborhoods to these pipeline-free areas.[11] In our model, we ignore the spillover effects to the control group so that the estimates should be interpreted as a lower bound of the true treatment effects.

Next, comparing borrowers' creditworthiness and demographic characteristics across the two groups, we find that borrowers' loan amount and annual gross income in the treatment group was higher than that in the control group by $11,820 and $10,100, respectively. However, the average loan-to-income ratio of 1.75 in the treatment group was slightly lower than the average of 1.82 in the control group. Also, 49% of the borrowers in the treatment group had co-applicants, which was higher than the percentage in the control group (46.1%). Moreover, the treatment group had a lower share of Latinos by 8.2 percentage points and whites by 0.2 percentage points. It also had a higher portion of male applicants by 2.6 percentage points and blacks by 0.4 percentage points.

We further show that the two groups differ from each other in the Census tract-level characteristics. By comparison, the pipeline-present areas featured a lower population density and a lower percentage of minority population. Besides, the pipeline-present areas had a higher Census tract median family income of $47,710 compared to the average of $44,480 in the pipeline-free

[11] We identify one Census tract in Stephens County that does not have any pipeline infrastructure but once experienced oil spill due to tank overfill in 2010.

areas. The treatment group also had more owner-occupied units and more 1- to 4- family units, which were higher than the average of the control group by 28% and 25%, respectively. Finally, on average the Census tracts in the treatment group are located closer to the petroleum refineries (110 km) relative to those in the control group (127 km).

Given the different observable characteristics between the treatment and the control groups, we use a difference-in-difference model to take into account the permanent difference in the credit availability across the two groups. We also control for a rich set of covariates to address the concerns about the omitted variable bias. Moreover, we demonstrate the common trends in the observable characteristics across the two groups over the years with versus without pipeline incidents to show that the estimated treatment effects are not confounded by the changes in borrowers' creditworthiness or demographic composition.

2.4 Baseline Results

We begin by estimating the effects of pipeline infrastructure and pipeline incidents on lenders' credit decisions using equation (1). Results are presented in Table 2. The dependent variables are whether a loan is originated (Column 1), whether it is denied (Column 2), whether a loan is sold to the secondary market if it has been originated (Column 3), and whether it is sold to the GSEs (Column 4). In the last two cases, we restrict the sample to all the originated loans. All standard errors are clustered at the Census tract level.

The main coefficients of interest in Table 2 include the effects of pipeline infrastructure, i.e., the coefficient of the treatment dummy $pipe_j$, and the treatment effects of new incidents, i.e., the coefficient of the interaction term $pipe_j * incid_{jt}$. The coefficient of the treatment dummy in

Column 1 indicates that a permanent difference existed in the origination rates between the treatment and the control groups, all else being equal. On average, an average loan in the pipeline-present areas was less likely to be originated by 1.9%, and the estimate is statistically significant at the 1% level. In Column 2, the likelihood of the loan being denied in the treatment group was higher by 1.3% at the significance level of 5%. Since the loans that failed to be originated could be either denied or approved but not accepted, the difference in the absolute value of the estimates between Column 1 and Column 2 indicates the effects on the probability of the loan being conditionally approved. Next, the coefficients of the interaction terms in the first two columns show that whenever new incidents happened lenders further decreased the likelihood of originating the loans in the affected Census tracts by 1.8% and increased the likelihood of denying the loans by 1.7%. Both estimates are statistically significant at the 5% level.

In Column 3 and Column 4, we restrict the sample to all the originated loans to examine the sales of the loans to the secondary market. The coefficient of the treatment dummy in Column 3 indicates that the loans originated in the pipeline-present areas were 1.8% less likely to be sold in the securitization market at the significance level of 5%. In Column 4, although the probability of being sold to the GSEs was slightly lower by 0.5% for the loans in the treatment group, the estimate is statistically insignificant. Next, the positive signs of the interaction terms in the two columns suggest that when new incidents happened lenders tended to sell more originated loans in the affected areas for securitization, but the evidence is statistically insignificant.

The coefficients of the control variables in Table 2 are consistent with our expectation. For example, the loans from those who were male and white and those who had a higher annual gross income and a lower loan-to-income ratio were significantly more likely to be originated, while the loans from the Latino and black applicants were less likely to be originated. Meanwhile, the likelihood of a loan being originated is also positively correlated with some Census tract-level characteristics such as population density and median family income at the significance level of 1%. The estimated coefficients of the control variables in Column 2 show a consistent pattern, taking the opposite signs compared to those in Column 1.

We also estimate the marginal effects of the monetary loss in pipeline incidents by replacing the incident dummy, $incid_{jt}$, with the total property damage for each Census tract in equation (1). The estimates of the permanent group difference in the four credit outcomes are consistent with the results in Table 2.[12] However, we do not identify strong evidence showing significant marginal effects of the property damage on individual loans' credit outcomes, which indicates that lenders did not base their credit decisions on the severity of the incidents. Compared with the findings in Table 2, our results suggest that it could be the information shock itself rather than the salience of the incidents that affected lenders' risk perceptions. However, as it usually takes weeks or even months to investigate an incident and disclose the results to the public, lenders' weak response to the real loss of the incidents could also result from the delay in information disclosure.

2.5 Timing of The Incidents

2.5.1 Before and After an Incident

We expect that the current and the lagged status of pipeline incidents could have effects on lenders' credit decisions while the future incidents should not. If the leads of the treatment were significantly different from zero, then the common trends assumption would be violated. It would

[12] We do not report the estimation results due to space limit. The table is available upon request.

suggest that there could be unobservable factors correlated with the occurrence of pipeline incidents that explained lenders' credit decisions. Thus, we add one lead and one lag of the interaction term to equation (1) and estimate the model using the data from 2006 to 2010.

In Table 3, the lead terms are statistically insignificant across the four columns, which indicate that within one year before the incidents there was no additional significant difference in the mortgage credit availability between the treatment and control groups. Hence, lenders' changing credit decisions in the years when pipeline incidents occurred were less likely to be driven by factors other than the exogenous shock of the incidents. Moreover, although the occurrence of pipeline incidents significantly affected the likelihoods of the loans being originated and denied in the treatment group, we do not identify any significant persistent effects on lenders' credit decisions following the incidents. The lag terms of the incidents are statistically insignificant across all the columns, which means that the incidents shock quickly died out one year after. The short-lived incidents effects are likely to result from lenders' uncertainties about the post-incident remediation responsibility and costs, which would diminish as long as the cleanup liability is assigned. Jackson (2001) has similar findings that remediation status has a statistically significant effect on lenders' risk perceptions for a contaminated property. In his study, the percentage of lenders that would not provide a mortgage loan due to excessive environmental risks decreases from 93.2% before the cleanup to 4.2% by the time when the remediation has been completed.

2.5.2 Early versus Late Incidents

In the baseline analysis, we aggregate the loan applications and the pipeline incidents to the annual observations at the Census tract level, since the exact credit decision date is not available in the public HMDA data. However, since pipeline incidents could happen across all the months of a year, those occurred later of the year would only affect a portion of the loans issued that year. Thus, the analysis on an annual basis would result in underestimates of the true effects of pipeline incidents. In this section, we demonstrate the potential bias by comparing the effects of the incidents that happened at the beginning of each year with the baseline treatment effects.

Specifically, we only include the Census tracts that experienced the first pipeline incident as early as in the first quarter of the year while excluding the Census tracts that had their first incident later than March. We expect these early incidents to affect most of the loans issued during the year and result in the estimated treatment effects closer to the true effects. Since the PHMSA data only report the date for the incidents occurred before 2010, we run the robustness check using the sampling period of 2005-2009. In Table 4, we report the treatment effects of both the incidents that happened in the first quarter of the year (Panel A) and those that happened at any time of the year (Panel B). We find that the former generates a treatment effect of a larger magnitude on the likelihood of a loan being originated (-3.2%) compared to the latter (-1.9%). Consistently, the treatment effect from the early incidents on the loans' denial rate (2.9%) is also larger in magnitude than the effect estimated using the whole set of the incidents (2.2%). Meanwhile, the early incidents were also more influential on the likelihood of selling an originated loan to the GSEs compared to the average effect of all the incidents (4.2% versus 1.5%).

2.5.3 No Evidence of Sorting

The baseline results suggest that lenders further adjusted their lending decisions whenever new incidents happened on top of the permanent difference in the credit availability between the pipeline-present and the pipeline-free areas. Our explanation is that lenders denied the risky loans due to the elevated risk perceptions following the incidents. In this subsection, we rule out the possibility that the pool of potential borrowers in the treatment group could have changed after the incidents when less creditworthy borrowers sorted to the pipeline-present neighborhoods. Due to the limitation of the HMDA data, we cannot observe all the factors contributing to a borrower' creditworthiness (Munnel *et al.* 1996). However, we can make use of the available characteristics of the borrowers and the loans in the dataset to test all the potential consequences of sorting that could be observed by us. Our robustness checks also help to exclude the possibility of sorting caused by borrowers' unobservable features that are correlated with the observable measures of their creditworthiness.

2.5.4 Borrowers' Characteristics

We first test the relative changes in the creditworthiness and the demographic characteristics of the borrowers in the treatment group when pipeline incidents happened compared to the trend in the control group. Specifically, we replace the dependent variables of equation (1) with borrowers' individual-level characteristics, including loan amount, annual gross income, the loan-to-income ratio, the presence of co-applicants, gender, ethnicity, and race, taking into account the same set of Census tract-level characteristics, lender fixed effects, county fixed effects, and year fixed effects. While we allow the existence of a permanent difference between the treatment and the control groups, we expect that new incidents did not lead to further changes in these characteristics. In Table 5, we do not find any significant permanent group difference in either the creditworthiness or the demographic features of the borrowers except that the average loan amount and annual gross income of the borrowers in the treatment group was significantly higher by $10,150 and $9,340 respectively than that in the control group. Moreover, when new incidents happened, there was no further change in the average quality of borrowers in the areas affected by the incidents. Our findings indicate that the pool of borrowers in the years when pipeline incidents happened was not significantly different from that in the incidents-free years, which provides evidence for the common trends of the covariates. The results support that the treatment effects were unlikely to be caused by the changes in the creditworthiness or the demographic composition of borrowers after the incidents.

2.5.5 Loans' Denial Reasons

Another concern is that the treatment effects could also be confounded by factors that are unobservable to researchers but correlated with lenders' credit decision. For example, lenders could deny the loans from those risk-loving borrowers who are willing to bear the environmental hazards in the neighborhoods affected by pipeline incidents. In this subsection, we make use of the denial reasons provided by the HMDA data to cast light on the changes in the unobservable characteristics of borrowers in the incidents-affected areas. We restrict the sample to all the denied loans to which lenders provided at least one denial reason. Then we re-estimate equation (1) by replacing the dependent variable with a dummy indicating each denial reason to examine the change in the likelihood of the loans being rejected for each particular reason when pipeline incidents happened.

In Table 6, we only find weak evidence in Column 2 showing that lenders were more likely to reject the loans for the reason of employment history by 2.1% when pipeline incidents occurred at the significance level of 10%. This evidence indicates that the occurrence of pipeline incidents led lenders to be more cautious about the borrowers who had temporary and irregular employment. Although some of these borrowers might show decent current income, lenders concerned more about the uncertainties embedded in the unstable employment that could affect borrowers' ability to make regular payments over the whole term of a loan. Compared to the insignificant change in the likelihood of denying a loan for the reason of the debt-to-income ratio (Column 1), our findings suggest lenders' worries in the borrowers' repayment ability in the long run rather than in the short term when environmental hazards are identified.

Across all the other columns, we do not find any significant change in the likelihood of denying the loans for any other reason in the incidents-affected areas. For instance, Column 1 shows that there was no significant change in the likelihood of the loans being rejected due to unqualified debt-to-income ratio, which suggests that the risk preference of borrowers in the treatment group did not change disproportionally when pipeline incidents occurred compared to that of the control group. In Column 4, we find that the denied loans in the treatment group had 3.2% higher likelihood of being rejected due to insufficient collateral value at the significance level of 5%. The results indicate that systematic difference in housing values could exist between the treatment and the control groups, part of which could be attributed to the easement and restrictions on land titles and the inconvenience brought about by the utility services from pipeline operators. However, we find no further change in the probability of the loans being rejected due to insufficient collateral value when real incidents occurred, which suggests that collateral depreciation did not happen right after the incidents. The robustness check helps rule out simultaneous changes in the unobservable factors such as borrowers' risk preference and immediate housing devaluation that could affect lenders' credit decisions when pipeline incidents happened.

2.5.6 Cumulative Sorting Effects

Since the occurrence of pipeline incidents usually signals the insecurity of pipeline infrastructure in a neighborhood, the history of pipeline incidents could gradually shape the potential homebuyers' risk perceptions towards the neighborhood in the long run. If the lower origination rates of the loans in the pipeline-present areas were caused by the selection of risky households into these communities, then the selection problem would be more severe in the Census tracts of higher intensity of pipelines, since historically pipeline incidents could have happened more frequently in these areas. Likewise, the selection issue would also be more noticeable in the Census tracts with higher frequency of pipeline incidents over a given period, since more regular and intense incidents could send stronger signals to the potential homebuyers seeking less attractive neighborhoods.

In this subsection, we test whether it is the cumulative sorting effects that explain our baseline results. First, we split the pipeline-present areas by the length of pipeline infrastructure in each Census tract into four quartile groups. In Table 7, we examine the observable characteristics of both borrowers and neighborhoods across the pipeline-free Census tracts and the first and the fourth quartile groups for all the loans from 2005 to 2011. We report the mean characteristics for each group in Columns 1-3 with the standard deviation in parenthesis. We also compare the characteristics between the pipeline-free areas and the first quartile group and between the first and the fourth quartile groups respectively, reporting the difference with the p-value in parenthesis for an unpaired two-sample mean-comparison t-test in Column 4 and Column 5. If sorting existed, we would identify riskier borrowers in the pipeline-present areas and especially in the higher quartile group where pipeline infrastructure was more intense. However, we do not find any evidence of sorting during this period. For example, the t-test results in Column 4 show that the loans in pipeline-free areas had a higher loan-to-income ratio indicating higher credit risk. Moreover, the applicants in the first quartile group were more likely to be black and less likely to be white, but they were more likely to live in the neighborhoods featuring a lower minority percentage, a higher Census tract median family income, more owner-occupied units, and more 1- to 4- family units. Next, the t-test results in Column 5 show that borrowers in the fourth quartile group had a significantly higher loan amount and annual gross income but had a loan-to-income ratio statistically comparable to that in the first quartile group. They also had a higher likelihood of having co-applicants and living in more favorable neighborhoods. All the available measures of borrowers' creditworthiness indicate that the borrowers in the upper quartile group turned out to be less risky than those in the group with a lower intensity of pipelines.

We further categorize all the Census tracts by the total number of pipeline incidents that they encountered from 2005 to 2011 into four groups, in which the Census tracts did not experience any incident (Column 1), experienced one incident (Column 2), two incidents (Column 3), or three or more incidents (Column 4), respectively.[13] If homebuyers did select themselves to less attractive neighborhoods in response to the signals sent by the incidents over the seven years, then by the end of the period riskier homebuyers would have gradually sorted to the communities where incidents occurred more frequently. In Table 8, we examine the same set of features across the four groups for all the loan applications in 2011, which is the last year of our sampling period. We report the mean characteristics for each group in Columns 1-4 with the standard deviation in

[13] We also split the Census tracts by the quartile of their total property loss from pipeline incidents between 2005 and 2011. We obtain similar conclusions as we get from Table 8. The results are available upon request.

parenthesis. We also report the difference of the characteristics between the incidents-free Census tracts and the areas with one incident as well as the difference between the Census tracts with one incident and those with three or more incidents in Column 5 and Column 6 respectively with the p-value for the t-test in parenthesis. Again, we do not find strong evidence of sorting. For example, the t-test results in Column 5 indicate that the average loan amount and borrowers' annual gross income in the Census tracts with one incident were significantly higher than those in the incidents-free areas while the loan-to-income ratios were statistically indistinguishable across the two groups. Moreover, the borrowers in the one-incident group were more likely to be male, white, and had a co-applicant. They also had a higher chance of living in the neighborhoods with a lower population density and minority percentage, a higher median family income, and more owner-occupied units and 1- to 4- family units. Next, the t-test results in Column 6 indicate that the borrowers in the one-incident group and the three- or more-incidents group were statistically similar in terms of loan amount, annual gross income, the loan-to-income ratio, and the presence of co-applicants. While the neighborhoods in the Census tracts with three or more incidents had a higher minority percentage, we do not find strong statistical evidence showing that the two groups had significantly different median family income. Above all, the evidence indicates that the loans in the areas experiencing more frequent incidents were not riskier than those with less or none incidents at the end of the period of 2005-2011.

2.6 Heterogeneous Treatment Effects

2.6.1 Different Income Groups

In this subsection, we examine whether pipeline hazards had the same effects on the credit access across different income groups. All else being equal, the required costs of environmental investigation and remediation for a contaminated property could impose a greater financial burden on the lower-income borrowers in the incidents-affected areas. For this reason, we expect lenders to have particular concerns about the repayment capability of the lower-income borrowers who have weaker financial viability. Thus, we follow the rule provided by the Federal Financial Institutions Examination Council to categorize the baseline sample into the low- to middle-income group and the upper-income group according to whether the borrower's annual gross income is below or above 120% of the Texas nonmetropolitan median family income.

We estimate the heterogeneous treatment effects using the two subsamples respectively and report the results in Table 9. In Column 1, we find that the low- to middle-income borrowers in the pipeline-present areas had a lower origination rate by 2.5% compared to their counterparts in the pipeline-free areas at the 5% level. Whenever new incidents occurred, the difference in the origination rates was further enlarged by 3.6% at the 1% level. The corresponding pattern appears in the denial rates. By contrast, we do not find statistically or economically significant difference in the origination rates and denial rates among the upper-income borrowers as a result of the presence of pipeline infrastructure or pipeline incidence. The findings are consistent with our hypothesis that part of lenders' perceived risks of pipelines come from their concerns about borrowers' inability to repay the loan especially when cleanup liability is required. On the other hand, the results indicate that lenders tended to exclude the lower-income borrowers instead of the higher-income borrowers from their portfolio to manage the potential credit risk.

We also find that the loans from the low- to middle-income borrowers were 1.9% less likely to be sold to securitizers due to pipeline hazards (Column 3) while the pipeline-related loans from the upper-income borrowers were 1.6% less likely to be packaged for securitization (Column 7) at the 10% level. In the securitization market, although lenders may sometimes hold risky loans in their portfolio out of the reputational considerations (Agarwal, Chang and Yavas 2012), more often they adversely select the risky loans for securitization to transfer the credit risk (Jimenez and Saurina 2006; Dell'Ariccia, Igan and Laeven 2012; Keys, Seru and Vig 2012; Simkovic 2013). On the other hand, while the secondary market investors are usually unaware of the localized environmental risks, they can make purchase decisions following standards such as the "obligation ratios" that relate the applicant's housing expense and total debt burden to total income (Munnel *et al.* 1996). In our study, we do not have enough information to distinguish the actions chosen by lenders. Neither can we identify which party plays a dominant role since securitization involves the actions of both sellers and buyers. Nevertheless, our findings suggest slightly lower marketability of the loans related to pipeline hazards from the low- to middle-income borrowers who are usually deemed as having higher default risk in case of environmental hazards. We infer that even though purchasers in the secondary market cannot recognize the environmental risks hidden behind they can always screen the loans relying on the observable characteristics of borrowers.

2.6.2 Before, During, and After the Subprime Mortgage Crisis

In this subsection, we examine how lenders' risk-taking evolved with the stringency of securitizers' guidelines. Specifically, we follow the business cycle reference dates provided by the National Bureau of Economic Research (NBER) to split the baseline sample by the years before (2005-2007), during (2008-2009), and after (2010-2011) the late 2000s subprime mortgage crisis.[14] We further divide each period's subsample by borrowers' income to identify the target population for which lenders' risk management strategy was designed in each period. We estimate equation (1) using each specified subsample and report the results in Table 10.

Before the financial crisis (Panel A), the origination rate in the treatment group was insignificantly different from that in the control group among both the low- to middle-income borrowers and the upper-income borrowers. Whenever pipeline incidents happened, lenders reduced the origination rate in the affected areas by 3.5% for the low- to middle-income borrowers at the 5% level. By contrast, lenders did not adjust the loan origination rate for the upper-income counterparts in response to the incidents. In the secondary market, however, lenders managed to sell the risky loans from the upper-income borrowers to the GSEs, as is shown that in the incidents-affected areas the originated loans of this group were 2.9% more likely to be sold to the GSEs at the 5% level. The different risk management strategies towards the two income groups further indicate lenders' concerns about the default risk of the lower-income borrowers. Moreover, our findings demonstrate lenders' ability to package and sell the incidents-affected loans successfully even to the GSEs that are supposed to be most cautious about environmental hazards. Meanwhile, as lenders were only able to sell the risky loans from the upper-income borrowers, the findings suggest that the purchasers in the secondary market relied on the observable characteristics to screen the loans for sale when they only had limited information about the localized risks.

During the subprime mortgage crisis (Panel B), we do not find statistically significant evidence showing lenders' differential credit decisions towards pipeline hazards among the low-

[14] According to NBER, the most recent financial crisis lasted from December 2007 to June 2009. See more at http://www.nber.org/cycles.html

to middle-income borrowers regardless of whether real incidents happened, though all the coefficients show the expected signs and the magnitudes of the estimated effects are even greater compared to those in the previous period. However, during this period lenders managed the pipeline risks associated with the loans from the upper-income borrowers by decreasing the origination rate in the pipeline-present areas by 3.5% at the 5% level.

After the crisis (Panel C), lenders' credit decisions started to be significantly different across the treatment and the control groups especially among the low- to middle-income borrowers. As is shown in Column 1 and Column 2, the origination rate in the treatment group was significantly lower by 6.1% compared to that in the control group at the 5% level. Correspondingly, the denial rate was significantly higher by 6.7% at the 5% level. The magnitudes of both estimates are more than three times bigger than those for the full sample. By contrast, such significant group differences in the origination rates and the denial rates did not exist among the upper-income borrowers, which reconfirms lenders' concerns about the lower-income borrowers' repayment ability. We do not identify strong evidence showing lenders' response to the occurrence of new incidents after the crisis. Neither do we find significant evidence showing more sales of the risky loans to the securitization market during this period.

In sum, our results indicate that before the Great Recession lenders did not take systematically different actions towards the loans subject to potential pipeline risks. Instead, lenders exploited the originate-to-distribute model to manage pipeline hazards only when real incidents happened. It was only after the crisis that lenders started to distinguish the origination rate in the pipeline-present areas from that in the pipeline-free areas in a systematic way. The cyclicality of lenders' underwriting standards reflects the stringency of mortgage securitizers' guidelines across different periods. In particular, lenders' reliance on securitization as a risk management strategy coincided with the period of the mid-2000s, during which the deteriorating standards in the secondary market significantly reduced lenders' incentives to carefully screen and monitor borrowers' creditworthiness (Jimenez and Saurina 2006; Rajan, Seru and Vig 2011; Dell'Ariccia, Igan and Laeven 2012; Keys, Seru and Vig 2012; Simkovic 2013). The tightened securitization after the crisis forced lenders to manage the pipeline-related credit risk by discriminating the loans in the pipeline-present areas aggressively, which led to a particular credit crunch among the low- to middle-income borrowers during the period.

2.7 Conclusions

This study provides empirical evidence on mortgage lenders' perceptions of environmental risks using the 2005-2011 home mortgage loans and pipeline incidents in Texas. We find permanent difference in lenders' origination rates and denial rates between the pipeline-present areas and the pipeline-free areas. We interpret the difference in terms of lenders' perceived risks of pipelines, which is shown to come from their concerns about both collateral value and borrowers' repayment ability. We also find that lenders further lowered the credit availability to the pipeline-present areas whenever new incidents happened, which reflects their increased risk perceptions in response to the incidents. We interpret lenders' changing risk perceptions as evidence of their aversion to the uncertainties about the potential environmental liabilities, which would diminish soon after the cleanup responsibilities are assigned after the incidents. Finally, we show that lenders' risk-taking reflected the stringency of securitizers' guidelines across different periods over the financial crisis. Before the crisis, the originate-to-distribute business model helped lenders manage pipeline hazards passively, while securitizers in the secondary market screened the loans by the observable characteristics when they lacked sufficient information about the localized risks. After the tightening of securitization following the crisis, lenders managed pipeline hazards more aggressively by avoiding the lower-income borrowers in the pipeline-present areas due to the concerns about the default risk. Although our analysis is based on the observable characteristics of the loans in the HMDA data, we run a series of robustness checks to test the potential consequences of sorting due to both the observable factors and the unobservable factors that are correlated with the observable measures of borrowers' creditworthiness. Our robustness checks rule out the possibility of both simultaneous and cumulative sorting of risky borrowers into the neighborhoods with densely constructed pipelines and frequent incidents.

This study contributes to the policy debate on mortgage lenders' liability dilemma under the current environmental laws. While the federal Superfund and a series of state analogs have been designed to create a safe harbor from cleanup liabilities for lenders, a qualified exemption is still subject to when and how lenders interact with the contaminated properties as well as judicial interpretation on a case-by-case basis. Moreover, as the state statutes may differ from the federal scheme, lenders also have to be cautious about the state-specific qualifications for the liability exemption to apply within the state's jurisdiction (Ahrens and Langer 2008; Sigel and Bandza 2014). As we show in this study, a significant number of lenders are still reluctant to make loans to the neighborhoods subject to potential environmental risks. From a policy perspective, lenders' fear of environmental liabilities could lead to a credit crunch to otherwise creditworthy borrowers. More importantly, it could impede the real estate investment and slow down the redevelopment of those historically contaminated sites. This study calls for policy makers' involvement in seeking ways to reduce the deterrent effects of environmental contamination and the associated cleanup liabilities.

