ABSTRACT

Title of dissertation: ESSAYS ON INDIVIDUAL RESPONSES TO LABOR MARKET CONDITIONS AND POLICIES

Riley Edward Wilson
Doctor of Philosophy, 2018

Dissertation directed by: Professor Melissa S. Kearney
Department of Economics

This dissertation examines how individuals respond to changing conditions and policies in the labor market, with a particular interest in understanding economically motivated migration and labor force attachment.

I first turn to the question of labor mobility. There is long-standing academic and policy interest in the issue of economically motivated geographic mobility. I examine the recent context of localized "fracking" booms in the United States to explore the migration response to positive labor demand shocks. Using data from 1999 to 2013, I show that local fracking led to large increases in potential earnings and employment rates, as well as a sizable migration response. But, this average migration effect masks substantial underlying heterogeneity in migration behavior across both demographics and regions. Migrants to fracking areas were more likely to be male, unmarried, young, and less educated than movers more generally. Furthermore, both in- and out-migration rates increased with fracking and both flows were driven by the same demographic groups, suggesting fracking resulted in

short-term migration and increased churn. An instrumental variables analysis using fracking conditions to instrument for earnings suggests that a ten percent increase in average earnings increased in-migration rates by 3.8 percent in North Dakota fracking counties, as compared to only 2.4 percent in the West, 1.6 percent in the South, and 0.5 percent in the Northeast. The difference across regions is statistically significant; robust to housing market controls, geographic spillovers, and other various specifications; and is only partially explained by differences in commuting behavior, initial population characteristics, or a non-linear relationship between earnings and migration. There is some evidence that heterogeneous information flows might be driving the heterogeneous migration response. This implies that lack of information might be dampening rates of migration to economically favorable labor markets.

I next examine how labor market information affects these types of economically motivated migration decisions. Migration is a human capital investment that allows individuals to encounter more favorable labor markets. I exploit county-level variation in exposure to news about labor markets impacted by fracking, to show that access to information about potential labor market opportunities affects migration. I use pre-fracking newspaper circulation rates and content from national news outlets to capture exogenous variation in exposure to news about fracking in a particular destination. I then isolate the effect of news exposure by comparing migration flows to the same destination from differentially exposed origin counties. Exposure to newspaper articles about fracking increased migration to the areas mentioned in the news by 2.4 percent on average. News exposure also increases commuting to fracking counties. Exposure to TV news has a similar impact, and positive news about fracking increases migration more than negative news. As further evidence that news matters, Google searches for the term fracking and the names of states specifically mentioned spike after TV news broadcasts about fracking. Migration responses to news about fracking are largest from counties experiencing weak labor markets, suggesting these areas see the largest benefits to information provision.

Finally, I examine how a well known government policy aimed to incentize labor force participation – the Earned Income Tax Credit (EITC)– affects labor force transitions. Less-educated single women frequently transition in and out of the labor force. Although there is evidence that the Earned Income Tax Credit (EITC) increases annual labor force participation, it is unclear how it affects these high frequency, within year employment decisions and entry and exit. By exploiting the panel nature of the Current Population Survey, I overcome challenges associated with compositional changes and estimate the impact of increases in EITC generosity on employment transitions. EITC expansions induce less-educated single women who were previously attached to the labor force to work more months, leading stronger labor force attachment and more annual weeks worked. This leads to less annual exit, suggesting that the documented impact of the EITC on labor force participation rates in part operates by keeping previously employed single women in the labor force. This highlights the importance of understanding how income support programs affect not only labor force participation, but transitions as well. Employment decisions respond to increases in the maximum EITC credit eligible to *receive* in the current year, rather than the maximum credit eligible to *earn*, which differ because the EITC is a tax credit transferred with a one year lag. This would be consistent with workers basing their current work decisions on their lagged experience with the EITC. Further evidence additionally suggests that the employment response to the lagged EITC amount is likely due to information about the return to work, rather than to the relaxation of liquidity constraints.

ESSAYS ON INDIVIDUAL RESPONSES TO LABOR MARKET CONDITIONS AND POLICIES

by

Riley Edward Wilson

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Melissa S. Kearney, Chair/Advisor
Professor Katharine G. Abraham
Professor Judith K. Hellerstein
Professor Susan Parker
Professor Sergio Urzua

Dedication

For Allison, who stands by my side and made this possible. I am grateful she never tired of hearing about "labor mobility"; and for my parents, who convinced me early on I could achieve.

Acknowledgments

I am deeply grateful to everyone who has helped shape me into the researcher and economist I have become. First and foremost I would like to thank my advisor, Melissa Kearney, for helping me develop a love of research and an appetite for answering economic questions. She is the perfect example of someone who uses economics to inform us about the world. I am grateful for Judy Hellerstein and her countless helpful comments, re-reads, and hours delving into the world of fixed effects with me. I am grateful to Katharine Abraham for giving critical and timely feedback that has made my ideas more clear and to the point.

I thank Sergio Urzua and Susan Parker for their time and insight which has helped me to improve my work and increased my ability to think critically. I also thank Lesley Turner and Nolan Pope for helping me develop my research. My work has benefited from discussions with Cody Tuttle, Fernando Saltiel, Heath Witzen, Lucas Goodman, Ken Coriale, Stephanie Rennane, and other fellow graduate students at the University of Maryland.

My experience at Maryland would not be possible without Vickie Fletcher, Terry Davis, Amanda Statland, and Mark Wilkerson. I also acknowledge the irreplaceable financial support provided by the Economic Club of Washington DC through the Vernon E. Jordan Fellowship. Last of all, I acknowledge the support of my friends and family, my Mom and Dad who have always offered advice and encouragement, Allison my equal partner and source of strength, and Benson and Lincoln who still think I am the best dad even when I am busy.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

One of the starkest trends in the United States is the increasing disparity across education, race, and geography in outcomes ranging from labor market performance (Autor, Katz, & Kearney, 2008) to income and health inequality (Piketty & Saez, 2003; Chetty et al., 2016). This has led to a growing interest in understanding (1) why certain groups and places have seemed to be left behind and (2) if anything can be done about it. In this dissertation, I focus on two specific settings, to better understand some of the constraints, frictions, and policies that might affect individuals' decisions to respond to economic opportunities. In particular I explore the impact of local economic conditions – and frictions associated with migration– on geographic mobility, as well as the impact of a low-income tax credit policy with positive work incentives (the Earned Income Tax Credit) on the labor force attachment and transitions of less educated single women. By exploring these patterns, we can better understand some of the frictions people face when making labor market decisions, and shed light on how policy can relax some of these constraints.

First I explore recent patterns in