This study also sheds new light on mortgage lenders' risk management of environmental hazards. The empirical evidence suggests that lenders once managed the pipeline-related credit risk by denying the risky loans from the low- to middle-income borrowers and relying on the securitization market to sell the risky loans from the upper-income borrowers. Although we only examine the discrete lending decisions, lenders could always capitalize pipeline hazards into mortgage prices by requiring borrowers to pay the risk premium as a condition of closing the loan. Moreover, lenders may consider using tailored environmental insurance products to insure against the unknown environmental hazards (Bressler 2002; Davis and Levy 2012). Lenders may require borrowers to purchase the environmental insurance policies that mainly cover owners for their cleanup expenses above a certain amount for the unforeseen environmental conditions. Alternatively, lenders themselves may use the secured creditor environmental insurance policies, which could pay off the outstanding balance in the event of a default or cover the cleanup costs after foreclosing on the loan.

Finally, our study informs the public discourse about pipeline hazards and identifies the population at risk. We show that pipeline hazards have resulted in a direct loss of mortgage credit access to the low- to middle-income borrowers. Hence, environmental externalities of pipelines could become a potential barrier for those credit constrained families to obtain homeownership. Therefore, enhancing pipeline safety has significant policy implications for lower-income borrowers to improve their mortgage credit availability as well as homeownership opportunity, which is a primary vehicle for these households to accumulate wealth and access economic opportunity.

2.8 Tables and Figures

Table 2.1 Summary Statistics by the Treatment and the Control Groups, 2005-2011

	Pipeline-present Census tracts		Pipeline-free Census tracts	
	Mean	Std. Dev.	Mean	Std. Dev.
Dependent Variables				
Originated	0.723	0.448	0.736	0.441
Denied	0.204	0.403	0.186	0.389
Sold	0.434	0.496	0.495	0.500
Sold to the GSEs	0.223	0.417	0.267	0.442
Pipeline Incidents				
Number of pipeline incidents	0.059	0.235	0.001*	0.022
Property damage	101,036	2,685,651.50	0.119*	5.302
Borrowers' Characteristics				
Loan amount (in $10,000's)	11.707	10.142	10.525	8.867
Annual gross income (in $10,000's)	8.003	10.911	6.993	7.371
Loan-to-income ratio	1.746	1.830	1.818	3.691
Co-applicants	0.490	0.500	0.461	0.499
Latino	0.172	0.377	0.254	0.435
Male	0.770	0.421	0.744	0.436
Asian	0.008	0.092	0.010	0.098
Black	0.029	0.168	0.025	0.156
White	0.948	0.221	0.950	0.217
Census Tracts' Characteristics				
Population density (thousand per km^2)	1.429	2.691	4.527	4.078
Minority population%	0.294	0.213	0.385	0.253
Median family income (in $10,000's)	4.771	1.020	4.448	1.188
Number of owner-occupied units (in 1000)	1.465	0.671	1.144	0.416
Number of 1- to 4- family units (in 1000)	2.189	0.925	1.754	0.575
Distance to the nearest refineries (1000 km)	0.110	0.061	0.127	0.068
Number of Observations		55,344		7,988

Note: The treatment group includes all the nonmetro Census tracts in Texas where pipelines exist, and the control group includes the remaining nonmetro Census tracts free of any pipeline infrastructure. The sample includes the HMDA loan applications for the conventional, 1- to 4-family, owner-occupied, and first-lien home purchase loans in Texas from 2005-2011.
*One Census tract in Stephens County does not have any pipeline infrastructure but experienced an oil spill due to tank overfill in 2010.

Table 2.2 The Effects of Pipeline Incidents

	(1) Originated		(2) Denied		(3) Sold		(4) Sold to the GSEs	
	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.
$pipe_j$	-0.019***	-2.6	0.013**	2.1	-0.018**	-2.2	-0.005	-0.9
$pipe_j * incid_{jt}$	-0.018**	-2.0	0.017**	2.4	0.009	1.1	0.012	1.5
log(loan amount)	0.031***	5.2	-0.032***	-5.3	0.148***	10.5	0.029***	4.5
log(annual gross income)	0.049***	8.1	-0.059***	-9.5	-0.119***	-8.6	-0.037***	-5.6
Loan-to-income ratio	-0.007**	-2.4	0.007**	2.4	-0.041***	-4.6	-0.010***	-2.8
Co-applicants	0.001	0.2	0.006	1.6	0.003	0.7	0.011***	3.0
Latino	-0.081***	-13.9	0.077***	13.8	-0.065***	-8.2	-0.033***	-6.8
Male	0.019***	4.3	-0.017***	-4.1	0.002	0.4	0.003	0.6
Asian	-0.002	-0.1	-0.004	-0.2	0.015	0.6	0.031	1.3
Black	-0.145***	-6.8	0.145***	7.3	-0.015	-0.7	-0.061***	-3.3
White	0.049***	3.1	-0.052***	-3.4	0.006	0.3	-0.007	-0.5
Population density	0.005***	7.4	-0.005***	-6.5	0.002***	2.9	0.001	1.5
Minority population %	0.029	1.3	-0.030	-1.5	0.002	0.1	0.022	1.2
log(median family income)	0.112***	7.9	-0.105***	-7.8	0.042**	2.6	0.044***	2.8
Number of owner-occupied units	0.015	1.0	-0.009	-0.6	0.025*	1.9	0.007	0.6
Number of 1- to 4- family units	-0.006	-0.6	0.003	0.3	-0.011	-1.2	-0.004	-0.4
Distance to refineries	0.032	0.2	0.141	0.8	-0.175	-0.9	-0.243	-1.4
Constant	0.347***	8.1	0.493***	12.7	0.696***	14.8	0.738***	17.1
Year Fixed Effects	Yes		Yes		Yes		Yes	
County Fixed Effects	Yes		Yes		Yes		Yes	
Lender Fixed Effects	Yes		Yes		Yes		Yes	
Observations	62,244		62,244		45,104		45,104	
R-squared	0.166		0.188		0.536		0.583	

Notes: Columns 1 and 2 use the 2005-2011 full sample of loan applications, whereas Columns 3 and 4 use the 2005-2011 sample of the originated loans. The dependent variables are listed as the column titles. The variable $pipe_j$ is coded as 1 if pipelines are present in Census tract j, and 0 otherwise. The estimates of $pipe_j$ measure the permanent effects of pipeline infrastructure. The variable $incid_{jt}$ is coded as 1 if a pipeline incident occurred to Census tract j in year t, and 0 otherwise. The estimates of $pipe_j * incid_{jt}$ measure the effects of pipeline incidents. The estimates of the coefficients are presented in the table, with t-statistics reported using clustered standard errors by Census tract. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.3 The Effects Before and After a Pipeline Incident

	(1) Originated		(2) Denied		(3) Sold		(4) Sold to the GSEs	
	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.
$pipe_j$	-0.022***	-2.6	0.014*	1.9	-0.021**	-2.4	-0.010	-1.6
$pipe_j$*$incid_{j,t+1}$	-0.008	-0.9	0.008	1.0	-0.001	-0.1	-0.004	-0.4
$pipe_j * incid_{j,t}$	-0.018*	-1.7	0.017*	1.9	0.004	0.5	0.003	0.3
$pipe_j$*$incid_{j,t-1}$	-0.009	-0.8	0.007	0.7	-0.001	-0.1	0.001	0.1
Observations	46,817		46,817		33,604		33,604	
R-squared	0.178		0.202		0.544		0.613	

Notes: Columns 1 and 2 use the 2006-2010 full sample of loan applications, whereas Columns 3 and 4 use the 2006-2010 sample of the originated loans. The dependent variables are listed as the column titles. The estimates of $pipe_j$ measure the permanent effects of pipeline infrastructure. The estimates of $pipe_j * incid_{jt}$ measure the contemporaneous effects of pipeline incidents, the estimates of $pipe_j * incid_{j,t+1}$ measure the counterfactual effects before pipeline incidents, and the estimates of $pipe_j * incid_{j,t-1}$ measure the lagged effects of pipeline incidents. We use the same control variables and fixed effects as in the baseline case (Table 2.2). The estimates of the coefficients are presented in the table, with t-statistics reported using clustered standard errors by Census tract. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.4 Early versus Late Incidents

	(1) Originated		(2) Denied		(3) Sold		(4) Sold to the GSEs	
	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.
Panel A: Pipeline Incidents in the First Quarter of the Year								
$pipe_j$	-0.015*	-1.9	0.007	1.1	-0.014	-1.6	-0.001	-0.2
$pipe_j * incid_{jt}$	-0.032**	-2.0	0.029**	2.1	0.020	1.5	0.042***	2.7
Observations	50,979		50,979		37,173		37,173	
R-squared	0.170		0.194		0.528		0.572	
Panel B: Pipeline Incidents in the Whole Year								
$pipe_j$	-0.013*	-1.7	0.006	0.8	-0.014*	-1.7	-0.002	-0.3
$pipe_j * incid_{jt}$	-0.019**	-2.0	0.022***	2.8	0.009	1.0	0.015*	1.7
Observations	52,791		52,791		38,467		38,467	
R-squared	0.169		0.194		0.527		0.571	

Notes: The analysis in this table pertains to the 2005-2009 loan applications (Columns 1 and 2) and the sample of the originated loans over this period (Columns 3 and 4). Panel A includes a sample of the Census tracts that experienced incidents as early as in the first quarter of the year. Panel B includes a sample in which incidents occurred anytime of the year. The dependent variables are listed as the column titles. The estimates of $pipe_j$ measure the permanent effects of pipeline infrastructure. The estimates of $pipe_j * incid_{jt}$ measure the effects of pipeline incidents. We use the same control variables and fixed effects as in the baseline case (Table 2.2). The estimates of the coefficients are presented in the table, with t-statistics reported using clustered standard errors by Census tract. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.5 Borrowers' Characteristics

	(1) Loan Amount (in $10,000's)		(2) Annual Gross Income (in $10,000's)		(3) Loan-to-income Ratio		(4) Co-applicant		(5) Latino		(6) Male		(7) White	
	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.
$pipe_j$	1.015***	2.8	0.934***	3.4	-0.042	-1.1	0.014	1.4	0.007	0.6	0.005	0.7	-0.003	-0.4
$pipe_j * incid_{jt}$	0.142	0.5	0.283	1.2	-0.021	-0.8	0.012	1.1	-0.012	-1.3	0.011	1.4	0.008	1.6
Observations	63,332		62,244		62,244		63,332		63,332		63,332		63,332	
R-squared	0.147		0.032		0.020		0.068		0.256		0.012		0.034	

Notes: The table reports the estimation results using the 2005-2011 full sample of loan applications. The dependent variables are listed as the column titles. The estimates of $pipe_j$ measure the permanent effects of pipeline infrastructure. The estimates of $pipe_j * incid_{jt}$ measure the effects of pipeline incidents. We use the same control variables and fixed effects as in the baseline case (Table 2.2). The estimates of the coefficients are presented in the table, with t-statistics reported using clustered standard errors by Census tract. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.6 Lenders' Denial Reasons

	(1) Debt-to-income Ratio		(2) Employment History		(3) Credit History		(4) Insufficient Collateral		(5) Insufficient Cash		(6) Unverifiable information		(7) Credit Application Incomplete		(8) Mortgage Insurance Denied	
	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.
$pipe_j$	-0.011	-0.7	-0.010	-1.1	-0.018	-1.0	0.032**	2.1	-0.003	-0.3	0.007	0.7	0.010	0.9	0.004	1.3
$pipe_j * incid_{jt}$	-0.010	-0.6	0.021*	1.9	-0.001	-0.1	0.009	0.4	-0.004	-0.3	-0.018	-1.6	0.006	0.5	-0.002	-0.5
Observations	9,894		9,894		9,894		9,894		9,894		9,894		9,894		9,894	
R-squared	0.194		0.084		0.338		0.165		0.119		0.175		0.312		0.142	

Notes: The table reports the estimation results using the 2005-2011 sample of the denied loans with denial reasons available. The dependent variables are listed as the column titles. The estimates of $pipe_j$ measure the permanent effects of pipeline infrastructure. The estimates of $pipe_j * incid_{jt}$ measure the effects of pipeline incidents. We use the same control variables and fixed effects as in the baseline case (Table 2.2). The estimates of the coefficients are presented in the table, with t-statistics reported using clustered standard errors by Census tract. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.7 Cumulative Sorting Effects I: Census Tracts of Different Intensity of Pipelines, 2005-2011

	(1) Pipeline-free Census tracts	(2) The first quartile group	(3) The fourth quartile group	(4) Difference between (1) and (2) (*p*-value)	(5) Difference between (2) and (3) (*p*-value)
Loan amount (in $10,000's)	10.525	10.408	11.944	0.117	-1.536
	(8.867)	(9.577)	(9.528)	(0.360)	(0.000)
Annual gross income (in $10,000's)	6.993	7.186	8.348	-0.193	-1.162
	(7.371)	(8.100)	(10.909)	(0.075)	(0.000)
Loan-to-income ratio	1.818	1.718	1.714	0.100	0.003
	(3.691)	(1.145)	(2.912)	(0.019)	(0.908)
Co-applicants	0.461	0.451	0.515	0.011	-0.064
	(0.499)	(0.498)	(0.500)	(0.126)	(0.000)
Latino	0.254	0.260	0.162	-0.006	0.098
	(0.435)	(0.439)	(0.368)	(0.330)	(0.000)
Male	0.744	0.749	0.803	-0.004	-0.054
	(0.436)	(0.434)	(0.398)	(0.476)	(0.000)
Asian	0.010	0.013	0.005	-0.003	0.007
	(0.098)	(0.111)	(0.073)	(0.057)	(0.000)
Black	0.025	0.031	0.021	-0.006	0.011
	(0.156)	(0.174)	(0.142)	(0.006)	(0.000)
White	0.950	0.938	0.962	0.013	-0.024
	(0.217)	(0.242)	(0.192)	(0.000)	(0.000)
Population density	4.527	4.503	0.105	0.024	4.399
	(4.078)	(3.800)	(0.100)	(0.667)	(0.000)
Minority population %	0.385	0.375	0.284	0.010	0.091
	(0.253)	(0.239)	(0.199)	(0.003)	(0.000)
Median family income (in $10,000's)	4.448	4.650	4.738	-0.202	-0.088

Table 2.7 (cont.)

	(1.188)	(1.237)	(0.775)	(0.000)	(0.000)
Number of owner-occupied units	1.144	1.200	1.538	-0.056	-0.338
	(0.416)	(0.438)	(0.687)	(0.000)	(0.000)
Number of 1- to 4- family units	1.754	1.872	2.261	-0.118	-0.388
	(0.575)	(0.614)	(0.909)	(0.000)	(0.000)
Distance to refineries	0.127	0.119	0.095	0.008	0.024
	(0.068)	(0.061)	(0.057)	(0.000)	(0.000)
N	7,988	13,889	13,811		

Notes:

1. The table reports, for the 2005-2011 full sample, the mean characteristics of borrowers and neighborhoods for the pipeline-free Census tracts (Column 1), the Census tracts falling into the first quartile group (Column 2) and the fourth quartile group (Column 3) of the distribution of the pipelines' length in each Census tract. In parenthesis of each cell from Column 1 to Column 3 is the standard deviation of the characteristics of that cell.

2. The table also reports the difference of the characteristics between Column 1 and Column 2 (in Column 4) and the difference between Column 2 and Column 3 (in Column 5). In parenthesis of each cell from Column 4 and Column 5 is the p-value for an unpaired two-sample mean-comparison t-test testing the null hypothesis that the characteristics are equal across the two corresponding groups.

Table 2.8 Cumulative Sorting Effects II: Census Tracts of Different Frequencies of Pipeline Incidents, 2011

	(1) Incidents-free Census tracts	(2) Census tracts with one incident	(3) Census tracts with two incidents	(4) Census tracts with three or more incidents	(5) Difference between (1) and (2) (*p*-value)	(6) Difference between (2) and (4) (*p*-value)
Loan amount (in $10,000's)	11.561	14.099	12.437	12.930	-2.538	1.169
	(9.645)	(12.685)	(9.056)	(10.325)	(0.000)	(0.191)
Annual gross income (in $10,000's)	8.372	9.545	8.841	10.676	-1.173	-1.131
	(8.296)	(9.286)	(6.604)	(17.281)	(0.003)	(0.388)
Loan-to-income ratio	1.776	1.727	1.590	1.694	0.049	0.033
	(5.674)	(1.335)	(1.003)	(1.156)	(0.669)	(0.737)
Co-applicants	0.578	0.620	0.628	0.667	-0.042	-0.047
	(0.494)	(0.486)	(0.485)	(0.473)	(0.043)	(0.231)
Latino	0.197	0.207	0.177	0.219	-0.010	-0.012
	(0.398)	(0.405)	(0.383)	(0.414)	(0.563)	(0.720)
Male	0.776	0.837	0.787	0.849	-0.061	-0.012
	(0.417)	(0.370)	(0.411)	(0.359)	(0.000)	(0.689)
Asian	0.007	0.006	0.000	0.000	0.001	0.006
	(0.086)	(0.077)	(0.000)	(0.000)	(0.680)	(0.045)
Black	0.026	0.017	0.012	0.021	0.010	-0.004
	(0.161)	(0.128)	(0.110)	(0.143)	(0.085)	(0.712)
White	0.957	0.973	0.988	0.979	-0.015	-0.006
	(0.202)	(0.163)	(0.110)	(0.143)	(0.034)	(0.602)
Population density	1.793	0.567	0.092	0.150	1.226	0.416
	(2.965)	(1.376)	(0.123)	(0.197)	(0.000)	(0.000)
Minority population %	0.310	0.275	0.225	0.345	0.035	-0.069
	(0.220)	(0.193)	(0.142)	(0.166)	(0.000)	(0.000)
Median family income (in $10,000's)	5.180	5.401	5.161	5.258	-0.221	0.143

Table 2.8 (cont.)

	(1.143)	(0.843)	(0.590)	(0.986)	(0.000)	(0.069)
Number of owner-occupied units	1.355	1.528	1.474	1.406	-0.173	0.122
	(0.600)	(0.646)	(0.581)	(0.756)	(0.000)	(0.042)
Number of 1- to 4- family units	2.055	2.211	2.210	2.036	-0.156	0.175
	(0.851)	(0.795)	(0.825)	(1.089)	(0.000)	(0.039)
Distance to refineries	0.114	0.099	0.095	0.098	0.015	0.001
	(0.061)	(0.058)	(0.065)	(0.044)	(0.000)	(0.816)
N	3,097	663	164	192		

Notes:

1. The table reports, for the 2011 sample, the mean characteristics of borrowers and neighborhoods for the Census tracts without experiencing any incident (Column 1), experiencing one incident (Column 2), two incidents (Column 3), and three or more incidents (Column 4) from 2005 to 2011. In parenthesis of each cell from Column 1 to Column 4 is the standard deviation of the characteristics of that cell.
2. The table also reports the difference of the characteristics between Column 1 and Column 2 (in Column 5) and the difference between Column 2 and Column 4 (in Column 6). In parenthesis of each cell from Column 5 and Column 6 is the p-value for an unpaired two-sample mean-comparison t-test testing the null hypothesis that the characteristics are equal across the two corresponding groups.

Table 2.9 Heterogeneous Treatment Effects I: Different Income Groups

	Low- to Middle-income Borrowers								Upper-income Borrowers							
	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)	
	Originated		Denied		Sold		Sold to the GSEs		Originated		Denied		Sold		Sold to the GSEs	
	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.
$pipe_j$	-0.025**	-2.2	0.018**	2.1	-0.019*	-1.8	-0.009	-1.1	-0.008	-0.8	0.003	0.4	-0.016*	-1.8	-0.003	-0.4
$pipe_j * incid_{jt}$	-0.036***	-2.8	0.025**	2.1	0.004	0.3	0.008	0.6	-0.008	-0.8	0.015	1.6	0.008	0.8	0.010	1.0
Observations	26,023		26,023		17,074		17,074		36,221		36,221		28,030		28,030	
R-squared	0.187		0.207		0.598		0.596		0.157		0.177		0.518		0.586	

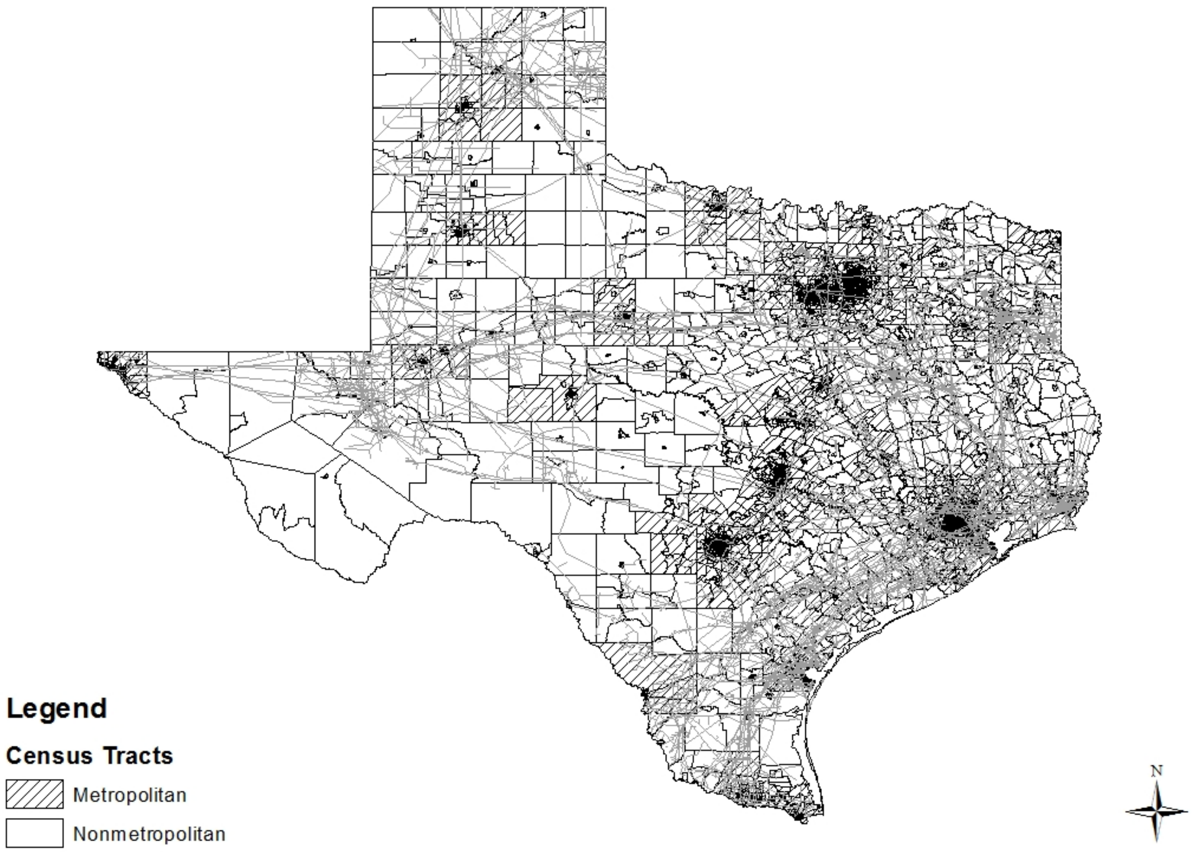
Notes: Columns 1-4 use the 2005-2011 low- to middle-income borrowers' loan applications, whereas Columns 5-8 use the 2005-2011 upper-income borrowers' loan applications. Columns 3, 4, 7, and 8 use the sample of the originated loans in the corresponding income group. The dependent variables are listed as the column titles. The estimates of $pipe_j$ measure the permanent effects of pipeline infrastructure. The estimates of $pipe_j * incid_{jt}$ measure the effects of pipeline incidents. We use the same control variables and fixed effects as in the baseline case (Table 2). The estimates of the coefficients are presented in the table, with t-statistics reported using clustered standard errors by Census tract. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.10 Heterogeneous Treatment Effects II: Before, During, and After the Subprime Mortgage Crisis by Income Group

	Low- to Middle-income Borrowers								Upper-income Borrowers							
	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)	
	Originated		Denied		Sold		Sold to the GSEs		Originated		Denied		Sold		Sold to the GSEs	
	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.	Coeff.	t-Stat.
Panel A: Before the Subprime Mortgage Crisis, 2005 - 2007																
$pipe_j$	-0.017	-1.4	0.009	0.9	-0.018	-1.6	-0.018	-1.6	0.008	0.7	-0.010	-0.9	-0.011	-0.9	0.002	0.2
$pipe_j * incid_{jt}$	-0.035**	-2.4	0.021	1.6	-0.003	-0.2	0.005	0.3	0.004	0.2	0.018	1.3	0.018	1.3	0.029**	2.2
Observations	16,631		16,631		11,106		11,106		22,001		22,001		16,986		16,986	
R-squared	0.184		0.212		0.574		0.567		0.175		0.202		0.479		0.547	
Panel B: During the Subprime Mortgage Crisis, 2008-2009																
$pipe_j$	-0.020	-0.9	0.003	0.1	-0.016	-0.7	0.012	0.8	-0.035**	-2.1	0.021	1.4	-0.002	-0.1	0.011	0.8
$pipe_j * incid_{jt}$	-0.040	-1.0	0.033	0.9	0.016	0.5	0.014	0.6	-0.025	-1.1	0.018	0.9	0.012	0.6	0.022	1.4
Observations	5,481		5,481		3,610		3,610		8,678		8,678		6,765		6,765	
R-squared	0.293		0.297		0.696		0.766		0.201		0.228		0.605		0.746	
Panel C: After the Subprime Mortgage Crisis, 2010-2011																
$pipe_j$	-0.061**	-2.0	0.067**	2.5	-0.030	-1.0	-0.003	-0.1	-0.024	-0.9	0.034	1.5	-0.022	-0.8	-0.011	-0.7
$pipe_j * incid_{jt}$	-0.060	-1.6	0.040	1.2	0.079*	1.9	0.010	0.4	-0.003	-0.1	-0.001	-0.0	0.000	0.0	-0.014	-0.8
Observations	3,911		3,911		2,358		2,358		5,542		5,542		4,279		4,279	
R-squared	0.280		0.287		0.650		0.745		0.220		0.226		0.597		0.700	

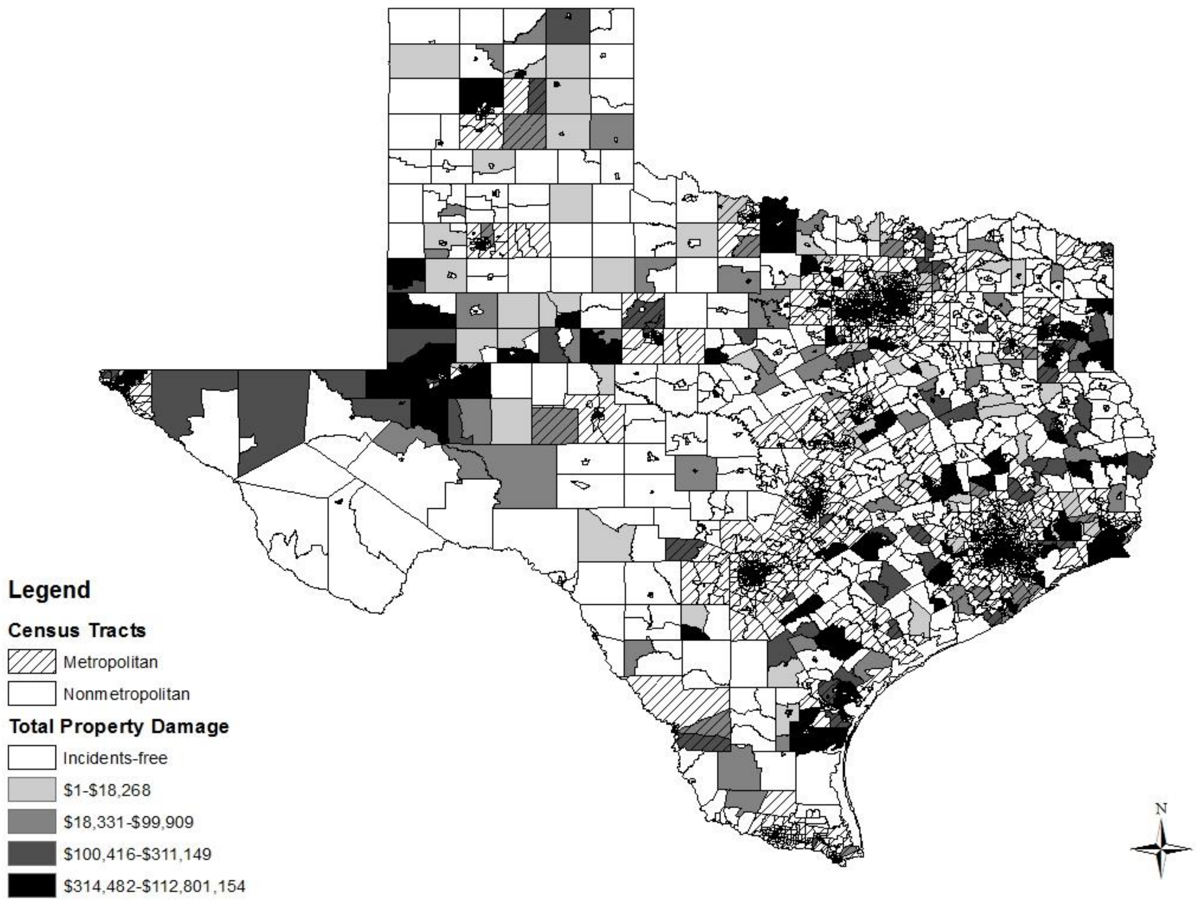
Notes: Panel A, B, and C report the estimation results using the 2005-2007, 2008-2009, and 2010-2011 sample, respectively. Columns 1-4 use the low- to middle-income borrowers' loan applications, whereas Columns 5-8 use the upper-income borrowers' loan applications. Columns 3, 4, 7, and 8 use the sample of the originated loans in the corresponding income group. The dependent variables are listed as the column titles. The estimates of the coefficients are presented in the table, with t-statistics reported using clustered standard errors by Census tract. *** p<0.01, ** p<0.05, * p<0.1.

Figure 2.1 Texas Pipeline Network



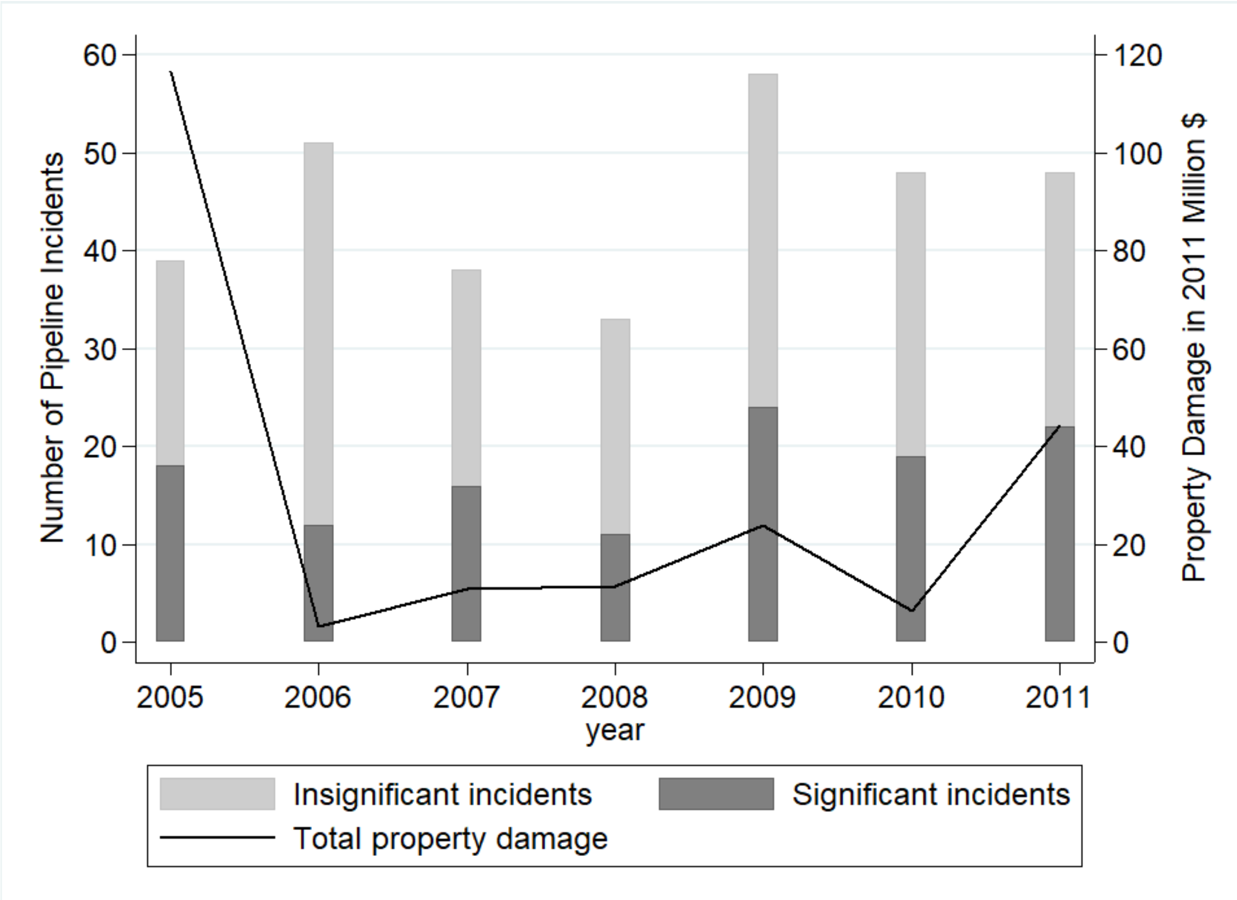
Source: Generated by the authors using data from the U.S. Energy Information Administration.

Figure 2.2 Quartile Distribution of Total Property Damage of Pipeline Incidents in Texas, 2005-2011



Source: Generated by the authors using data from the U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration. We convert the total property damage into 2011 U.S. dollars.

Figure 2.3 Pipeline Incidents in the Nonmetro Texas from 2005 to 2011



Source: Generated by the authors based on the data from the Pipeline and Hazardous Materials Safety Administration (PHMSA). According to the PHMSA, significant incidents refer to those associated with fatality, injury, or total property damage over $50,000. We convert the total property damage into 2011 U.S. dollars.

2.9 References

Agarwal, S., Chang, Y. and Yavas, A., 2012. Adverse selection in mortgage securitization. *Journal of Financial Economics*, *105*(3), pp.640-660.

Ahrens, M.H. and D.S. Langer. 2008. Environmental Risks for Lenders Under Superfund: A Refresher for the Economic Downturn. *Bloomberg Corporate Law Journal*, 3, pp.482-493.

Jim, B. and Peter, Z., 1996. How complete is HMDA?: HMDA coverage of Freddie Mac purchases. *Journal of Real Estate Research*, *11*(1), pp.39-55.

Berndt, A. and Gupta, A., 2009. Moral hazard and adverse selection in the originate-to-distribute model of bank credit. *Journal of Monetary Economics*, *56*(5), pp.725-743.

Bond, S., Kinnard Jr, W.N., Worzala, E.M. and Kapplin, S.D., 1998. Market participants' reactions toward contaminated property in New Zealand and the USA. *Journal of Property Valuation and Investment*, *16*(3), pp.251-272.

Bressler, A. 2002. Environmental Insurance Changes the Game for Commercial Lenders. International Risk Management Institute. https://www.irmi.com/articles/expert-commentary/environmental-insurance-changes-the-game-for-commercial-lenders

Carroll, T.M., Clauretie, T.M., Jensen, J. and Waddoups, M., 1996. The economic impact of a transient hazard on property values: The 1988 PEPCON explosion in Henderson, Nevada. *The Journal of Real Estate Finance and Economics*, *13*(2), pp.143-167.

Davis, A.N. and A.D. Levy. 2012. Lending on Contaminated Properties: Using Environmental Insurance to Manage Environmental Risk and Get the Deal Closed. Corporate LiveWire. October 18, 2012. http://www.shipmangoodwin.com/files/17329_CorporateLiveWire.pdf

Dell'Ariccia, G., Igan, D. and Laeven, L.U., 2012. Credit booms and lending standards: Evidence from the subprime mortgage market. *Journal of Money, Credit and Banking*, *44*(2-3), pp.367-384.

Gamble, H.B. and Downing, R.H., 1982. Effects of nuclear power plants on residential property values. *Journal of Regional Science*, *22*(4), pp.457-478.

Gracer, J. and Leas, C., 2008. Lender Beware: Navigating the Superfund Safe Harbor During Workouts and Foreclosures. *New York Law Journal*, (July).

Hansen, J.L., Benson, E.D. and Hagen, D.A., 2006. Environmental hazards and residential property values: Evidence from a major pipeline event. *Land Economics*, *82*(4), pp.529-541.

Healy, P.R. and Healy Jr, J.J., 1992. Lenders' perspectives on environmental issues. *The Appraisal Journal*, *60*(3), p.394.

Igan, D., Mishra, P. and Tressel, T., 2012. A fistful of dollars: lobbying and the financial crisis. *NBER Macroeconomics Annual*, *26*(1), pp.195-230.

Islam, A., Ahmed, A., Hur, M., Thorn, K. and Kim, S., 2016. Molecular-level evidence provided by ultrahigh resolution mass spectrometry for oil-derived DOC in groundwater at Bemidji, Minnesota. *Journal of hazardous materials*, *320*, pp.123-132.

Jackson, T.O. 2001. Environment risk perceptions of commercial and industrial real estate lenders. *Journal of Real Estate Research*, *22*(3), pp.271-288.

Jimenez, G. and J. Saurina. 2006. Credit cycles, credit risk, and prudential regulation. *International Journal of Central Banking*, 2(2), pp.65-98.

Keys, B.J., Seru, A. and Vig, V., 2012. Lender screening and the role of securitization: evidence from prime and subprime mortgage markets. *The Review of Financial Studies*, *25*(7), pp.2071-2108.

Kiel, K.A., 1995. Measuring the impact of the discovery and cleaning of identified hazardous waste sites on house values. *Land Economics*, pp.428-435.

Kohlhase, J.E., 1991. The impact of toxic waste sites on housing values. *Journal of urban Economics*, *30*(1), pp.1-26.

Matheny, K. 2016. 30 Years Later, Contamination Remained at Site of Pipeline Spill. Detroit Free Press. May 10, 2016. http://www.freep.com/story/news/local/michigan/2016/05/07/enbridge-line5-oil-spill-hiawatha-national-forest/83507228/

McCluskey, J.J. and Rausser, G.C., 2001. Estimation of perceived risk and its effect on property values. *Land Economics*, *77*(1), pp.42-55.

McCluskey, J.J. and Rausser, G.C., 2003. Hazardous waste sites and housing appreciation rates. *Journal of Environmental Economics and Management*, *45*(2), pp.166-176.

McCoy, S.J. and Walsh, R.P., 2014. *WUI on fire: Risk, salience & housing demand* (No. w20644). National Bureau of Economic Research.

Medina, D.A. 2016. Sioux' Concerns Over Pipeline Impact On Water Supply 'Unfounded,' Company Says. NBC News. September 13, 2016. http://www.nbcnews.com/news/us-news/sioux-s-concerns-over-pipeline-impact-water-supply-unfounded-company-n647576

Michaels, R.G. and Smith, V.K., 1990. Market segmentation and valuing amenities with hedonic models: the case of hazardous waste sites. *Journal of Urban Economics*, *28*(2), pp.223-242.

Munnell, A.H., Tootell, G.M., Browne, L.E. and McEneaney, J., 1996. Mortgage lending in Boston: Interpreting HMDA data. *The American Economic Review*, pp.25-53.

Naoi, M., Seko, M. and Sumita, K., 2009. Earthquake risk and housing prices in Japan: Evidence before and after massive earthquakes. *Regional Science and Urban Economics*, *39*(6), pp.658-669.

Nelson, J.P., 1981. Three Mile Island and residential property values: Empirical analysis and policy implications. *Land Economics*, *57*(3), pp.363-372.

Pope, J.C., 2008. Do seller disclosures affect property values? Buyer information and the hedonic model. *Land Economics*, *84*(4), pp.551-572.

Purnanandam, A., 2010. Originate-to-distribute model and the subprime mortgage crisis. *The Review of Financial Studies*, *24*(6), pp.1881-1915.

Railroad Commission of Texas. 2017. Pipeline Safety. Retrieved from http://www.rrc.state.tx.us/pipeline-safety/ on March 15, 2017

Rajan, U., Seru, A. and Vig, V., 2015. The failure of models that predict failure: Distance, incentives, and defaults. *Journal of Financial Economics*, *115*(2), pp.237-260.

Sigel, G.A.B.R.I.E.L.L.E. and Cane, S.M., 2007. Lender Liability Under Environmental Laws for Real Estate and corporate transactions. *Environmental Law In Illinois Corporate And Real Estate Transactions, IICLE*, pp.4-21.

Simkovic, M., 2013. Competition and crisis in mortgage securitization. *Ind. LJ*, *88*, p.213.

Simons, R.A., 1999. The effect of pipeline ruptures on noncontaminated residential easement-holding property in Fairfax County. *Appraisal Journal*, *67*, pp.255-263.

Simons, R.A., Winson-Geideman, K. and Mikelbank, B.A., 2001. The effects of an oil pipeline rupture on single-family house prices. *The Appraisal Journal*, *69*(4), p.410.

Wilson, A.R. and Alarcon, A.R., 1997. Lender attitudes toward source and nonsource impaired property mortgages. *The Appraisal Journal*, *65*(4), p.396.

Worzala, E.M. and Kinnard, W.N., 1997. Investor & Lender Reactions to Alternative Sources of Contamination. *Real Estate Issues*, *22*, pp.42-48.

Xu, Y. and Zhang, J., 2013. Nonlocal mortgage lending and the secondary market involvement. *Journal of Real Estate Literature*, *20*(2), pp.307-322.

CHAPTER 3

DO UNCONVENTIONAL OIL AND GAS WELLS GENERATE MORE WASTEWATER? A LIFESPAN PERSPECTIVE

3.1 Introduction

Wastewater disposal has always been a critical issue in the oil and gas industry due to the potential risks of causing environmental contamination. The chemical constituents identified in oil and gas wastewater could lead to mortality of aquatic species and severe damage to ground vegetation and human health when present in high concentrations (Veil et al., 2004). Unconventional oil and gas production combines fracking and horizontal drilling to extract shale resources, which involves more complex chemicals and produces more hazardous wastewater compared with conventional production methods (Orem et al., 2017). From 2003 to 2016, the annual total wastewater produced by horizontal wells in shale formations ("UOG wells") increased by 13.6 times while the volume of vertical wells ("COG wells") remained constant over the period (Top left panel of Figure 1). As the rapid growth of UOG wastewater production has outpaced the handling and disposal capacity in many regions (Lutz et al., 2013), the untreated and inadequately treated wastewater not only contaminates soil, surface water, and shallow groundwater near the drilling locations but also threatens the drinking water quality downstream of the wastewater disposal sites (Wilson and Van Briesen, 2013; States et al, 2013; Vengosh et al., 2014). In the long term, the large volume of wastewater produced by UOG wells may further lead to radium accumulation and persist as environmental and health hazards by gradually releasing toxic elements and radiation in the impacted areas (Warner et al., 2013a; Vengosh et al., 2014).

Although oil and gas wastewater has long been controversial because of its sheer volume and the associated environmental risks, comprehensive and accurate quantification of the impacts of unconventional techniques on wastewater production has been absent to date. To understand the wastewater production pattern of UOG wells, it is first worth noting that the increased total volume coincided with a rising number of UOG wells nationwide, which has increased from 4,882 in 2003 to 73,410 in 2016 (Top right panel of Figure 1). Moreover, the average per-well UOG wastewater has also increased rapidly since the late 2000s and exceeded that of COG wells by 2014 (Bottom left panel of Figure 1). Although a naïve comparison between the two types of wells suggests that on average a UOG well produces more wastewater than a COG one, the conclusion could be misleading since it does not take into account well-level heterogeneities in resource type, energy production, and production age. In particular, engineering evidence shows that UOG wastewater usually drops rapidly in volume after the initial production stage and then remains at a considerably lower daily flow rate throughout the rest of life (Veil, 2015; Kondash, Albright, and Vengosh, 2017). Thus, a well's production age is an essential factor to understand its wastewater production pattern. As is illustrated in the bottom right panel of Figure 1, while new wells that just started energy production in each year only accounted for about 5% among COG wells, their share among UOG wells was much higher and volatile, ranging from the peak of 31% in 2007 to 4% in 2016. Moreover, as the fracking technology has been improving continuously over time, simply pooling wells of different cohorts may confound the distinct production pattern of each cohort. Above all, to discern the precise impact of fracking on wastewater production requires an econometric analysis to account for the lifespan pattern, the cohort pattern, and the heterogeneity in other characteristics of a well.

In this study, we examine how fracking and horizontal drilling in UOG production has affected the pattern of wastewater production both contemporaneously and cumulatively over the lifespan of a well. We conduct an empirical analysis using proprietary panel data for 49,050 onshore oil and gas wells in the U.S. from 2008 through 2016. One challenge in identifying the use of the fracking technique by UOG wells is the lack of well-level data on the type of technology used for a national sample of wells. To circumvent this problem, we make use of each well's drill direction (horizontal or vertical) and geographic location (shale or non-shale) as an indicator of the use of fracking technology. Exploiting the fact that fracking is widely combined with horizontal drilling to extract energy from shale formations, we infer that horizontal wells in shale formations are UOG wells. We then define COG wells as all vertical wells regardless of location, which could be on either shale or non-shale formations.

Our study provides the first econometric analysis of the effects of unconventional stimulation and drilling techniques on wastewater production, using a comprehensive sample of wells with detailed geographic information and lifespan observations. We estimate the cohort- and age-varying differences in wastewater quantity and efficiency between UOG and COG wells for each cohort year from 2008 to 2016, controlling for wells' resource type, the dominant hydrocarbon type, energy production, county-level permanent disparities, state-specific time trends and business model of different operators. We find that UOG wells starting production before 2010 produced a statistically similar or a significantly lower amount of wastewater contemporaneously compared to COG wells throughout their entire lifespan. We further show that the difference between the two types of wells had reversed since 2010 when UOG wells, on average, started to produce more wastewater than COG wells in the first six months of their lifespan. The cohort-specific analysis shows that the contemporaneous gap in the initial months was larger in more recent cohorts, which indicates that the prevalence of the fracking technology over time has led wastewater production by UOG wells to be increasingly front-loaded. Regardless of cohort years, the estimated difference in contemporaneous wastewater production between the two types of wells always decreased with age in magnitude and turned statistically insignificant eventually, which indicates that both types of wells produced a statistically similar amount of wastewater as wells aged.

We also demonstrate similar lifespan and cohort patterns in cumulative wastewater volume of the two types of wells. The estimation results show that the cumulative wastewater produced by UOG wells that started production after 2010 was significantly higher than that of COG wells during their initial stage of production. However, the estimated difference tended to diminish with age and become significantly negative eventually. Furthermore, the "breakeven" point at which the cumulative wastewater production of UOG wells equaled that of COG wells occurred at an older age in more recent cohorts. For example, the positive difference between UOG and COG wells had turned negative since the 42^{nd} month in the 2010 cohort in contrast to the 47^{th} month in the 2011 cohort. For younger cohorts, the "breakeven" points were not yet achieved within the timeframe of our sample. These results indicate that the cumulative wastewater production of UOG wells may ultimately be lower than that for COG wells despite the growingly intensive wastewater production in the initial years of a well's lifespan.

Finally, the estimation results show that the wastewater-to-energy ratio of UOG wells was always significantly lower than that of COG wells. Although the difference was enlarged in magnitude in younger cohorts, it did not vary with a well's age as production continued, which suggests that the higher amount of wastewater in the initial production stage was accompanied by a proportionally larger volume of energy production. Our results also indicate that, relative to COG wells, UOG wells were growing more efficient over time measured by a lower wastewater-to-energy ratio in more recent cohorts. Our findings challenge the myths that fracking increases the overall oil and gas wastewater production and shows that the impact of fracking on wastewater production is nuanced. The real concern lies in the highly intensive production in the initial few years after a well begins production but not in the lifespan cumulative volume or the production efficiency.

This study extends the literature on the effects of fracking on water resources. Existing studies focus on the impacts of fracking on the quality of both surface water and groundwater. A rich body of engineering studies examine the chemical composition of fracking wastewater directly (Osborn et al. 2011; Warner et al. 2013a, 2013b, 2014; Darrah et al. 2014; Harkness et al., 2015), while ecologists and environmental economists also discuss the fracking-related water quality issue (Entrekin et al., 2011; Papoulias and Velasco 2013; Hitaj et al., 2014; Muehlenbachs and Olmstead, 2014). Some econometric studies use hedonic price models to show homebuyers' perceived risks of water contamination through a lower marginal willingness to pay to live near the drilling wells (Gopalakrishnan and Klaiber, 2013; Muehlenbachs et al., 2015).

Several studies have also estimated the effects of fracking on water quantity, which includes both freshwater as the input and wastewater as the byproduct of the fracking process. The focus of these studies has been centered on the water-use efficiency of fracking relative to other extraction methods (Goodwin et al., 2014; Kondash and Vengosh, 2015), the potential water stress issue in local areas (Freyman, 2014; Gallegos et al., 2015; Horner et al., 2016), and the challenges posed to regional wastewater disposal and treatment capacity (Lutz et al., 2013; Nicot et al., 2014). However, existing studies are limited to regional case studies of isolated shale basins. By contrast, our study uses a comprehensive national sample of wells and accounts for heterogeneities in well configuration and local geology. Our study also differs from previous research by considering the age-varying production pattern over a well's lifespan and the distinct production patterns across different cohorts.

3.2 Background

In UOG production, fracking serves as a recovery stimulation process to enhance the productivity of oil and gas wells. Before the innovation of horizontal drilling, fracking used to be applied to conventional geological formations and mature fields of low productivity (Morton, 2013; Wang and Krupnick, 2013). Horizontal drilling differs from vertical drilling by extending the range of fracking sideways along a rock formation rather than contacting it vertically, which significantly increases the area of reservoirs where hydrocarbons can be extracted and thus gives operators more economic incentives to use the fracking technology (Morton, 2013).[1] In 2002, a merger between Mitchell Energy and Devon Energy expedited the application of horizontal drilling, which led to a rapid increase in the combined use of the new drilling technique and fracking (Morton, 2013; Wang and Krupnick, 2013). The combination of the two technologies makes it technically and economically feasible to extract fossil fuels trapped in shale formations whose permeability is extremely low (Lutz et al., 2013).

The widespread use of fracking significantly affects the life cycle of wastewater production in the oil and gas industry. In UOG production, fracking occurs within a few days or weeks right after a well's drilling and completion phase during which up to six million gallons of water is

[1] Horizontal drilling discussed in this study also includes directional drilling. Thus, horizontal wells defined in this study refers to wells that are drilled either horizontally or at an angle relative to vertical wells.

injected into each well (Finkel, 2015; U.S. Environmental Protection Agency (EPA), 2016). The upfront water requirement for fracking leads to an extremely high rate of initial flowback, with up to 70 percent of the fluid discharging back out of the well shortly after the process (EPA, 2015). After the first few weeks, the volume of wastewater returning to the surface diminishes to a considerably lower level. Wastewater containing both residual fracking fluids and natural formation water is generated continuously but at a much lower flow rate as energy is produced over the lifespan of the well (Veil, 2015). By contrast, wastewater production of COG wells tends to be modest initially and may increase over time when water is used to enhance the energy productivity of mature wells.

3.3 Empirical Method

3.3.1 Difference-in-difference-in-differences Model

To test whether UOG and COG wells have different lifecycles of wastewater production, we estimate the following equation to find out how the contemporaneous and cumulative wastewater production of the two types of wells varies with production age while controlling for other well-level characteristics.

(1) $$y_{it} = \beta_0 + \beta_1 shale_i + \beta_2 gas_i + \beta_3 \log(energy_{it}) + \sigma uog_i + \sum_{k=1}^{K} \sum_{j=1}^{J} \sigma_{jk} uog_i * A_{ij} * C_{ik} + county_i + state_i * year_t + state_i * operator_i + \varepsilon_{it}$$

where i indexes well and t indexes month. The variable A_{ij} is an age dummy equal to one if well i has an age of j months. The variable C_{ik} is a cohort dummy that takes a value of one if well i starts production in the k^{th} cohort year. The coefficient σ measures the initial difference in wastewater between new UOG and COG wells after they start to produce oil and gas in the baseline cohort. The coefficients σ_{jk} captures the additional difference in wastewater production between the two types of wells in the k^{th} cohort year when they both have an age of j months.

The model includes a dummy indicating a well's resource type, $shale_i$, and a dummy indicating the dominant hydrocarbon type produced by the well, gas_i. It also controls for the effects of total oil and gas production by including $\log(energy_{it})$ as an explanatory variable. We further include the county fixed effects, $county_i$, to allow for county-level time-invariant differences in geology and regulations, the state by year fixed effects, $state_i * year_t$, to control for state-specific time trends in oil price, and the state by operator fixed effects, $state_i * operator_i$, to consider the state-specific business model of different operators.

3.3.2 Data and Descriptive

We conduct this analysis using a proprietary well-level dataset through an academic use agreement with the Drillinginfo, Inc. The Drillinginfo collects oil and gas production data for 30 states in the United States for each well and month. We focus on the fracking boom period of 2008-2016 to construct a sample of wells at the month level.[2] We retain the wells of 17 states in the U.S. that require oil and gas operators to report wastewater production.[3] We restrict our sample to active onshore oil and gas wells with identifiable drill direction type. As some operators did not report

[2] The US experienced the first jump in shale production in 2008 (US EIA, 2009). See more at https://www.eia.gov/dnav/ng/ng_prod_shalegas_s1_a.htm

[3] These 17 states include Alabama, Arkansas, California, Coloradio, Florida, Michigan, Mississippi, Montana, Nebraska, Nevada, New Mexico, New York, North Dakota, South Dakota, Texas, Utah, and Wyoming.

wastewater production for all the wells of certain states in specific years, we exclude these observations to avoid systematic data reporting error. We also exclude wells that did not report any wastewater production over the sampling period. Besides, since large-scale commercial horizontal drilling was not successful until the 1990s (King and Morehouse, 1993), we exclude horizontal and directional wells that were reported to start production before 1990. We finally exclude wells switching between active production and shutdown during the lifespan to retain a balanced sample of wells with continuous production.

We then project the geographic coordinates of each well onto the latest shale play shapefiles provided by the U.S. Energy Information Administration (EIA) to identify whether a well is in shale regions.[4] We define conventional wells as those with a drill direction of vertical and drilled in either shale or non-shale regions. We then define unconventional wells as those classified as horizontal or directional wells but located in shale regions only. Thereby, we exclude horizontal or directional wells drilled in non-shale regions, since oil and gas in these areas could be extracted with or without fracking with substantial uncertainty. If some UOG wells defined in this study did not use fracking or some COG wells did apply fracking in practice, then our estimated differences between the two types of wells would serve as a lower bound of the true impacts of unconventional techniques on wastewater production. Next, we define oil wells as those with an energy production type classification of Oil, and gas wells as those classified as Gas. We treat wells classified as "Oil&Gas" as either oil or gas wells depending on their dominant energy type during the sampling period. Finally, we convert each well's oil production in barrel (bbl) and

[4] The shapefiles are available at https://www.eia.gov/maps/maps.htm

gas production in thousand cubic feet (Mcf) to the energy equivalents in million British thermal units (mmBtu) to construct an aggregate measure of energy production.[5]

We define the cohort of a well by the calendar year when the well started energy production. We then calculate the well's age in month since the beginning of production. We trace each well's production starting from the first month after the initial month when its energy and wastewater production was observed for a full month for the first time.[6] For wells of cohort year t, we track their monthly wastewater and energy production up to $12 * (2016 - t)$ months so that all the wells of the cohort can be observed over the same number of months. Otherwise, wells starting production at the beginning of a cohort year would be observed over a longer lifespan compared to those starting production approaching the end of the year. Finally, we drop outlier wells whose wastewater production or wastewater-to-energy ratio was once greater than the 99th percentile and those whose energy production was once less than the 1st percentile of wells within the same production type, age, and cohort group.

[5] According to the energy equivalents conversion criteria published by the EIA, 1000 cubic feet (mcf) of natural gas equals 1.037 mmBtu and a barrel of oil (bbl) equals 5.717 mmBtu. See more at https://www.eia.gov/energyexplained/index.cfm/index.cfm?page=about_energy_conversion_calculator

[6] We find that in our sample wells started production almost equally likely on any day of a month. Thus, a well's total production in the initial month only equals to about a half of a full month's quantity. In addition, due to the existence of some "late-stage" wells that started production at the end of the initial month, the average energy and wastewater production we can observe in each month is higher than the true value because of the downwards trends of production. However, taking difference between UOG and COG wells across ages and cohorts can cancel out the bias.

Table 1 reports the summary statistics of the month-level sample over the period of 2008-2016, which summarizes the well-level characteristics by UOG and COG wells. On average, UOG wells produced a higher amount of both wastewater and energy but had a lower wastewater-to-water ratio (0.119 bbl/mmBtu) compared to that of COG wells (0.415 bbl/mmBtu). UOG wells were younger than COG wells measured by the average production age (30 months versus 35 months). Moreover, while gas wells were the dominant production type among UOG wells (64.7%), they consist of 40.1% of COG wells. Also, while 100% UOG wells were in shale regions by definition, the shale share in COG wells was about 54%. Finally, while the number of UOG wells were almost evenly distributed over the cohort years, the count of COG wells declined over cohorts.

3.4 Empirical Results

In this section, we report the estimated differences in wastewater production between UOG and COG wells over their lifespan using Equation (1). Table 2 reports the estimation results, where the dependent variables from Columns 1-3 are the log form of contemporaneous wastewater, the log form of cumulative wastewater, and the wastewater-to-energy ratio, respectively. We cluster the standard errors at the well level. Using the estimates obtained from each regression, we run a series of linear hypothesis tests on the null hypothesis of $\sigma + \sigma_{jk} = 0$ to estimate the cohort- and age-specific difference in each dependent variable between UOG and COG wells at age j in cohort k. For cohort in year t from 2008 to 2016, the tests generate $12 * (2016 - t)$ estimates as each well in the cohort can be observed over the same number of months.

3.4.1 Contemporaneous wastewater production

Figure 2 first plots the coefficients and the 95% confidence intervals of the estimated cohort- and age-specific difference in contemporaneous wastewater production between UOG and COG wells over their lifespan. A general pattern emerges from the figure that within a cohort the difference in contemporaneous wastewater production between the two types of wells always decreased with wells' production age while across cohorts the initial difference switched from negative to positive with the magnitude being enlarged continuously.

Specifically, among wells starting production before 2010, UOG wells produced a statistically similar or a significantly lower amount of wastewater than COG wells over their lifespan. In the cohort years of 2010 and 2011, the difference turned positive in the first six months of production, which were weakly significant, but remained significantly negative for the rest of lifespan. Since 2012, the positive difference has covered an increasingly more extended initial portion of lifespan at the significance level of 1%. The changing production pattern indicates that UOG wells have produced a higher amount of wastewater than COG wells in the initial stage of lifespan since then.

This pattern persisted and became even more prominent for wells in younger cohorts, where the magnitude of the initial difference in wastewater production has increased continuously. Among wells in the 2012 cohort, for instance, the difference in wastewater production of the first month in lifespan was 0.25 at the significance level of 1%, which means that the initial volume of UOG wastewater became significantly higher than that of COG wells by 25%. Compared with the estimated difference of 0.009 for wells at the same age in the 2011 cohort, the increased initial difference indicates that wastewater production became more front-loaded due to the upfront water requirements for fracking in UOG production. In cohorts younger than 2012, the initial difference between the two types of wells was even larger in magnitude. By cohort 2015, UOG wells had produced a significantly higher volume of wastewater than COG wells by over 70% in the first production month at the 1% significance level. The ever-growing initial difference across cohorts may result from the fact that a higher portion of new UOG wells defined in this study began to apply fracking in production as the technology matured and became more prevalent over time.

Although the post-2011 cohorts displayed a positive initial difference in production between UOG and COG wells, the gap tends to close as wells age. In the 2012 and 2013 cohorts, after roughly two years of production, the estimated difference dropped in magnitude over age and turned insignificant, which indicates a statistically similar amount of wastewater produced by the two kinds of wells when they aged. As we can only observe a limited timeframe for both types of wells in younger cohorts, we cannot find the closure of the production gap between them for the last two cohorts. However, the cohort trends project that the contemporaneous gap in the last two cohorts declined even faster than before. Given the higher initial contemporaneous difference and the faster decline in the production gap between the two types of wells, the relative difference in cumulative wastewater production remains unclear. Thus, we further investigate the cumulative difference between UOG and COG wells in the next subsection.

3.4.2 Cumulative wastewater production

Figure 3 plots the coefficients and the 95% confidence intervals of the estimated cohort- and age-specific difference in contemporaneous wastewater production between UOG and COG wells. The figure illustrates that, while the cumulative wastewater produced by UOG wells was significantly lower than that by COG wells throughout their lifespan in cohorts before 2010, it was not the case for younger cohorts in which the cumulative difference between the two types of wells turned from positive to negative with the "breakeven" point being gradually delayed across cohorts.

Specifically, for the 2008 and 2009 cohorts, the cumulative wastewater produced by UOG wells was statistically significantly lower than that by COG wells by 15% and 24% by the end of the observable lifespan respectively. These results are consistent with the findings that UOG wells produced a significantly lower contemporaneous amount of wastewater compared to COG wells in these cohorts. In the 2010 cohort, the estimated difference in cumulative wastewater turned positive in the initial months of production, which aligns with the findings in the contemporaneous difference that UOG wastewater production became more front-loaded in lifespan starting from this cohort. The estimation results show that the cumulative wastewater production by UOG wells exceeded that by COG wells by 8% at the first 12 months. We find consistent evidence that the initial difference between UOG and COG wells increased in magnitude continuously in younger cohorts. For example, at the end of the first year, while the cumulative gap between the two types of wells was 7.8% for the 2011 cohort, it rose to 86.2% for the 2015 cohort.

However, the cumulative difference between UOG and COG wells declined in magnitude and finally turned negative as production continued over the lifespan of the well. In the 2010 cohort, for instance, this estimated difference turned negative in the 42^{nd} month of production. By the end of the observable lifespan in the cohort, i.e., the 72^{nd} month, the cumulative wastewater by UOG wells was significantly lower than that produced by COG wells by 13.2% at the significance level of 1%. The findings indicate that, despite a higher cumulative volume at the initial stage of production, the cumulative lifespan production of UOG wells was significantly lower than that of COG wells. Moreover, we find that the "breakeven" point occurred later in more recent cohorts. While the estimated difference turned negative in the 47^{th} month for the 2011 cohort, they stayed positive throughout the entire observable lifespan for the 2012 cohort though the difference was decreasing toward zero. The delayed "breakeven" point is probably caused by the fact that UOG wells produced an increasingly higher amount of wastewater at the initial stage of lifespan in younger cohorts.

3.4.3 Wastewater-to-energy ratio

Finally, we test the cohort- and age-varying difference in wastewater-to-energy ratio between UOG and COG wells. We run a model specification excluding the variable of $\log(energy_{it})$ from Equation (1). Using the parameter estimates obtained, Figure 4 plots the coefficients and the 95% confidence intervals for the estimates of the linear hypothesis test results of $\sigma + \sigma_{jk} = 0$ at age j in cohort k. Overall, the results demonstrate that UOG wells were significantly more efficient than COG wells, as measured by the amount of wastewater per unit of energy produced, regardless of wells' production age or cohort year. For example, among wells in the 2008 cohort, the ratio for UOG wells was lower than that of COG wells by -0.11 bbl/mmBtu at the significance level of 1%. The difference in ratio persisted over the entire lifespan of wells' production, which indicates that wastewater volume always varied proportionally with energy production. Thus, we infer that the energy production of UOG wells was also concentrated within the initial few years of lifespan, which is similar to the distribution of wastewater production by these wells.

For wells in younger cohorts, the difference in efficiency between the two types of wells was further enlarged in magnitude. In the 2015 cohort, the ratio for UOG wells was significantly lower than that for COG wells by -0.3 bbl/mmBtu. The findings indicate that, while both wastewater and energy production of UOG wells increased across cohorts due to the prevalence of fracking over time, energy production increased to a greater extent than wastewater output, which led UOG production to be increasingly more efficient than COG wells concerning wastewater generation.

3.5 Conclusions

This study uses a comprehensive well-level dataset to estimate the impacts of unconventional production techniques of fracking and horizontal drilling on oil and gas wastewater production. We distinguish unconventional wells from conventional ones by wells' drill type and resource type given that fracking is typically used in conjunction with horizontal drilling to extract shale energy. We estimate the cohort- and age-specific differences in wastewater production between UOG and COG wells to show how unconventional energy production has affected the quantity of wastewater produced by the two types of wells regarding the contemporaneous amount, the cumulative volume, and the water-to-energy ratio over a well's lifespan and across different cohorts.

This study contributes to the public discussions about the environmental costs and the management challenges presented by the fracking technique that has gained increased prevalence in the oil and gas industry. While it is perceived that shale energy development is relatively more wastewater intensive, we show the importance of taking a lifespan perspective in assessing the production of wastewater. Contrary to the prevailing perception, UOG wells turn out to be more efficient in production with a significantly lower wastewater-to-energy ratio than COG wells. The higher production efficiency implies a lower operating cost associated with each unit of energy extracted borne by energy operators. Moreover, we show that the cumulative wastewater amount produced by an average UOG well could be significantly lower than that of COG wells by the end of their lifespan. These findings point out that the rapid growth in wastewater generation since the recent energy boom has been a consequence of the growth in the scale of production instead of the stimulation and drilling methods applied.

Despite a lower per-well lifespan volume, we show that the wastewater production by UOG wells has become more concentrated and occurs upfront during the initial stage of a well's life. We also find that there is a gradual delay in the point of reversal during a well's life at which the difference in cumulative production between UOG and COG wells turns negative in more recent cohorts. Furthermore, considering the immense size of the ongoing fracking activities, UOG wells still generate a much higher overall amount of wastewater than COG wells, which may impose massive pressure on local infrastructure and wastewater treatment facilities especially when a large number of wells are in their initial stage of production concurrently. The issue is even more critical given that existing wastewater disposal capacity is being quickly saturated while only a small share of shale resources has been extracted to date (U.S. EIA, 2015). Through recognizing the cyclical pattern of wastewater production of UOG wells over their lifespan, local governments may design more effective regulations and management practices governing the permitting of new drilling activity. Regulations can be implemented on the number of new wells that can be drilled at a point in time to manage the outflow of wastewater and reduce the total volume of wastewater being disposed of within a given timeframe. Moreover, permitting of new wells should also incorporate the age distribution of existing wells along with the regional wastewater treatment capacity to help alleviate the impacts of fracking on local infrastructure and water quality.

3.6 Tables and Figures

Table 3.1 Summary Statistics

	UOG Wells		COG Wells	
	Mean	Std. Dev.	Mean	Std. Dev.
log (1+produced water) (bbl)	6.208	1.975	5.852	2.267
log (energy production) (mmBtu)	9.462	1.110	8.214	1.234
Wastewater-to-energy ratio (bbl/mmBtu)	0.119	0.287	0.415	0.965
Production age (month)	29.901	22.162	34.681	24.045
Gas well	0.647	0.478	0.401	0.490
Shale	1.000	0.000	0.536	0.499
Cohort of 2008	0.156	0.362	0.278	0.448
Cohort of 2009	0.109	0.312	0.155	0.362
Cohort of 2010	0.142	0.349	0.163	0.370
Cohort of 2011	0.153	0.360	0.145	0.353
Cohort of 2012	0.163	0.370	0.119	0.324
Cohort of 2013	0.137	0.344	0.074	0.262
Cohort of 2014	0.102	0.302	0.050	0.219
Cohort of 2015	0.039	0.192	0.015	0.120
N		1,463,385		871,439

Notes: The table reports the summary statistics of the 2008-2016 monthly well-level sample.

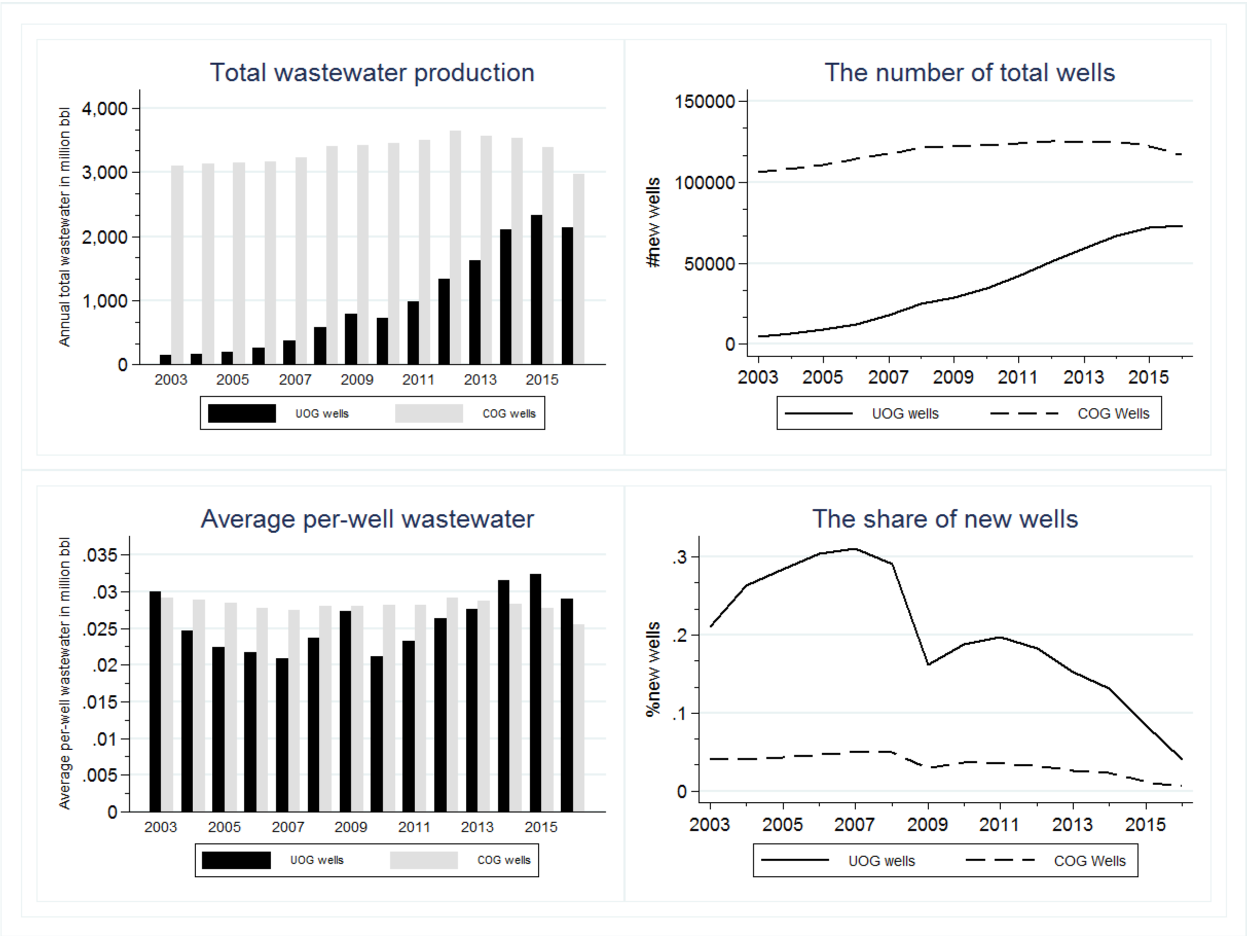
Table 3.2 Age-varying Difference in Wastewater Volume and Efficiency

	(1) Log(1+wastewater)	(2) Log (1+cumulative wastewater)	(3) Wastewater-to-energy ratio
Shale region	-0.005	0.022	-0.110***
	(0.059)	(0.061)	(0.030)
Gas well	-0.887***	-0.936***	-0.107***
	(0.050)	(0.054)	(0.012)
Log(energy)	0.704***	0.460***	
	(0.007)	(0.007)	
UOG dummy (σ)	Yes[i]	Yes[ii]	Yes[iii]
UOG * Age * Cohort dummies (σ_{jk})	Yes[i]	Yes[ii]	Yes[iii]
County FE	Yes	Yes	Yes
State by year FE	Yes	Yes	Yes
Operator by state FE	Yes	Yes	Yes
Observations	2,334,803	2,334,803	2,334,803
R-squared	0.589	0.608	0.377

Notes: The table reports the estimation results using the 2008-2016 monthly well-level sample. The dependent variables are listed as the column title. Standard errors clustered at the well level are reported in parentheses *** p<0.01, ** p<0.05, * p<0.1.

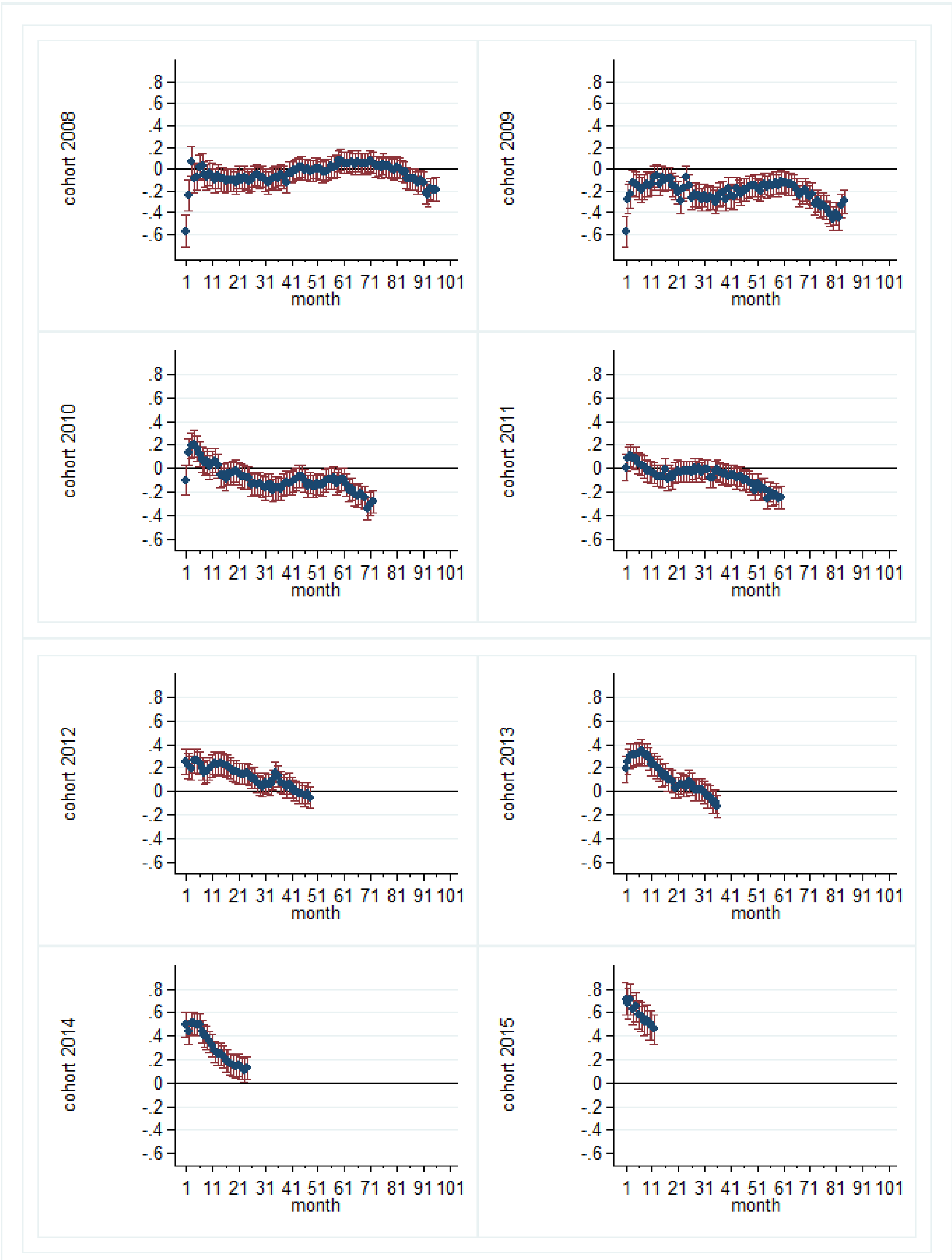
i. The estimates of $\sigma + \sigma_{jk}$ are plotted in Figure 2.
ii. The estimates of $\sigma + \sigma_{jk}$ are plotted in Figure 3.
iii. The estimates of $\sigma + \sigma_{jk}$ are plotted in Figure 4.

Figure 3.1 The U.S. UOG and COG Wastewater Production, Well Counts, and New Wells' Share



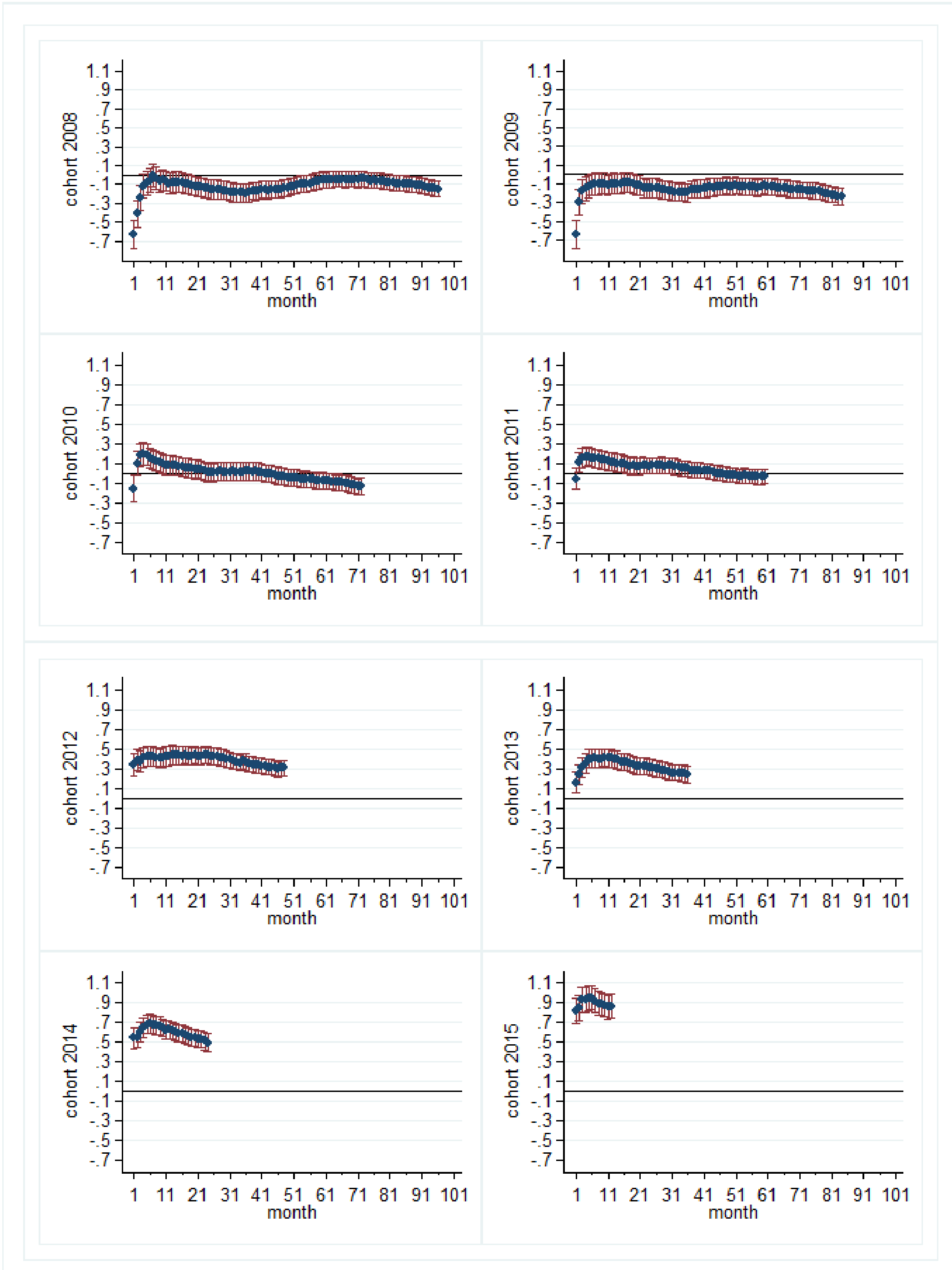
Data sources: Summarized by authors using the data from the DrillingInfo.
Notes: UOG wells refer to unconventional oil and gas wells, which are defined as horizontal wells in shale formations. COG wells refer to conventional oil and gas wells, which are defined as vertical wells regardless of location.

Figure 3.2 Log Difference in Contemporaneous Wastewater Production between UOG and COG Wells



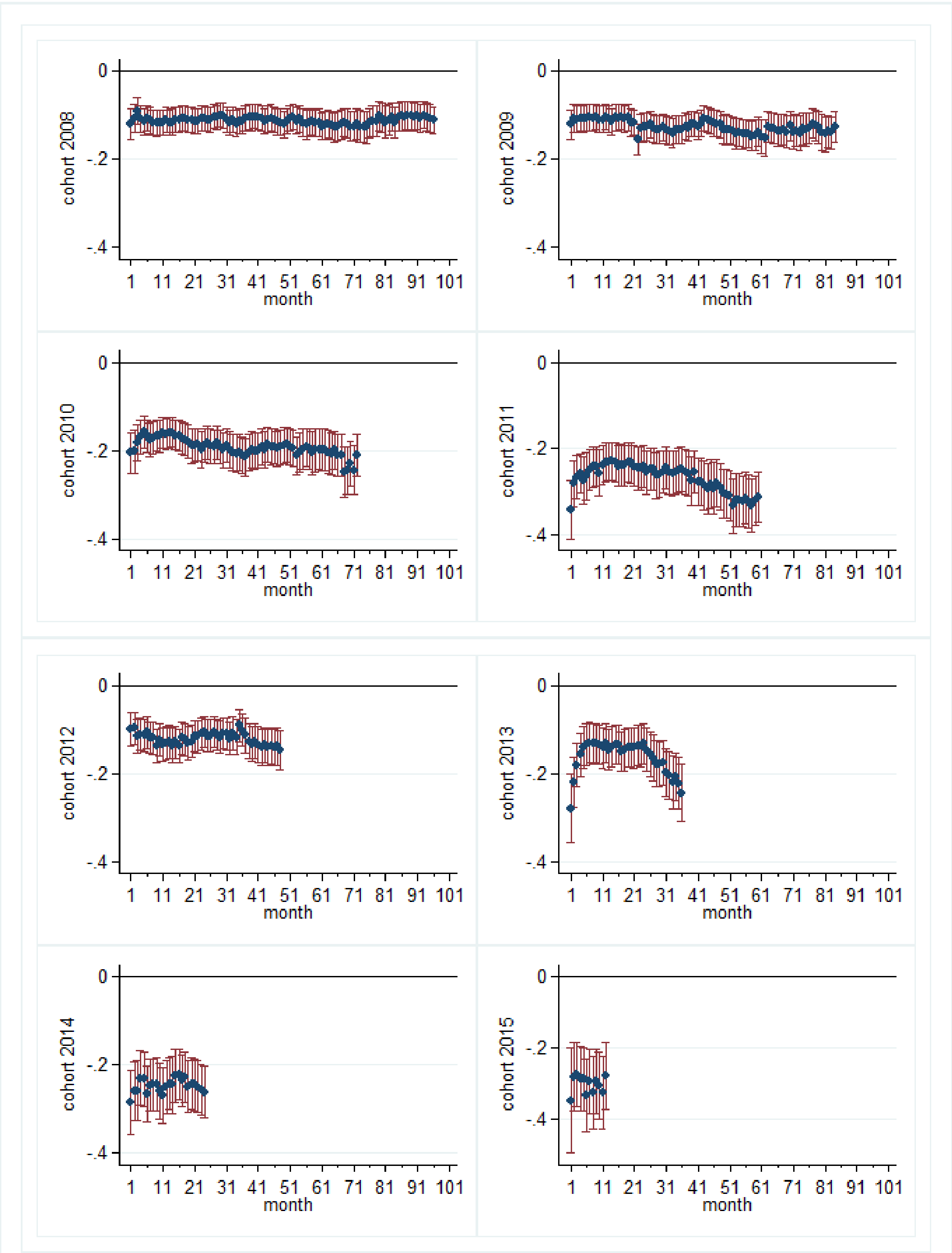
Notes: The figure plots the coefficient estimates and the 95% confidence intervals for the cohort- and age-varying log difference in contemporaneous wastewater production between UOG and COG wells. The model controls for a well's resource type, the dominant hydrocarbon type, log of energy production, the county fixed effects, the state-specific time trends, and the state- and operator-specific business model. The coefficient is interpreted as the difference in percentage terms ((UOG-COG)/COG).

Figure 3.3 Log Difference in Cumulative Wastewater Production between UOG and COG Wells



Notes: The figure plots the coefficient estimates and the 95% confidence intervals for the cohort- and age-varying difference in cumulative wastewater production between UOG and COG wells. The model controls for a well's resource type, the dominant hydrocarbon type, log of energy production, the county fixed effects, the state-specific time trends, and the state- and operator-specific business model. The coefficient is interpreted as the difference in percentage terms ((UOG-COG)/COG).

Figure 3.4: Difference in Wastewater-to-energy Ratio (in bbl/mmBtu) between UOG and COG Wells



Notes: The figure plots the coefficient estimates and the 95% confidence intervals for the cohort- and age-varying difference in wastewater-to-energy ratio between UOG and COG wells. The model controls for a well's resource type, the dominant hydrocarbon type, the county fixed effects, the state-specific time trends, and the state- and operator-specific business model. The coefficient is interpreted as the difference in barrels of wastewater per mmBtu of energy produced (UOG-COG).

3.7 References

Darrah, T.H., Vengosh, A., Jackson, R.B., Warner, N.R. and Poreda, R.J., 2014. Noble gases identify the mechanisms of fugitive gas contamination in drinking-water wells overlying the Marcellus and Barnett Shales. *Proceedings of the National Academy of Sciences*, *111*(39), pp.14076-14081.

Finkel, Madelon L. 2015. *The human and environmental impact of fracking: How fracturing shale for gas affects us and our world*. Santa Barbara: ABC-CLIO.

Entrekin, S., Evans-White, M., Johnson, B. and Hagenbuch, E., 2011. Rapid expansion of natural gas development poses a threat to surface waters. *Frontiers in Ecology and the Environment*, *9*(9), pp.503-511.

Freyman, M., 2014. Hydraulic fracturing and water stress: water demand by the numbers. A Ceres Report.

Gallegos, T.J., Varela, B.A., Haines, S.S. and Engle, M.A., 2015. Hydraulic fracturing water use variability in the United States and potential environmental implications. *Water Resources Research*, *51*(7), pp.5839-5845.

Goodwin, S., Carlson, K., Knox, K., Douglas, C. and Rein, L., 2014. Water intensity assessment of shale gas resources in the Wattenberg field in northeastern Colorado. *Environmental science & technology*, *48*(10), pp.5991-5995.

Gopalakrishnan, S. and Klaiber, H.A., 2013. Is the shale energy boom a bust for nearby residents? Evidence from housing values in Pennsylvania. *American Journal of Agricultural Economics*, *96*(1), pp.43-66.

Harkness, J.S., Dwyer, G.S., Warner, N.R., Parker, K.M., Mitch, W.A. and Vengosh, A., 2015. Iodide, bromide, and ammonium in hydraulic fracturing and oil and gas wastewaters: environmental implications. *Environmental science & technology*, *49*(3), pp.1955-1963.

Hitaj, C., Boslett, A. and Weber, J.G., 2014. Shale development and agriculture. *Choices*, *29*(4), pp.1-7.

Horner, R.M., Harto, C.B., Jackson, R.B., Lowry, E.R., Brandt, A.R., Yeskoo, T.W., Murphy, D.J. and Clark, C.E., 2016. Water use and management in the Bakken shale oil play in North Dakota. *Environmental science & technology*, *50*(6), pp.3275-3282.

Kondash, A.J., Albright, E. and Vengosh, A., 2017. Quantity of flowback and produced waters from unconventional oil and gas exploration. *Science of the Total Environment*, *574*, pp.314-321.

Kondash, A. and Vengosh, A., 2015. Water footprint of hydraulic fracturing. *Environmental Science & Technology Letters*, *2*(10), pp.276-280.

Lutz, B.D., Lewis, A.N. and Doyle, M.W., 2013. Generation, transport, and disposal of wastewater associated with Marcellus Shale gas development. *Water Resources Research*, *49*(2), pp.647-656.

Morton, M.Q., 2013. Unlocking the Earth-A Short History of Hydraulic Fracturing. *GEOExPro*, *10*(6), pp.86-90.

Muehlenbachs, L. and Olmstead, S., 2014. Hydraulic fracturing and water resources. *Choices*, *29*(4).

Muehlenbachs, L., Spiller, E. and Timmins, C., 2015. The housing market impacts of shale gas development. *American Economic Review*, *105*(12), pp.3633-59.

Nicot, J.P., Scanlon, B.R., Reedy, R.C. and Costley, R.A., 2014. Source and fate of hydraulic fracturing water in the Barnett Shale: a historical perspective. *Environmental science & technology*, *48*(4), pp.2464-2471.

Orem, W., Varonka, M., Crosby, L., Haase, K., Loftin, K., Hladik, M., Akob, D.M., Tatu, C., Mumford, A., Jaeschke, J. and Bates, A., 2017. Organic geochemistry and toxicology of a stream impacted by unconventional oil and gas wastewater disposal operations. *Applied Geochemistry*, *80*, pp.155-167.

Osborn, S.G., Vengosh, A., Warner, N.R. and Jackson, R.B., 2011. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *proceedings of the National Academy of Sciences*, *108*(20), pp.8172-8176.

Papoulias, D.M. and Velasco, A.L., 2013. Histopathological analysis of fish from Acorn Fork Creek, Kentucky, exposed to hydraulic fracturing fluid releases. *Southeastern Naturalist*, *12*(4), pp.92-111.

States, S., Cyprych, G., Stoner, M., Wydra, F., Kuchta, J., Monnell, J. and Casson, L., 2013. Marcellus Shale drilling and brominated THMs in Pittsburgh, Pa., drinking water. *Journal: American Water Works Association*, *105*(8).

King, R.F. and Morehouse, D., 1993. Drilling sideways—a review of horizontal well technology and its domestic application. *Energy Information Administration Technical Report*.

U.S. Energy Information Administration, 2015. Annual energy outlook 2015 with projections to 2040. Office of Integrated and International Energy Analysis, U.S. Department of Energy, Washington, DC.

U.S. Environmental Protection Agency, 2015. Review of state and industry spill data: characterization of hydraulic fracturing-related spills. Office of Research and Development, Washington, DC. Retrieved from https://www.epa.gov/sites/production/files/2015-05/documents/hf_spills_report_final_5-12-15_508_km_sb.pdf

U.S. Environmental Protection Agency, 2016. Hydraulic fracturing for oil and gas: Impacts from the hydraulic fracturing water cycle on drinking water resources in the United States. Office of Research and Development, Washington, DC. Retrieved from https://www.epa.gov/sites/production/files/2016-12/documents/hfdwa_executive_summary.pdf

Veil, J.A., Puder, M.G., Elcock, D. and Redweik Jr, R.J., 2004. *A white paper describing produced water from production of crude oil, natural gas, and coal bed methane* (No. ANL/EA/RP-112631). Argonne National Lab., IL (US).

Veil, J., 2015. US Produced Water Volumes and Management Practices in 2012; report prepared for the Groundwater Protection Council. *GWPC: Oklahoma City, OK*.

Vengosh, A., Jackson, R.B., Warner, N., Darrah, T.H. and Kondash, A., 2014. A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the United States. *Environmental science & technology*, *48*(15), pp.8334-8348.

Wang, Z. and Krupnick, A., 2013. A retrospective review of shale gas development in the United States: What led to the boom? Resources for the Future, RFF DP 13-12.

Warner, N.R., Christie, C.A., Jackson, R.B. and Vengosh, A., 2013a. Impacts of shale gas wastewater disposal on water quality in western Pennsylvania. *Environmental science & technology*, *47*(20), pp.11849-11857.

Warner, N.R., Kresse, T.M., Hays, P.D., Down, A., Karr, J.D., Jackson, R.B. and Vengosh, A., 2013b. Geochemical and isotopic variations in shallow groundwater in areas of the Fayetteville Shale development, north-central Arkansas. *Applied Geochemistry*, *35*, pp.207-220.

Warner, N.R., Darrah, T.H., Jackson, R.B., Millot, R., Kloppmann, W. and Vengosh, A., 2014. New tracers identify hydraulic fracturing fluids and accidental releases from oil and gas operations. *Environmental science & technology*, *48*(21), pp.12552-12560.

Wilson, J.M. and Van Briesen, J.M., 2013. Source water changes and energy extraction activities in the Monongahela River, 2009–2012. *Environmental science & technology*, *47*(21), pp.12575-12582.

CHAPTER 4

FRACCIDENTS: THE IMPACT OF FRACKING ON ROAD TRAFFIC DEATHS

4.1 Introduction

The increased prevalence of hydraulic fracturing, a.k.a. fracking, has brought about a nationwide shale oil and gas extraction boom since the late 2000s. Accompanied by the massive economic benefits, however, the fracking boomtowns also bear a growing burden on local services and infrastructure. Over the decade of 2006-2015, the national fatal traffic crashes decreased by 17%, while many boomtowns witnessed a surge in road fatalities.[7] Part of the spike might be explained by the growth in migrants and commuters in these towns since more traffic results in more accidents (Vickrey, 1968). However, a more concerning cause could be the heavy trucks servicing the energy sector, which give rise to more severe crashes and impose immense externalities on other road users (Li, 2012; Jacobsen, 2013; Anderson and Auffhammer, 2014; Muehlenbachs, Staubli, and Chu, 2017). Besides usual transport needs in drilling and production phases, fracking wells demand additional trucks to transport fracking fluid and liquid wastes, adding up to two to three times the truck trips required by conventional wells (Moss, 2008; Podulka and Podulka,

[7] We calculate the change in national fatal crashes using data from the National Highway Traffic Safety Administration (NHTSA) at https://www-fars.nhtsa.dot.gov/Main/index.aspx. Meanwhile, Fractracker Alliance (2014) reports that from 2009 to 2013 the drilling counties in North Dakota, West Virginia, Texas, Pennsylvania, and New Mexico all experienced faster increase in traffic fatalities compared to non-drilling counties within each state. See more at https://www.fractracker.org/2014/09/truck-counts/

2010; New York State Department of Transportation (NYSDOT), 2011). Thus, the truck traffic associated with fracking *per se* may have brought massive accident externalities to road safety.

This study estimates a fixed-effect Poisson model to disentangle the effects of fracking-related transport from that of confounding activities on the increased traffic crashes in energy boomtowns. We focus on a very local area surrounding each road segment. With road segment fixed effects, we effectively exploit the temporal variations in truck traffic within a road segment induced by the exogenous timing of fracking operations nearby to identify the impacts of fracking-induced water trucking. Fracking, usually combined with horizontal drilling, occurs within a few days at the end of well completion to release oil and gas from shale rocks (Drillinginfo, 2014; Finkel, 2015; Muehlenbachs, Staubli, and Chu, 2017). Trucks are in high demand both during and after the process to convey water, fracking materials, and liquid wastes, which account for over 90% of all the truck traffic from a well over its productive life (NYSDOT, 2011). Within a buffer area of a given road segment, we count the number of horizontal wells completed in the current month ("fracking wells") to capture the truck trips associated with fracking fluid injection. We then count the number of nearby horizontal wells one month after completion ("post-fracking wells") to measure the truck use related to fracking wastewater hauling. The identification strategy assumes that the timing of fracking is exogenous to the existing traffic status and that a road segment is more likely to be used by nearby wells than distant ones. We control for other traffic vehicles servicing wells' drilling and production phases using the number of spudding and producing wells respectively that include both horizontal and vertical wells. As monthly varying control variables at the road segment level rarely exist, we include county-specific time trends to control for the time-varying population and local policies and year-specific seasonality to control for market conditions and weather.

This study takes empirical evidence from the 2006-2014 oil and gas wells and fatal highway crashes in North Dakota, which experienced significant shale oil reserves development since 2006 (Nordeng, 2010). The estimation results of a fixed-effect Poisson model show a significant increase in fatal accidents at the post-fracking phase. Precisely, within a baseline buffer of six miles of a road segment, every additional post-fracking well increased the count of fatal crashes involving large trucks by 7.5%. Besides, the post-fracking activity also led to more severe accidents measured by a higher monetary loss per person.

We further show that in the baseline buffer each additional post-fracking well was associated with an additional 3,289 vehicles in the year when it was fracked but did not affect the accident rate. Although the traffic volume data are not available at the month level, we believe that the increased traffic counts were more likely to occur immediately after well completion rather than being evenly distributed across the year. We also show that wells' post-fracking transport significantly increased the incidence of daytime accidents during both rush and non-rush hours but not nighttime crashes. Our findings indicate that the increased fatal crashes were due to a higher traffic amount induced by fracking activities rather than a higher accident rate. Although we find weak evidence of a higher chance of drowsy driving with post-fracking operations nearby, we do not identify any significant association between well operations and other risky behaviors of the crash drivers.

This study expands the literature that explores the factors influencing the incidence and severity of road fatalities. Existing studies focus on various aspects of road safety, such as the amount of traffic (Romen and Shurtz, 2016), text messaging bans (Abouk and Adams, 2013), minimum wage (Adams, Blackburn, and Cotti, 2012), drunk driving (Levitt and Porter, 2011), compulsory insurance regulations and no-fault liability laws (Cohen and Dehejia, 2004), seat belt use (Cohen and Einav, 2003), and police enforcement (Deangelo and Hansen, 2014). A particularly active research area is the accident externality of heavyweight vehicles (Li, 2012; Jacobsen, 2013; Anderson and Auffhammer, 2014; Muehlenbachs, Staubli, and Chu, 2017). The study that is most related to ours is Muehlenbachs, Staubli, and Chu (2017), who estimate the traffic impacts of heavy trucks related to fracking using quarterly variations in traffic volume and crashes on the routes that connect wells, water withdrawal points, and wastewater disposal points. Our analysis differs from their study by exploiting more precise variations at the road segment-month level where the fracking-related traffic is proxied by the number of fracking and post-fracking wells in a buffer zone. We show that it is the first month after fracking that is subject to most externalities of truck transport, which complements Muehlenbachs, Staubli, and Chu (2017) who examine the traffic impacts arising from the whole season of fracking. While they find that fracking increased both truck and non-truck crashes but did not affect the severity of the accidents, we identify that fracking not only increased the count of large-truck crashes but also led to more fatalities and more critical injuries. Moreover, our driver-level analysis examines driver behaviors in shale energy extraction, which has not been investigated in the current studies.

4.2 Empirical Method

4.2.1 Operational Phases of A Fracking Well

Fracking is a stimulation technology intended to enhance the productivity of oil and gas wells. The process involves pumping a high-pressure mixture of sand, chemicals, and large volumes of water to prop open underground cracks and allow the flow of gas, oil, salt water, and fracking fluids to wells. Its application used to be limited to mature fields and conventional geological formations until the early 2000s when fracking started to be combined with horizontal drilling to explore shale resources (Morton, 2013; Wang and Krupnick, 2013). The rapid increase in the combined use of the two technologies prompted large-scale shale exploration in the Bakken Formation of North Dakota since 2006 (Nordeng, 2010).

Fracking is a water-intensive process – up to six million gallons of water is required to be injected into a wellhead as the base fluid (U.S. Environmental Protection Agency (EPA), 2016). Shortly after the process, up to 70 percent of the injected fluids discharge back out of the well together with naturally occurring water trapped in formations (EPA, 2015). Both fracking water supply and waste collection depend mainly on heavy trucks; hence, fracking generates a substantial freight demand during and after the process.

A conventional oil and gas well experiences three operational phases, including drilling, well completion, and production (DrillingInfo, 2014). Panel A of Figure 1 illustrates that trucks are needed to serve the stage-specific transport purpose throughout the entire lifespan of a well. Operators start drilling a well from removing rock, dirt and other sedimentary material with drill bits, which is called spud. The whole drilling phase lasts about 80 days (Muehlenbachs, Staubli, and Chu, 2017), during which multiple truck trips are needed to transport drilling rigs and other equipment. After finishing drilling a well, operators may wait for favorable market conditions to prepare the newly drilled wells for production in the phase of well completion (Smith, 2017). When extracting conventional formations, well completion consists of casing and perforation only, after which oil and gas can flow out readily (DrillingInfo, 2014).

Shale energy extraction differs from conventional practice in that fracking stimulation is necessary to release trapped oil and gas before production begins. As is shown in Panel B of Figure 1, the fracking process lasts about 3-5 days at the end of well completion until oil and gas start to flow from the formations into the wells (Coloradans for Responsible Energy Development

(CRED), 2014). Over an average fracking period of 3 days, up to 2,300 truck trips are needed to fulfill the logistics demand of water supply (NYSDOT, 2011). Moreover, the demand for trucks continues to be high within a few weeks after well completion, as a vast volume of wastewater needs to be collected and transported to disposal sites. On average, the volume of fracking wastewater at the initial months of production is more than twice the amount generated by conventional wells in the same production phase (Xu, Xu, and Khanna, 2017). Later on, wastewater diminishes in volume rapidly and keeps being returned as byproducts of energy at a considerably lower daily flow rate (Veil, 2015), which indicates that the truck demand becomes much less intense throughout the rest of the well's production.

4.2.2 Exogenous Variations in Truck Traffic at the Road Segment Level

This study exploits the temporal shocks to truck traffic induced by fracking activities near a road segment. Unlike Muehlenbachs, Staubli, and Chu (2017), we do not explicitly identify the routes used by fracking wells to transport water from the withdrawal sites and haul wastewater to the disposal sites. Instead, we measure fracking-related truck traffic by the intensity of fracking activities within a buffer zone of each road segment. Figure 2 shows that a buffer zone is defined by the distance from a buffer's border to a road segment. Our analysis depends on three identification assumptions. First, we assume that the timing of a well's drilling, fracking, and production depends largely on local geology and market conditions. In our sampling area that belongs to the Bakken Formation of North Dakota, the fracking technique is mostly combined with horizontal drilling to extract shale resources (Nordeng, 2010). To identify the onset of fracking activities, we use horizontal wells' completion month to infer when fracking occurs, exploiting the fact that fracking takes place within a few days at the end of well completion phase (CRED, 2014;

DrillingInfo, 2014). Thus, we treat horizontal wells completed in the current month as fracking wells and those completed in the previous month as wells fracked one month before. Since the transport demand of horizontal and vertical wells are similar in drilling and production phases, we consider wells of both drill types when counting spudding wells and producing wells.

Second, we assume that wells closer to a road segment are more likely to use it for transport purpose compared to more distant wells. Admittedly, wells at a distance may still affect the traffic nearby when drivers avoid heavy traffic near the well sites or trucks servicing multiple well sites converge to an interstate far away from the well sites (Muehlenbachs, Staubli, and Chu, 2017). However, we expect the probability to decrease with the distance from the road segment to the wells. To demonstrate this point, we show distance-decaying effects of fracking activities on the count of fatal crashes as a robustness check. Alternatively, if wells included in the buffer areas did not use the road in reality, our method would produce a lower-bound estimate of the true impact of fracking per well.

Third, we assume that the associated traffic demand in each phase is a linear function of the well counts of different statuses given the fairly homogeneous geology in the Bakken Shale formation. While the traffic associated with each well may also capture commute vehicles, the frequency of commute traffic is much lower than that of truck traffic servicing the fracking process. Particularly, since over 98% of the oil and gas workforce are in the pre-drilling and drilling phase (Brundage et al., 2011), the traffic impacts near the fracking stage is less likely to be confounded by commute vehicles. In our analysis, we use county-specific time trends to capture the confounder of commute traffic. We further narrow down to large-truck crashes to pinpoint the impact of fracking-related truck traffic. We also examine the incidence of fatal accidents in the rush hours when commute vehicles were concentrated on the road to rule out the potential confounding effects of commute traffic.

4.2.3 Empirical Model

We estimate the effects of fracking-related truck traffic on the expected count of the fatal auto-vehicle crash using the following fixed-effect Poisson model:

(1) $$E(y_{it}|X) = \exp(\beta_1 comp_{it} + \beta_2 comp_{it-1} + \beta_3 spud_{it} + \beta_4 prod_{it} + \alpha_i + \eta_t + \mu_{ct})$$

where each observation is a road segment associated with the number of wells in each operational phase within the defined buffer area and the count of fatal traffic crashes on the road. The index i denotes road segment, and the index t signifies year-month. The dependent variable is the count of fatal traffic accidents by vehicle type. The coefficients of β_1, β_2, β_3, and β_4 are interpreted as the differences in logs of the expected crash counts for a one-unit increase in the number of wells in the corresponding status. Specifically, we count the number of horizontal wells completed in the current month, $comp_{it}$, to capture the logistics activities in well completion at the end of which fracking stimulation occurs. Note that fracking wastewater transport may also take place during the completion month for some wells whose fracking process was finished earlier in the month. Next, the number of horizontal wells completed in the last month, $comp_{it-1}$, captures the traffic related to wastewater collection after fracking for most wells, as large volumes of wastewater return to the surface within a few weeks after well completion. For a small share of wells that were completed at the very end of last month, however, the variable also captures part of their fracking fluid delivery that was not finished until the current month.

We further capture the equipment transport needs arising from the drilling phase by the number of newly drilled wells, $spud_{it}$. The variable helps capture the confounding effects caused by labor commute vehicles since a dominant percentage of oil and gas employment is created at drilling stage (Brundage et al., 2011). We further control the traffic related to regular oil and gas production by the number of wells that have been producing oil and gas actively, $prod_{it}$. Older producing wells require fewer truck trips due to a lower wastewater volume generated (Xu, Xu, and Khanna, 2017). As the average age of producing wells increases with the number of wells accumulated over months, a higher number of producing wells could be associated with a lower traffic impact. However, road segments without any producing wells nearby should be safer overall compared to those with some producing wells, since the former is free of any traffic related to energy production. Thus, we expect the sign of the coefficient of β_4 to be uncertain.

We control for the road segment fixed effects, α_i, to eliminate the permanent differences across road segments, such as road capacity, the functional classifications, the access to disposal sites and water withdrawal sites, etc. We also control for the year by season fixed effects, η_t, to capture the seasonality in weather conditions and market conditions that are common to the whole drilling areas.[8] Finally, we control for the county by year fixed effects, μ_{ct}, to deal with the county-specific time trends. The county by year fixed effects enable us to disentangle the traffic directly related to fracking from confounders in the energy boomtowns, such as the construction of new hotels, the influx of migrants and commuters, the year-specific changes in county-level

---

[8] While the year-month fixed effects can control for monthly variations in market and weather conditions, the fixed-effects Poisson model controlling for them together with the road segment and the year-county fixed effects doesn't converge well. Thus, we use the year-season fixed effects instead, though the results of using year-month fixed effects with either road segment or year-county fixed effects are also consistent with our baseline findings. For all the linear models we run later of the study, we use the year-month fixed effects, which can capture the more precise monthly trends than the seasonal trends.

demography and local regulations and policies, etc. Note that controlling for fixed effects in a non-linear model may lead the observations without any variations within the fixed-effect groups to drop out of the sample. Thus, we test the sensitivity of the baseline model to different samples selected by the fixed-effect groups in the robustness check section.

4.3 DATA AND DESCRIPTIVE STATISTICS

4.3.1 Well-level Data

We use proprietary well-level data from the Drillinginfo Inc., which provides comprehensive information for all the U.S. oil and gas wells, including geographic location, drill type, spud date, completion date, and monthly energy and wastewater production. We focus on the period of 2006-2014 when fracking and horizontal drilling had been used combinedly to develop the oil reserves in the Bakken Formation in North Dakota. We restrict the sample to active oil and gas wells with non-zero energy production reported in the Drillinginfo data.

Due to the lack of a direct measure of the fracking status of each well, we identify wells' drilling and fracking status using the spud date and the completion date respectively. The former indicates the commencement date to drill a new well, while the latter denotes the date when a well finishes the completion phase. On average, wells' spud date is four months ahead of the completion date. We then identify producing wells by wells' production status. Over 95% of wells are recorded to start production from the beginning of the month of completion.[9] We exclude the wells that entered the production phase in the recent two months from producing wells to avoid double

[9] As all the wells' production date starts from the first day of a month, which might be inexact, we do not use the information to infer when fracking happens.

counting since the transport demand of these wells has been captured by wells completed in the current and the previous month.

4.3.2 Traffic Crash Data

We obtain the data regarding fatal traffic crashes from the Fatality Analysis Reporting System (FARS) of the National Highway Traffic Safety Administration (NHTSA). The FARS data provide a nationwide census of the fatal auto-vehicle crashes that involve a motor vehicle traveling on a public trafficway and result in the death of a motorist or a non-motorist within 30 days of the accident. For each traffic accident, the FARS data report the timing and location, the environmental conditions at the time of the crash, the in-transport motor vehicles involved, and the demographic features of drivers, passengers, pedestrians, and pedal cyclists. The FARS data also distinguish the accidents including at least one large truck from those involving passenger cars, light trucks (pickup, utility, van, or other), motorcycles, or buses only. In our analysis, we define the former as large-truck crashes and the latter as non-truck crashes, which account for roughly 35% and 65% of the total crashes, respectively.

We calculate the monetary loss of the crashes using the Value of a Statistical Life (VSL). The VSL measures individuals' marginal willingness to pay to reduce the expected number of fatalities by one, which is estimated to be 9.2 million in 2014 dollars by the U.S. Department of Transportation (DOT) (2016). The benefits of preventing injuries of each degree are then assessed at an estimated fraction of the VSL depending on the severity extent, as indicated by minor, moderate, serious, severe, or critical. Since the FARS data classify the severity of injuries only by non-incapacitating or incapacitating, we treat the former as minor or moderate injuries and the latter as serious, severe, or critical injuries, and then take the average costs of injuries in each category. The costs of non-incapacitating and incapacitating injuries are thus estimated to be 0.025 and 0.349 of the VSL respectively. The total monetary loss of a crash, therefore, equals the aggregated VSL of the fatalities and the injuries of different severity in the accident. We also measure the severity of an accident by the average monetary loss per person, which includes the deaths, the injured, and the noninjured in the accident.

Figure 3 plots the 2001-2014 annual count of fracking wells and fatal traffic crashes by vehicle type in the oil and gas producing counties of North Dakota. Both large-truck crashes and non-truck crashes show an increasing trend over the years, which reflects the massive influx of population and vehicles in the boomtowns since the mid-2000s. By comparison, the count of crashes involving large trucks follows the time trends of fracking wells more closely. In 2013, large-truck crashes outnumbered non-truck accidents for the first time, suggesting that heavy trucks have imposed significant accident externalities on these towns.

We project the location of all the 2006-2014 fatal auto-vehicle crashes and active oil and gas wells onto the North Dakota highway system in Figure 4.[10] The covered highway road segments include interstate, state, and county highways. We focus on the crashes that occurred within 300 feet from the centerline of each road segment, which might be right on the highways or on the highway-entrance or -exit ramps. In total, we identify 340 fatal auto-vehicle crashes involving 565 motor vehicles for analysis. Moreover, we find that most wells are located in the Bakken and Three Forks Shale formations in North Dakota, in which roughly 78% of the wells are horizontal while the remaining 22% are vertical. The longest distance between a well and the nearest highway road segment is approximately twelve miles. Thus, we first choose a maximum

[10] We use the shapefiles provided by the Highway Performance Monitoring System (HPMS) public release at https://www.fhwa.dot.gov/policyinformation/hpms/shapefiles.cfm.

buffer of twelve miles so that each well belongs to the buffer area of at least one road segment. After we apply the search radius, we have 1,392 road segments with at least one well in the buffer zone.

4.4 Main Results

4.4.1 Fatal Traffic Crashes Increased After Fracking

We begin by estimating the effects of fracking operations within the estimated baseline buffer of six miles of a road segment on the expected count of fatal auto-vehicle crashes using equation (1). Table 2 reports the results of a fixed-effect Poisson model, where we control for the road segment fixed effects, the county by year fixed effects, and the year by season fixed effects. The dependent variables are the count of total fatal traffic crashes (Column 1), large-truck crashes (Column 2), and non-truck crashes (Column 3). The number of observations is dropped automatically when the outcome variables have all zero values within the fixed-effect groups. We cluster the standard errors by road segment to account for correlations over time within each road segment.

In Table 2, the coefficients are interpreted as the differences in the logs of the expected traffic crash count for a one-unit increase in the number of wells in each status, all else being equal. The results in Column 1 show that across well operations in each phase it is the post-completion phase that is associated with a significant increase in fatal auto-vehicle crashes. Specifically, given a one-unit increase in the number of post-fracking wells, the total crash count is expected to increase by a factor of 1.068 ($= \exp(0.066)$), holding the other variables constant. The estimate is statistically significant at the 1% level. Meanwhile, every one-unit increase in the number of fracking wells decreased the expected total crash count slightly by a factor of 0.996 ($= \exp(-0.004)$), which is statistically insignificant. We interpret the surge in fatal crashes one month after well completion as a result of the intense truck trips collecting and transporting wastewater after fracking. Although we do not find any significant traffic impacts in the fracking phase, the results are probably caused by the exclusion of local roads from our sample. While highways usually constitute fracking waste disposal routes, local roads serve as water withdrawal routes more often (Muehlenbachs, Staubli, and Chu, 2017). As this study focuses on highways, we cannot capture the effects of water sourcing on local roads during the fracking process.

By contrast, a one-unit increase in the number of spudding wells increased the mean crash count by a factor of 1.027 ($= \exp(0.027)$), which is statistically insignificant. We also find that each additional producing well slightly decreased the crash count by a factor of 0.995 ($= \exp(-0.005)$) at the significance level of 10%. The negative coefficient of producing wells should be taken with caution, however, because of the small magnitude of the estimate and the weak statistical evidence. As the direction of the effects of producing wells is expected to be uncertain, we do observe that the coefficient becomes positive and statistically insignificant in some other model specifications.

We further distinguish the fatal crashes by the type of vehicles involved to examine the role played by large trucks in well operations. If the increased fatal accidents indeed resulted from the trucks conveying fracking wastewater after well completion, we would identify a significant increase in large-truck crashes. However, non-truck crashes may also increase as passenger car drivers try to avoid trucks even though trucks are not directly involved in the crashes. The coefficient estimates in Column 2 show that, given a one-unit increase in the number of post-fracking wells, the count of large-truck crashes is expected to increase by a factor of 1.075 ($= \exp(0.072)$) at the 5% level. The findings support our interpretation that the increased fatal accidents at the post-fracking phase were caused by intense truck trips hauling wastewater. Similar to the findings in total traffic crashes, we do not identify any significant impacts on fatal crashes due to wells' drilling and production activities. Likewise, although the number of fracking wells is expected to increase large-truck crashes by a factor of 1.014 ($= \exp(0.014)$, the estimate is statistically insignificant. Column 3 shows that wells' activities across all the operational phases did not have any significant effects on non-truck crashes. However, by comparison, the estimates of the impacts on non-truck crashes are less precisely measured as shown by the larger standard errors, which prevents us from drawing a firm conclusion about the effects of fracking operations on non-truck crashes.

4.4.2 Robustness Tests for the Total Crash Count

Including Road Segments that Never Had an Accident. As we use the fixed-effect Poisson model in the baseline estimation, the road segments without any fatal crashes are dropped out of the sample due to lack of intertemporal variations. Thus, the baseline model does not incorporate the information from these observations. To understand the extent to which excluding these road segments could bias our estimates, we estimate a Poisson model excluding the road segment fixed effects while keeping the county by year and the season by year fixed effects (Column 1 of Table 3). The results show that an additional post-fracking well increased the expected count of total fatal crashes by a factor of 1.061 ($= \exp(0.059)$) at the 5% level. The estimate is of similar magnitude compared with the baseline estimate. Hence, including the road segments that never experienced any accident does not change our conclusion about the impacts of post-fracking transport on road safety.

Vertical Wells As Counterfactual. Next, we expect that without fracking process well operations during and after the well completion phase would be much less truck intensive and thus affect the traffic outcomes to a lesser extent. We use vertical wells' completion activities as a counterfactual. Since fracking a vertical well is less profitable than fracking a horizontal one, operators have lower economic incentives to apply fracking to vertical wells (Morton, 2013). We add the number of vertical wells completed in the current and the previous month to the baseline model and report the results in Column 2 of Table 3. The estimates of the two variables, 0.061 and -0.597 respectively, are statistically insignificant, implying that operations of vertical wells did not have any significant effects on traffic crashes during or following the completion phase. Meanwhile, the number of horizontal wells completed in the last month still significantly increased the count of fatal accidents. Our findings show that the increased fatal crashes in the first month after well completion were less likely to be caused by any activities other than fracking-related trucking.

Persistent Effects of Fracking Waste Trucking. The baseline model assumes that most wells' fracking activities occurred in the same month when they were completed based on the fact that fracking occurs within a few days at the end of well completion. However, exceptions may occur to a small share of wells fracked at the end of a month and finishing the completion phase at the beginning of the next month. Thus, we test the validity of our assumption by adding one lead term of the number of completed wells to the baseline model. Meanwhile, we explore the persistent effects of transport following the fracking process by further adding the number of horizontal wells completed two months ago. We expect wells' demand for truck trips to decrease rapidly over time since the volume of fracking wastewater reduces greatly within a few weeks after well completion (Veil, 2015). Column 3 of Table 3 shows that, while we identify significant traffic impacts of wells' activities one month after completion, we do not find statistical evidence to conclude any persistent effects on traffic outcomes beyond the first month. Neither do we identify any significant effects of well operations one month before the completion month. The results reveal the temporally intense nature of fracking-related water trucking.

Distance-decaying Effects of Fracking Waste Trucking. Next, we vary the buffer distance mile by mile up to twelve miles away from each road segment and re-run the baseline model using each buffer. We expect that including more distant wells in larger buffers would lower the average effects of wells' activities within the buffer if wells farther away are less likely to use a road segment. We compile the estimates and the 95% confidence intervals for the coefficients of post-fracking wells from twelve regressions in Figure 7. The figure shows that, given a one-unit increase in the number of post-fracking wells within one mile, the count of fatal crashes increased by a factor of 1.185 ($= \exp(0.170)$). As the buffer's border is extended, the marginal effects drop to 1.068 ($= \exp(0.066)$) at six miles and finally to 0.986 ($= \exp(-0.014)$) at twelve miles. Moreover, when buffers are large enough, the magnitudes of the estimates become too small to conclude any significant effects. The results show that the traffic impacts of wells' fracking operations decayed with the distance to a road segment, which implies a distance-decreasing chance to use the road.

4.4.3 Increased Monetary Loss and Intensified Severity Due to Fracking

In this subsection, we estimate the monetary loss of the fracking-related traffic crashes associated with one more post-fracking well. In Table 4, we examine both the total monetary loss of fatalities and injuries (Columns 1-3) and the average loss per person on each road segment (Columns 4-6). Since both variables are right-skewed continuous, we take the inverse hyperbolic since (IHS) transformation of the variables following Muehlenbachs, Staubli, and Chu (2017). We then estimate a log-linear model that controls for the same fixed effects as the baseline model.

Column 1 shows that every additional post-fracking well increased the total monetary loss of fatal crashes by 1.1% at the 5% level. Columns 2-3 further show that, while wells' post-fracking activities led to significantly higher financial damage in large-truck accidents, the effects were statistically insignificant for non-truck crashes. Next, Column 4 shows that every additional post-fracking wells increased the average monetary loss per person by 1% at the 5% level. The results indicate that the intense truck trips serving fracking-related water hauling not only increased the count of traffic accidents but also led to more severe crashes that involved more fatalities and more critical injuries. The findings align with the fact that the larger weight, frame, and height, as well as the rigidity of heavy trucks can elevate the extent of damage inflicted on the crash vehicles (Anderson and Auffhammer, 2014). Columns 5-6 show that wells' post-fracking transport only intensified the severity of large-truck crashes but not non-truck accidents, which is consistent with the findings in the total monetary loss and crash count.

4.4.4 Search for the Baseline Buffer Cutoff

We assume that the traffic on a road segment can be affected by the operating activities of wells nearby but to a varying degree depending on the distance to the well. We seek a baseline buffer for each road segment so that beyond the buffer's border well operations only have minimal effects or no effects on traffic of the road. To search for the ideal distance cutoff, we draw twelve buffers with the borders ranging from one mile to twelve miles surrounding every road segment. We treat the areas between the boundaries of two adjacent buffers as a ring. We then estimate the number of fracking wells, post-fracking wells, spudding wells, and producing wells in each of the twelve rings on the count of fatal crashes in one fixed-effect Poisson model.

The estimation results show that the number of post-fracking wells in closer rings significantly affected the count of large-truck crashes. In Figure 5, we plot the estimates and the 95% confidence intervals for the coefficients of the number of post-fracking wells in each of the twelve rings. For instance, the traffic impacts were significant at the 5% level for wells 1 mile away from the road segments and were statistically insignificant for wells 2 miles away. Overall, the figure shows that the post-fracking activities farther away from a road segment had smaller effects on fatal crashes on the road. Beyond six miles from a road segment, the number of post-fracking wells in each ring turns to be negatively correlated with the crash count with the magnitude of the estimates close to zero. Driven by the pattern emerging from the results, we first select six miles as a baseline distance cutoff to estimate the average effects of well operations within the buffer. In a robustness test, we relax the cutoff restriction to show how the selection of different buffers may affect the estimates.

4.4.5 Summary Statistics

Table 1 reports the summary statistics of the baseline sample. We find that the sample mean of total traffic crashes is 0.002, while the sample variance also approximates to 0.002 ($= 0.048^2$). The distribution of the dependent variable thus obeys the assumption required by the Poisson distribution (mean equals variance). The same property holds for truck and non-truck crashes as well, which justifies the selection of a Poisson model. Also, we find that the count of daytime crashes, nighttime crashes, rush-hour crashes, and non-rush crashes all ranges between 0 and 1 with an average of 0.001. The table also shows that on average each road segment is subject to a total monetary loss of $26,362 due to traffic fatalities and injuries every month, among which $11,028 occurred to truck crashes while $15,334 occurred to non-truck crashes. Moreover, the average monetary loss per person on each road segment is $12,778, while that of truck and non-truck crashes is $4,494 and $8,315, respectively. Within six miles of each road segment, we find that horizontal wells account for more than 98% of both spudding and completed wells, while its share is about 60% among producing wells. The findings indicate that horizontal wells constitute a dominant percentage in new wells. Moreover, less than one well is drilled and completed every month, while about 40 wells are engaged in active oil and gas production on average.

Figure 6 demonstrates the seasonal pattern of fatal traffic crashes and wells' drilling and fracking activities, in which we calculate the monthly number of wells and fatal crashes in our sample over a one-year cycle. The figure shows that non-truck crashes took place less frequently in winter, staying at around ten times every month from December to the next March and rising to about twenty counts per month throughout the rest of the year. By comparison, the incidence of large-truck crashes is distributed more evenly across the months, which may result from the continuous truck trips serving the oil and gas industry throughout the whole year. The figure also shows that the number of spudding wells stayed at a constant level in the first three seasons and slightly decreased in the last season. Meanwhile, the number of fracking wells tended to be smaller in the first two seasons and then rose to a higher level in the second half of the year. Overall, well operations were smooth across the year without a sharp increase in magnitude or concentration in certain months.

4.5 Mechanisms

4.5.1 Traffic Volume vs. Accident Rate

This section explores the potential mechanisms through which the fatal traffic crashes took place. We start with investigating the impacts of fracking on traffic volume and traffic accident rate, as the count of accidents on the road is a function of both traffic amount and per-vehicle risk of getting into an accident (Romen and Shurtz, 2016). Since the total traffic counts data are available only *annually* from 2011 to 2014, we perform the analysis at the annual level using a subsample over the period. The key explanatory variable is the number of wells ever completed during the year, which measures the intensity of wells' fracking activities. The control variables include the number of wells that were drilled and those that had started regular production since the last year.

We first estimate the effects of fracking on traffic volume using a log-linear model controlling for the county by year fixed effects and the road segment fixed effects. We find that, given a one-unit increase in the number of fracking wells, the annual total traffic counts increased by 0.3% at the significance level of 1% (Column 1 of Table 5). The estimated percentage change is equivalent to an increase in the yearly cumulative traffic counts by 3,289, which could involve the vehicles trucking water for both fracking injection and wastewater collection. These vehicles were highly likely to be concentrated within a tight time frame near the well completion date instead of being evenly distributed across the months, which thus may generate a substantial impact on the roads with limited traffic capacity. Also, we find that an additional producing well near a road also increased the annual traffic counts by 0.1% at the significance level of 5%.

We then estimate the same model using three types of accident rate as dependent variables, including the count of total crashes per million vehicle miles traveled (VMT), truck crashes per million VMT, and non-truck crashes per million VMT. The results are reported in Columns 2-4 of Table 5 respectively. We do not find any significant effects of fracking activities on any accident rate, which indicates that fatal crashes grew proportionately with traffic volume without elevating the accident rate. We thus conclude that the increased fatal crashes resulted from a higher traffic amount induced by fracking-related vehicles rather than a higher chance of accident given the traffic intensity.

4.5.2 Timing of the Increased Fatal Crashes

Next, we investigate how fracking operations have affected the timing of fatal crashes. We first distinguish daytime crashes from night crashes by defining the former to be the crashes taking place between 5am-9pm and the latter to be those occurring in the rest of the day. Next, for each road segment observed in a particular month, we count daytime and night crashes. Since both variables turn out to take values of either 0 or 1, we treat them as dummy variables and run a linear probability model to estimate the effects of well operations on the likelihood of each type of crashes. We rescale the count of wells in each status to be in 100 to facilitate reporting the estimated coefficients.

Table 6 shows that each additional post-fracking well significantly increased the likelihood of fatal daytime crashes by 0.042 percents at the significance level of 5% (Column 1). As the sample average incidence of fatal daytime crashes is roughly 0.15%, each post-fracking well increased the likelihood of this type of crashes by 28% ($= 0.042/0.15$). In particular, given a one-unit increase in post-fracking wells, the likelihood of daytime truck crashes increased by 0.029 percents at the significance level of 5%, while the probability of daytime non-truck crashes did not change significantly (Columns 2-3). By contrast, we do not find any significant changes in the probability of night crashes due to an additional post-fracking well (Columns 4-6). The results indicate that fracking-related fatal accidents were mainly concentrated in the daytime rather than the night. The findings align with our previous conclusion that the increased crashes were caused by a higher traffic amount since daytime traffic is usually much higher than the nighttime volume (Romen and Shurtz, 2016).

We then define rush hours to be anytime during 6am-10am or 4pm-8pm of a day inclusively. We use the same model specification to examine changes in the likelihoods of rush-hour crashes and non-rush crashes. Column 7 of Table 6 shows that an additional post-fracking well increased the likelihood of rush-hour crashes by 0.037 percents, which is statistically significant at the 10% level. In particular, the probability of large-truck crashes in rush hours increased by 0.026 percents with an additional post-fracking well nearby at the significance level of 10% (Column 8). The findings indicate that fracking-related transport significantly increased the chance of fatal crashes involving large trucks during the rush hours. An additional post-fracking well also increased the incidence of non-rush crashes by 0.027 percents at the 10% level, which could include both large-truck or non-truck crashes during the non-rush hours. The lack of evidence of crashes concentrating in the rush hours indicates that the traffic impacts of fracking-related transport were unlikely to be affected by commute vehicles that usually emerged from the rush hours of the day.

4.5.3 Driver Characteristics

Finally, we examine the potential changes in drivers' characteristics due to the effects of fracking operations. As oil and gas operations take place near a road segment, an increasing number of drivers working in the energy industry may be concentrated on the road, which could alter the composition of the pool of drivers. In particular, since oil and gas workers are often young men, drivers near the fracking well sites may be more likely to run into traffic crashes (Massie, Campbell, and Williams, 1995). Thus, we estimate the association between well operations and the demographic features, driving behaviors, license qualification, and driving history of the drivers involved in the accidents. Specifically, we merge the vehicle- and the driver-level information with the crashes to obtain a sample of drivers. We then merge the well operations status with the driver-level sample by the road segments where the crashes happened. Thus, each observation is a driver involved in a crash, including 565 drivers in 340 crashes on 228 road segments in total. While we control for the county by year and the month by year fixed effects, we exclude the road segment fixed effects since among all the crash roads 174 road segments only had one crash during the sampling period.

Table 7 reports the results of the baseline model using the driver-level sample. We do not identify any significant effects of well operations on the probability of such demographic features as male driver, work injuries, and drivers younger than 24 (Columns 1-3). Instead, in Column 7 we do find that one more post-fracking well increased the likelihood of drowsy driving by 0.32 percent at the 10% level, which means that drivers were more likely to be sleepy, asleep, or fatigued when the fatal accidents happened. However, we do not find any significant increase in the likelihoods of other risky driving behaviors such as speeding, drinking, omitting safety belt use, and careless driving (talking, eating, or car phones). Neither do we find any significant increase in the chance of holding an invalid license (Columns 9-11) or inferior driving history (Columns 12-16).

4.6 Conclusions

This study identifies a causal relationship between fracking-related trucking and road traffic deaths using evidence from North Dakota. We find that transport activities after fracking a well not only increased the count of fatal crashes but also resulted in more severe accidents. The impacts can be explained by a higher traffic volume but not by a higher accident rate or risky driving behaviors. Our analysis indicates that fracking-induced truck traffic significantly increased the incidence of road fatalities in the oil and gas industry, where the traffic death rate had already been six times the rate in all sectors before the fracking boom (Centers for Disease Control and Prevention, 2011). The findings call for the active involvement of policymakers and oil and gas operators in seeking ways to improve workplace safety of the energy sector as well as to mitigate the externalities of fracking operations to fellow road users.

Our findings shed light on risk management strategies for both operators and local governments to manage the threats of fracking-induced trucking to road safety. We show that the peak traffic effects take place right after the well completion phase during which a substantial number of trucks are needed to transport fracking liquid waste. To reduce the road use impacts, oil and gas operators may redistribute the traffic loads over time to avoid intense water hauling during the peak hours. Operators can also improve water supply system to reduce the trucking demand by constructing water wells serving multiple well pads via a piping system. Moreover, they may want to develop the onsite wastewater treatment and disposal facilities as opposed to trucking wastewater over a long distance. Since we identify that the fracking-related crashes mostly happened in the daytime, local governments may take approaches such as reserved truck lanes at certain hours, an analogy to high-occupancy vehicle (HOV) lanes. An active traffic alert and warning system with live well operations updates could also help road users to monitor traffic and avoid exposure to hazards. Finally, our study contributes to the discussion on the well-level Pigouvian tax that may be charged to internalize the costs of fracking-induced water transport (Black, McCoy, and Weber, 2018).

This study draws the public attention to the hidden costs of truck transport borne by the fracking boomtowns besides the economic gains on income and employment (Allcott and Keniston, 2017; Maniloff and Mastromanaco, 2014; Feyrer, Mansur, and Sacerdote, 2017), which contributes to a broader picture of the overall costs and benefits of the U.S. oil and gas boom. In addition to the direct fatality and injury loss quantified in this study, indirect costs occur as well regarding local emergency medical services and the maintenance and replacement of traffic infrastructure, which are beyond the scope of this study but demand future research. Overall, our findings are informative for policymakers to weigh the costs and benefits when deciding whether to allow the use of fracking and planning future shale development.

4.7 Tables and Figures

Table 4.1 Summary Statistics

	Mean	Standard Deviation	Min	Max
Fatal Auto-vehicle Crashes				
#Total crash	0.002	0.048	0	2
#Truck crash	0.001	0.031	0	2
#Non-truck crash	0.001	0.036	0	1
#Daytime crash	0.001	0.038	0	1
#Night crash	0.001	0.028	0	1
#Rush-hour crash	0.001	0.033	0	1
#Non-rush hour crash	0.001	0.034	0	1
#Total crash loss ($)	26,362	605,252	0	37,030,000
#Truck crash loss ($)	11,028	394,098	0	37,030,000
#Non-truck crash loss ($)	15,334	457,263	0	33,736,400
#Total crash severity ($)	12,778	291,639	0	9,200,000
#Truck crash severity ($)	4,494	155,751	0	9,200,000
#Non-truck crash severity ($)	8,315	247,008	0	9,200,000
#Wells in the baseline buffer of six miles				
#Spudding	0.574	1.615	0	36
#Spudding, horizontal	0.565	1.612	0	36
#Spudding, vertical	0.009	0.114	0	14
#Completed	0.577	1.679	0	36
#Completed, horizontal	0.567	1.672	0	36
#Completed, vertical	0.01	0.143	0	27
#Producing	39.686	52.309	0	697
#Producing, horizontal	21.609	42.783	0	697
#Producing, vertical	18.078	23.375	0	174
				150,336

Notes: This table reports the summary statistics of a balanced panel of road segments by year-month from 2006 to 2014. The fatal traffic crashes in our sample occurred within 300 feet from the centerline of a road segment. The baseline buffer of six miles is estimated to be the approximate distance cutoff of an average road segment beyond which well operations no longer affect the traffic status on the road.

Table 4.2 Effects of Fracking-related Transport on Crash Count

	(1)	(2)	(3)
	#Total crash	#Large-truck crash	#Non-truck crash
#Fracking	-0.004	0.014	-0.030
	(0.025)	(0.034)	(0.040)
#Post-fracking	0.066***	0.072**	0.058
	(0.024)	(0.031)	(0.040)
#Spudding	0.027	0.026	0.036
	(0.024)	(0.038)	(0.041)
#Producing	-0.005*	-0.004	-0.004
	(0.002)	(0.003)	(0.004)
Observations	21,720	7,368	14,268
Year-season FE	YES	YES	YES
Year-county FE	YES	YES	YES
Road segment FE	YES	YES	YES

Notes: This table reports the estimation results of a fixed-effect Poisson model using the 2006-2014 road segment by year-month panel data. The estimates are based on the baseline buffer of six miles. The dependent variables are the count of each type of crashes. Fracking wells are defined as horizontal wells completed in the current month. Post-fracking wells are defined as horizontal wells completed in the last month. Spudding and producing wells refer to newly drilled wells and actively producing wells, respectively, which include both horizontal and vertical wells. Observations are dropped in the cases where the county of a year did not have any crashes on any road, or the road segments never had any crash during the sampling period. Standard errors clustered at the road segment level are reported in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 4.3 Robustness Checks

	(1)	(2)	(3)
	Poisson model without road segment FE	Vertical wells as counterfactual	Persistent effects of fracking
$\#Comp_{i,t}$ (H)	-0.011	-0.004	0.008
	(0.023)	(0.026)	(0.025)
$\#Comp_{i,t}$ (V)		0.061	
		(0.163)	
$\#Comp_{i,t-1}$ (H)	0.059**	0.066***	0.067***
	(0.023)	(0.024)	(0.025)
$\#Comp_{i,t-1}$ (V)		-0.597	
		(0.846)	
$\#Spud_{it}$	0.047*	0.027	0.026
	(0.024)	(0.024)	(0.025)
$\#Prod_{i,t}$	0.002	-0.005*	-0.003
	(0.001)	(0.002)	(0.003)
$\#Comp_{i,t+1}$(H)			0.001
			(0.026)
$\#Comp_{i,t-2}$ (H)			-0.026
			(0.024)
Observations	118,707	21,720	21,426
Year-season FE	YES	YES	YES
Year-county FE	YES	YES	YES
Road segment FE	NO	YES	YES

Notes: This table reports the estimation results of a fixed-effect Poisson model using the 2006-2014 road segment by year-month panel data. The estimates are based on the baseline buffer of six miles. The dependent variables are the count of total fatal crashes. Column 1 reports the estimation results of a Poisson model excluding road segment fixed effects, where observations are dropped in the cases where the season of a year or the county of a year did not have any crashes on any road. Column 2-3 report the estimation results of the fixed-effect Poisson models, where observations are further dropped in the cases where the road segments never had a crash during the sampling period. Standard errors clustered at the road segment level are reported in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 4.4 Effects of Fracking-related Transport on Monetary Loss

	(1)	(2)	(3)	(4)	(5)	(6)
	Total Monetary Loss			Monetary Loss per Person		
	Total crash	Large-truck crash	Non-truck crash	Total crash	Large-truck crash	Non-truck crash
#Fracking	-0.002	-0.000	-0.001	-0.002	-0.000	-0.001
	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
#Post-fracking	0.011**	0.007**	0.004	0.010**	0.007**	0.003
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
#Spudding	0.004	0.003	0.002	0.004	0.002	0.002
	(0.005)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)
#Producing	0.00002	0.00007	-0.00007	0.00003	0.00007	-0.00006
	(0.00023)	(0.00019)	(0.00015)	(0.00022)	(0.00018)	(0.00015)
Observations	150,336	150,336	150,336	150,336	150,336	150,336
Year-month FE	YES	YES	YES	YES	YES	YES
Year-county FE	YES	YES	YES	YES	YES	YES
Road segment FE	YES	YES	YES	YES	YES	YES

Notes: This table reports the estimation results of a log-linear model using the 2006-2014 road segment by year-month panel data. The estimates are based on the baseline buffer of six miles. The dependent variables in Columns 1-3 are the Inverse Hyperbolic Sine (IHS) transformation of the total monetary loss of fatalities and injuries involved in each type of crashes. The dependent variables in Columns 4-6 are the IHS transformation of the monetary loss per person of the crashes. Each type of wells follow the same definition in Table 2. Standard errors clustered at the road segment level are reported in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 4.5 Effects of Fracking-related Transport on Traffic Volume and Accident Rate

	(1) Annual traffic counts	(2) Total crash rate	(3) Truck crash rate	(4) Non-truck crash rate
#Fracking	0.003***	0.003	0.003	-0.000
	(0.001)	(0.004)	(0.003)	(0.003)
#Spudding	-0.001	-0.002	-0.001	0.001
	(0.001)	(0.004)	(0.003)	(0.003)
#Producing	0.001**	-0.002	-0.002	0.000
	(0.000)	(0.002)	(0.001)	(0.001)
Observations	5,543	5,541	5,541	5,541
Year-county FE	YES	YES	YES	YES
Road segment FE	YES	YES	YES	YES

Notes: This table reports the estimation results of a log-linear model using the 2011-2014 road segment by year panel data. The estimates are based on the baseline buffer of six miles. The dependent variables in Columns 1-4 are the IHS transformation of the annual traffic counts, total crash rate, truck crash rate, and non-truck crash rate, respectively. The crash rates are measured by the count of fatal crashes per million vehicle miles traveled (VMT). Fracking wells are defined as horizontal wells completed in the current year. Spudding and producing wells refer to newly drilled wells in the current year and actively producing wells since the last year, respectively, which include both horizontal and vertical wells. Standard errors clustered at the road segment level are reported in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 4.6 Timing of the Increased Fatal Crashes

	(1) Daytime crash	(2) Daytime truck crash	(3) Daytime non-truck crash	(4) Night crash	(5) Night truck crash	(6) Night non-truck crash
#Fracking	-0.0240	-0.0107	-0.0132	0.0161	0.0113	0.0048
	(0.0177)	(0.0149)	(0.0097)	(0.0127)	(0.0103)	(0.0079)
#Post-fracking	0.0422**	0.0291**	0.0131	0.0215	0.0126	0.0089
	(0.0208)	(0.0145)	(0.0135)	(0.0131)	(0.0102)	(0.0073)
#Spudding	0.0308	0.0195	0.0113	-0.0033	-0.0054	0.0021
	(0.0229)	(0.0194)	(0.0149)	(0.0139)	(0.0109)	(0.0101)
#Producing	0.0006	0.0007	-0.0000	-0.0006	-0.0003	-0.0003
	(0.0010)	(0.0008)	(0.0007)	(0.0008)	(0.0005)	(0.0006)
Observations	150,336	150,336	150,336	150,336	150,336	150,336
Year-month FE	YES	YES	YES	YES	YES	YES
Year-county FE	YES	YES	YES	YES	YES	YES
Road segment FE	YES	YES	YES	YES	YES	YES

Notes: This table reports the estimation results of a linear probability model using the 2006-2014 road segment by year-month panel data. The estimates are based on the baseline buffer of six miles. The dependent variables are the count of fatal crashes of each type labeled in the column title. Each type of wells follows the same definition in Table 2. Standard errors clustered at the road segment level are reported in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 4.6 Timing of the Increased Fatal Crashes (Continued)

	(7) Rush-hour crash	(8) Rush-hour truck crash	(9) Rush-hour non-truck crash	(10) Non-rush crash	(11) Non-rush truck crash	(12) Non-rush non-truck crash
#Fracking	-0.0303**	-0.0188	-0.0115	0.0224	0.0194	0.0030
	(0.0150)	(0.0121)	(0.0091)	(0.0147)	(0.0119)	(0.0085)
#Post-fracking	0.0365*	0.0263*	0.0103	0.0272*	0.0154	0.0118
	(0.0194)	(0.0136)	(0.0125)	(0.0150)	(0.0110)	(0.0091)
#Spudding	0.0297	0.0148	0.0149	-0.0021	-0.0006	-0.0015
	(0.0202)	(0.0156)	(0.0145)	(0.0180)	(0.0156)	(0.0103)
#Producing	0.0010	0.0008	0.0002	-0.0009	-0.0004	-0.0005
	(0.0009)	(0.0008)	(0.0006)	(0.0009)	(0.0006)	(0.0007)
Observations	150,336	150,336	150,336	150,336	150,336	150,336
Year-month FE	YES	YES	YES	YES	YES	YES
Year-county FE	YES	YES	YES	YES	YES	YES
Road segment FE	YES	YES	YES	YES	YES	YES

Notes: This table reports the estimation results of a linear probability model using the 2006-2014 road segment by year-month panel data. The estimates are based on the baseline buffer of six miles. The dependent variables are the count of fatal crashes of each type labeled in the column title. Each type of wells follows the same definition in Table 2. Standard errors clustered at the road segment level are reported in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 4.7 Driver Characteristics

	(1) Male	(2) Work injury	(3) Younger than 24	(4) Speeding	(5) Drinking	(6) No belt	(7) Drowsy driving	(8) Careless driving
#Fracking	-0.0066	-0.0057	-0.0034	0.0010	0.0093	0.0006	-0.0008	-0.0012
	(0.0071)	(0.0069)	(0.0084)	(0.0007)	(0.0110)	(0.0014)	(0.0006)	(0.0010)
#Post-fracking	0.0082	0.0007	-0.0167	-0.0009	-0.0044	-0.0016	0.0032*	0.0010
	(0.0093)	(0.0058)	(0.0103)	(0.0006)	(0.0118)	(0.0039)	(0.0018)	(0.0014)
#Spudding	0.0021	0.0060	0.0056	0.0003	-0.0040	-0.0003	-0.0004	-0.0010
	(0.0057)	(0.0053)	(0.0088)	(0.0004)	(0.0107)	(0.0014)	(0.0003)	(0.0011)
#Producing	0.0002	0.0000	-0.0002	-0.0000	0.0005	-0.0000	-0.0001*	0.0002
	(0.0004)	(0.0004)	(0.0006)	(0.0000)	(0.0007)	(0.0002)	(0.0001)	(0.0001)
Observations	565	565	565	565	565	565	565	565
R-squared	0.388	0.294	0.355	0.708	0.389	0.704	0.475	0.757
Year-month FE	YES	YES	YES	YES	YES	YES	YES	YES
Year-county FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes: This table reports the estimation results of a linear probability model using the driver-level sample. The estimates are based on the baseline buffer of six miles. The dependent variables are dummies indicating each type of drivers labeled in the column title. Each type of wells follows the same definition in Table 2. Standard errors clustered at the road segment level are reported in parentheses *** p<0.01, ** p<0.05, * p<0.1

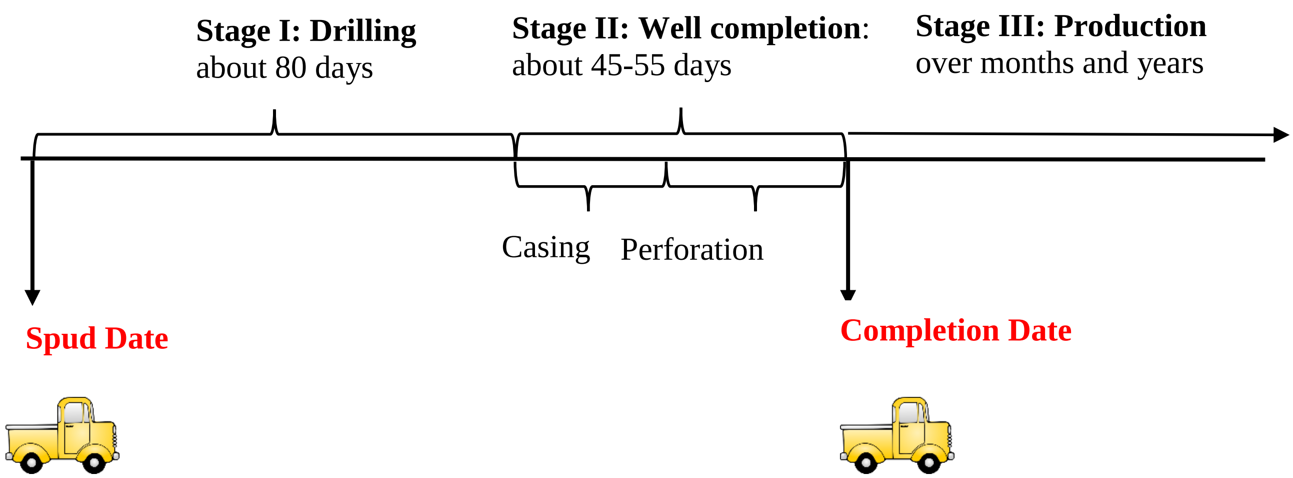
Table 4.7 Driver Characteristics (Continued)

	(9) Valid license	(10) License compliance with class of vehicle	(11) License compliance with restrictions	(12) Previous recorded crashes	(13) Previous suspensions &revocations	(14) Previous DWI convictions	(15) Previous speeding convictions	(16) Previous other convictions
#Fracking	0.0022	0.0016	0.0030	0.0014	-0.0039	0.0009	-0.0034	0.0045
	(0.0095)	(0.0094)	(0.0022)	(0.0091)	(0.0071)	(0.0054)	(0.0091)	(0.0084)
#Post-fracking	0.0101	0.0079	0.0007	-0.0044	-0.0022	-0.0038	0.0008	0.0040
	(0.0079)	(0.0084)	(0.0020)	(0.0079)	(0.0110)	(0.0066)	(0.0117)	(0.0097)
#Spudding	-0.0011	-0.0046	-0.0003	0.0001	0.0025	0.0026	-0.0000	-0.0126*
	(0.0105)	(0.0098)	(0.0010)	(0.0055)	(0.0074)	(0.0065)	(0.0100)	(0.0069)
#Producing	-0.0010*	-0.0009	0.0000	0.0002	0.0001	0.0000	-0.0004	0.0001
	(0.0005)	(0.0006)	(0.0001)	(0.0005)	(0.0005)	(0.0003)	(0.0006)	(0.0005)
Observations	565	565	565	565	565	565	565	565
R-squared	0.388	0.373	0.287	0.317	0.362	0.486	0.319	0.321
Year-month FE	YES	YES	YES	YES	YES	YES	YES	YES
Year-county FE	YES	YES	YES	YES	YES	YES	YES	YES

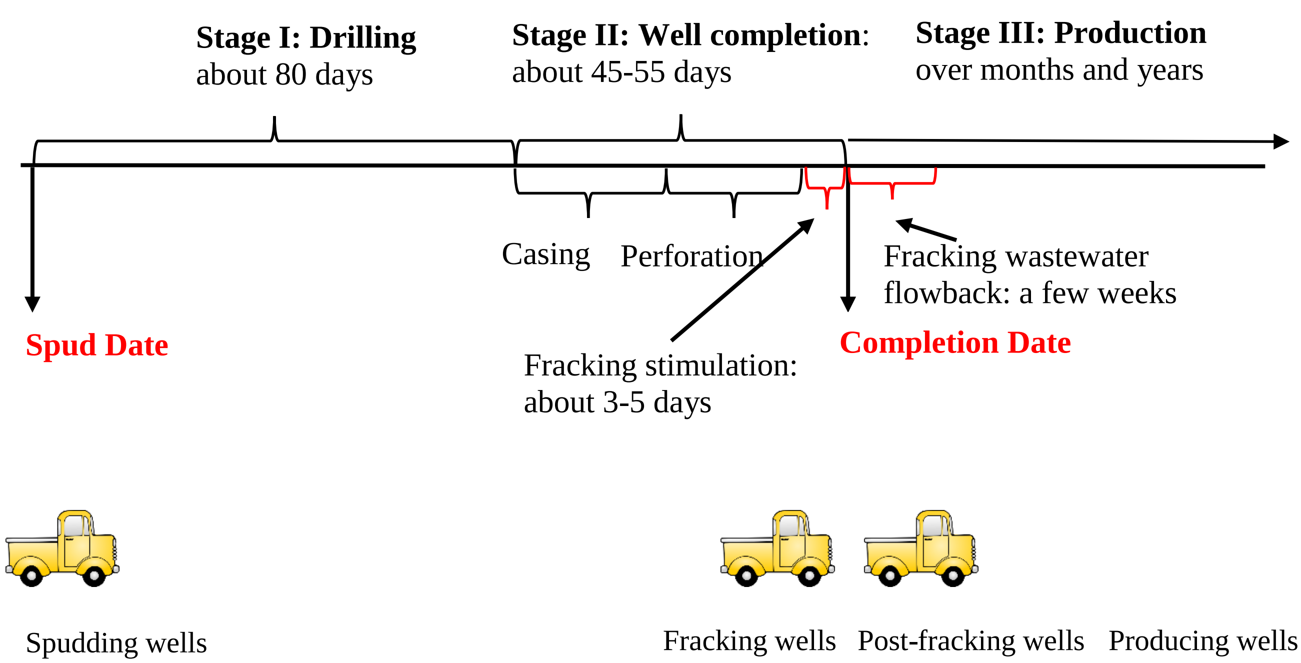
Notes: This table reports the estimation results of a linear probability model using the driver-level sample. The estimates are based on the baseline buffer of six miles. The dependent variables are dummies indicating each type of drivers labeled in the column title. Each type of wells follows the same definition in Table 2. Standard errors clustered at the road segment level are reported in parentheses *** p<0.01, ** p<0.05, * p<0.1

Figure 4.1

A. Operational Phases of Conventional Oil and Gas Wells

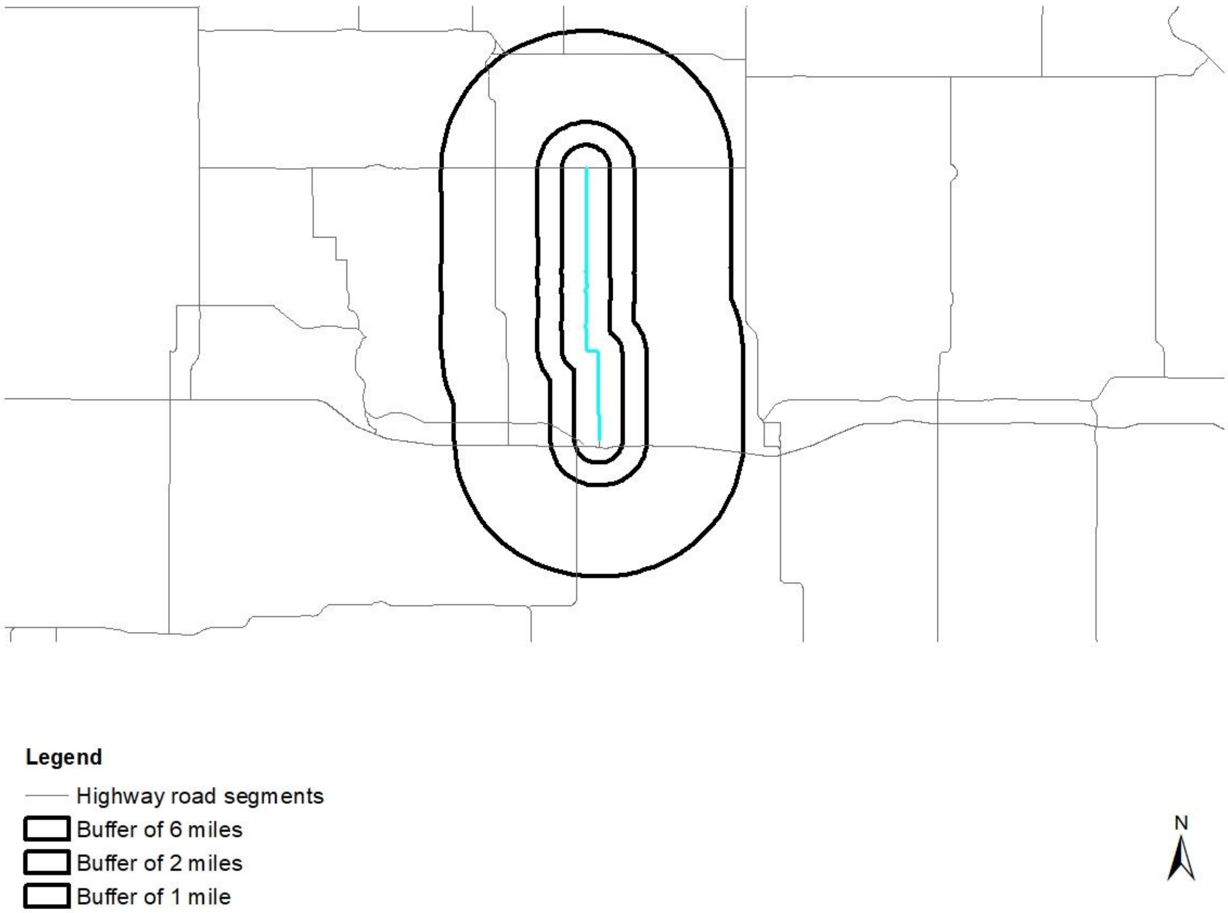


B. Operational Phases of Shale Oil and Gas Wells



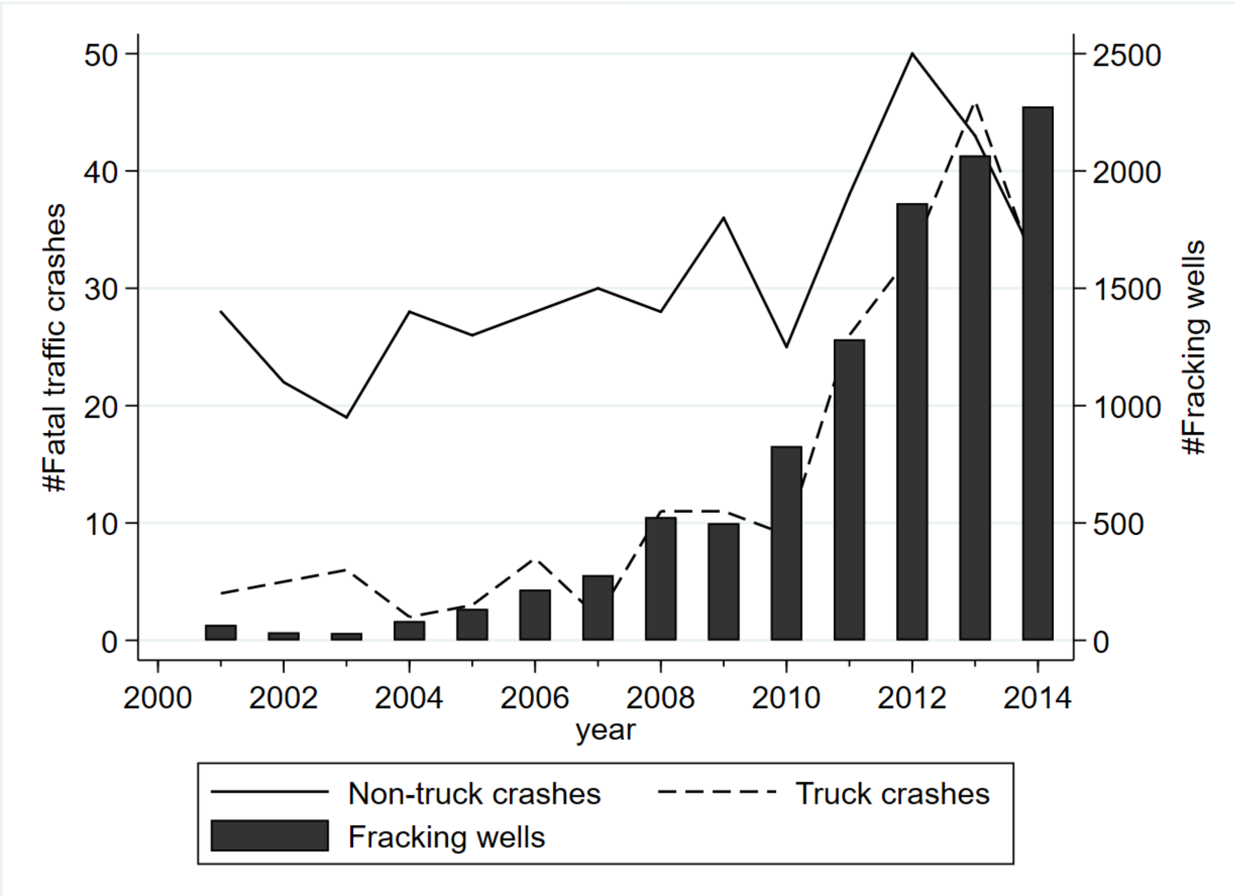
Notes: Created by authors using information by DrillingInfo (2014) and Muehlenbachs, Staubli, and Chu (2017)

Figure 4.2 Illustration of Buffers of A Road Segment by Distance



Notes: The figure illustrates buffers by distance for the selected road segment.

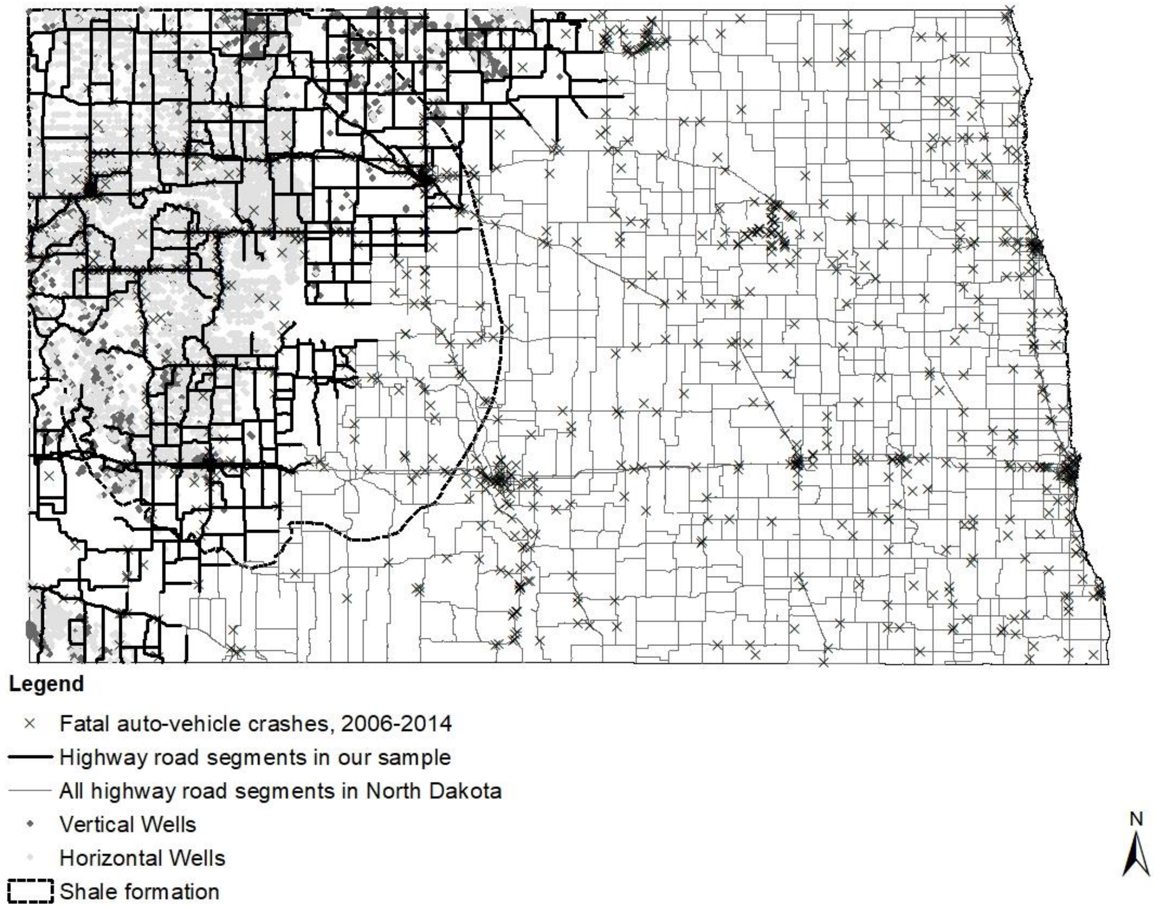
Figure 4.3 Time Trends of the Number of Fracking Wells and Fatal Auto-vehicle Crashes, 2001-2014



Source: Created by authors using data from the National Highway Traffic Safety Administration and the Drillinginfo.

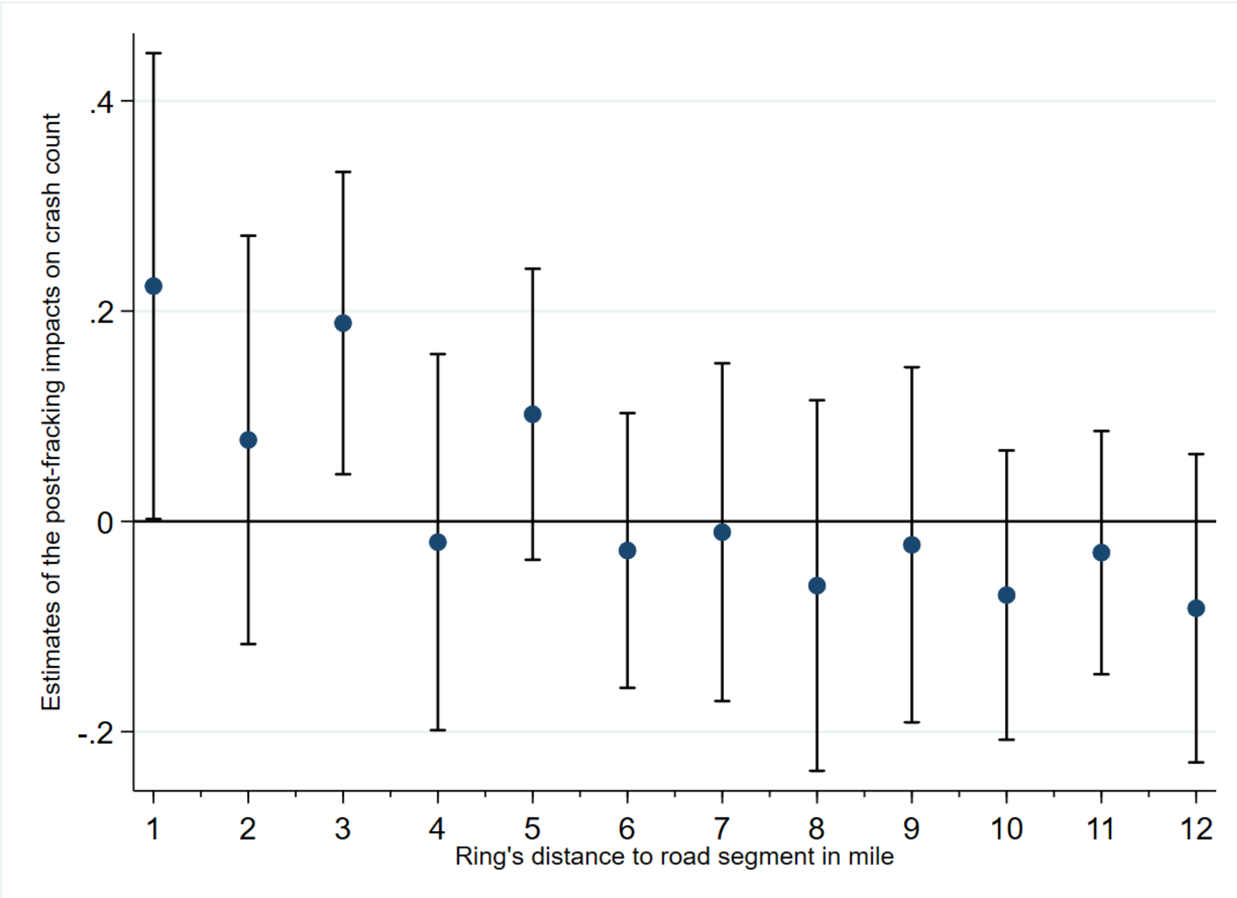
Notes: The figure plots the annual number of fracking wells and the annual count of fatal traffic crashes by vehicle type in the oil and gas producing counties of North Dakota. Fracking wells are defined as the horizontal wells completed in each year.

Figure 4.4 Oil and Gas Wells and Fatal Auto-vehicle Crashes on Highways in North Dakota



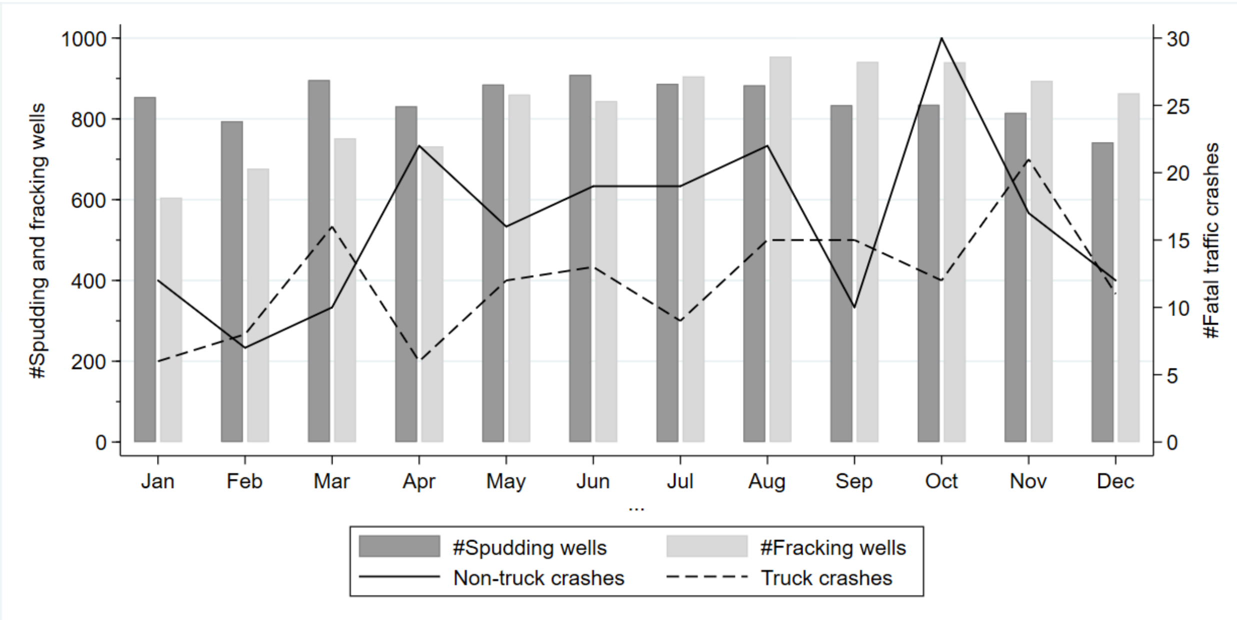
Source: Created by authors using the highway shapefiles from the Highway Performance Monitoring System, the fatal auto-vehicle crashes data from the Fatality Analysis Reporting System, and the geographic location of wells from the DrillingInfo data.

Figure 4.5 Search for the Baseline Buffer Cutoff



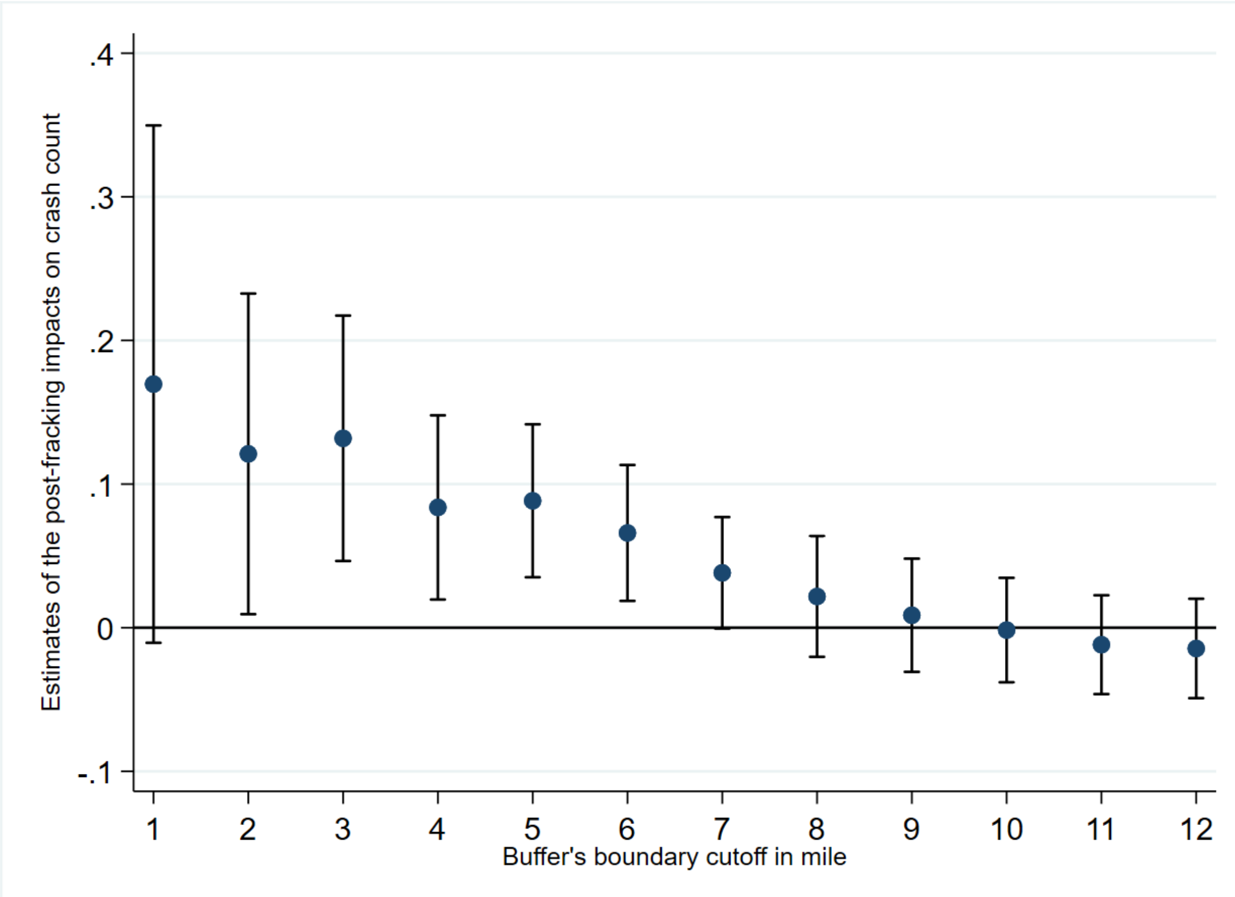
Notes: We divide the buffer of each road segment by twelve rings and estimate the effects of well operations in each ring on the count of large-truck crashes in one fixed-effect Poisson regression. The figure plots the coefficient estimates and the 95% confidence intervals of the number of post-fracking wells in each ring.

Figure 4.6 Seasonal Pattern in Fatal Traffic Crashes and Wells Operations



Notes: The figure plots the count of fatal traffic crashes and the number of spudding and fracking wells in our sample by month over a one-year cycle. Spudding wells are defined as wells newly drilled in each month. Fracking wells are defined as horizontal wells completed in each month.

Figure 4.7 The Distance-decaying Effects of Wells' Post-fracking Activities



Notes: The figure compiles the estimates from twelve fixed-effect Poisson regressions, which use twelve buffers with the borders ranging from one to twelve miles, respectively. The figure plots the coefficient estimates and the 95% confidence intervals for the number of post-fracking wells within the buffer.

4.8 References

Abouk, R. and Adams, S., 2013. Texting bans and fatal accidents on roadways: do they work? Or do drivers just react to announcements of bans?. *American Economic Journal: Applied Economics*, *5*(2), pp.179-99.

Adams, S., Blackburn, M.L. and Cotti, C.D., 2012. Minimum wages and alcohol-related traffic fatalities among teens. *Review of Economics and Statistics*, *94*(3), pp.828-840.

Allcott, H. and Keniston, D., 2017. Dutch disease or agglomeration? The local economic effects of natural resource booms in modern America. *The Review of Economic Studies*, *85*(2), pp.695-731.

Anderson, M.L. and Auffhammer, M., 2013. Pounds that kill: The external costs of vehicle weight. *Review of Economic Studies*, *81*(2), pp.535-571.

Black, Katie J., Shawn J. McCoy, and Jeremy G. Weber. 2018. When Externalities Are Taxed: The Effects and Incidence of Pennsylvania's Impact Fee on Shale Gas Wells. *Journal of the Association of Environmental and Resource Economists*, *5*(1), pp.107-153.

Brundage, Tracy L, Timothy W Kelsey, Janice Lobdell, Larry L Michael, Jeffrey Jacquet, James R Ladlee, Jeffrey F Lorson, and Thomas B Murphy. 2011. Pennsylvania Statewide Marcellus ShaleWorkforce Needs Assessment. Williamsport, PA: Marcellus Shale Education and Training Center.

Centers for Disease Control and Prevention (CDC, 2011. Occupational highway transportation deaths---United States, 2003--2008. *MMWR. Morbidity and mortality weekly report*, *60*(16), p.497.

Cohen, A. and Einav, L., 2003. The effects of mandatory seat belt laws on driving behavior and traffic fatalities. *Review of Economics and Statistics*, *85*(4), pp.828-843.

Cohen, A. and Dehejia, R., 2004. The effect of automobile insurance and accident liability laws on traffic fatalities. *The Journal of Law and Economics*, *47*(2), pp.357-393.

Coloradans for Responsible Energy Development. 2014. How long does fracking last and are wells fracked more than once? https://www.cred.org/fracking-frequency/

DeAngelo, G. and Hansen, B., 2014. Life and death in the fast lane: Police enforcement and traffic fatalities. *American Economic Journal: Economic Policy*, *6*(2), pp.231-57.

Drillinginfo. 2014. Well completion 101 part 3: Well stimulation. Petroleum Engineering Archives. https://info.drillinginfo.com/well-completion-well-stimulation/

Feyrer, J., Mansur, E.T. and Sacerdote, B., 2017. Geographic dispersion of economic shocks: Evidence from the fracking revolution. *American Economic Review*, *107*(4), pp.1313-34.

Finkel, M.L. ed., 2015. *The Human and Environmental Impact of Fracking: How Fracturing Shale for Gas Affects Us and Our World: How Fracturing Shale for Gas Affects Us and Our World*. ABC-CLIO.

Jacobsen, M.R., 2013. Fuel economy and safety: The influences of vehicle class and driver behavior. *American Economic Journal: Applied Economics*, *5*(3), pp.1-26.

Levitt, S.D. and Porter, J., 2001. How dangerous are drinking drivers?. *Journal of political Economy*, *109*(6), pp.1198-1237.

Li, S., 2012. Traffic safety and vehicle choice: quantifying the effects of the 'arms race'on American roads. *Journal of Applied Econometrics*, *27*(1), pp.34-62.

Maniloff, P. and Mastromonaco, R., 2014. The local economic impacts of hydraulic fracturing and determinants of dutch disease. *Colorado School of Mines, Division of Economics and Business Working Paper*, *8*.

Massie, D.L., Campbell, K.L. and Williams, A.F., 1995. Traffic accident involvement rates by driver age and gender. *Accident Analysis & Prevention*, *27*(1), pp.73-87.

Morton, M.Q., 2013. Unlocking the Earth: A Short History of Hydraulic Fracturing.

Moss, K., 2008. Potential development of the natural gas resources in the Marcellus Shale: New York, Pennsylvania, West Virginia and Ohio. *Denver, CO: National Park Service, US Department of the Interior, Geologic Resources Division*.

Muehlenbachs, L., Staubli, S. and Chu, Z., 2017. *The Accident Externality from Trucking* (No. w23791). National Bureau of Economic Research.

New York State Department of Transportation. 2011. Transportation impacts of potential Marcellus shale gas development. Draft Discussion Paper. http://www.slideshare.net/MarcellusDN/draft-discussion-paper-transportation-impacts-of-potential-marcellus-shale-gas-development

Nordeng, S.H., 2010. A brief history of oil production from the Bakken Formation in the Williston Basin. *Geo News: North Dakota, Department of Mineral Resources Newsletter*.

Podulka, Sandra G., and William J. Podulka. 2010. Comments on the science advisory board's 5/19/2010 draft committee report on the EPA's research scoping document related to hydraulic fracturing ("Report"). https://yosemite.epa.gov/sab/SABPRODUCT.nsf/F7A9F146CE2CE73E8525773E00653885/$File/EPA+SAB+Hydrof+Report+Comments+Sandra+and+William+Podulka.pdf

Romem, I. and Shurtz, I., 2016. The accident externality of driving: Evidence from observance of the Jewish Sabbath in Israel. *Journal of Urban Economics*, *96*, pp.36-54.

Smith, J. 2017. How and why petroleum leases are held by production: Analysis of a compound option. Paper presented at the seminar of the Department of Agricultural and Consumer Economics in the University of Illinois at Urbana-Champaign, Champaign, IL.

U.S. Department of Transportation. 2016. Revised departmental guidance on valuation of a statistical life in economic analysis. Washington, DC.

U.S. Environmental Protection Agency. 2015. Review of state and industry spill data: Characterization of hydraulic fracturing-related spills. Office of Research and Development, Washington, DC.

U.S. Environmental Protection Agency. 2016. Hydraulic fracturing for oil and gas: Impacts from the hydraulic fracturing water cycle on drinking water resources in the United States. Washington, DC: Office of Research and Development.

Veil, J., 2015. US produced water volumes and management practices in 2012. *Oklahoma City, OK: Ground Water Protection Council*.

Vickrey, W., 1968. Automobile accidents, tort law, externalities, and insurance: an economist's critique. *Law and Contemporary Problems*, *33*(3), pp.464-487.

Wang, Z. and Krupnick, A., 2013. A retrospective review of shale gas development in the United States: What led to the boom?.

Xu, M., Y. Xu, and M. Khanna. 2017. Do unconventional oil and gas wells generate more wastewater? A lifespan perspective. Working paper, University of Illinois at Urbana-Champaign.

CHAPTER 5

CONCLUSIONS

This dissertation applies econometric methods and spatial techniques to answer three questions regarding the socioeconomic consequences of the U.S. fracking boom: (1) how does oil and gas pipeline transport hazards affect mortgage credit access in the real estate market? (2) how does unconventional drilling and production techniques affect wastewater generation regarding the lifespan production pattern and the production efficiency? and (3) how does fracking-related trucking activities affect the accident externality?

The first essay demonstrates a permanently lower origination rate by 1.9% in the pipeline-present areas compared to the pipeline-free areas, which is shown to result from lenders' concerns about collateral value and borrowers' repayment ability. The permanent difference in credit access was further enlarged by 1.8% when pipeline incidents happened, which indicates lenders' aversion to the potential environmental liabilities. The essay identifies that it is the lower-income borrowers that are subject to the credit loss due to pipeline hazards while mortgage lenders' management strategies always responded to the tightening of the securitization market.

The second essay shows that horizontal wells drilled in shale regions after 2010 produced an increasingly larger amount of wastewater at the initial stage of production relative to vertical wells while the difference between them decreased with production age regardless of cohort years. However, unconventional wells tended to have a lower lifespan cumulative wastewater and a lower wastewater-to-energy ratio.

The third essay shows that wells' trucking activities within one month after the fracking phase increased the count of fatal crashes involving large trucks by 7.5% and induced more fatalities and more critical injuries. Moreover, the increased crashes emerged during both rush and non-rush hours in the daytime but not in the night. Evidence suggests that the increased accidents were due to a higher traffic amount rather than a higher accident rate or risky driving behaviors.

Several conclusions obtained from this dissertation are worthy of special attention. First, as this dissertation identifies the population who suffer from the credit loss when pipeline hazards are present, this essay promotes public discussion on the environmental injustice existing in the fracking boom. Future attempts on this topic can be made to further investigate the distribution of fracking wells and waste injection sites across neighborhoods with residents of different race, color, and income. Next, in contrast to the anecdotal evidence and prevailing perceptions, our rigorous lifespan analysis shows that the fracking technology turns out to be more efficient than conventional oil and gas production methods in the long run. Moreover, the constant technological improvements over time will eventually enable the U.S. to get rid of the dependence on oil import and dirty fuel resources. However, the real concerns should lie in the associated risk management strategies, which usually lag because of the unique production pattern and the unprecedented scale of the shale oil and gas boom. The elevated accident externality induced by fracking, for instance, may exist as a long-term road safety hazard considering the periodic nature of fracking activities, the large scale of drilling, and the huge resource potential throughout the nation. The findings of the dissertation thus inform policymakers of effective risk management practices to alleviate the socioeconomic costs that may occur in the future shale energy development.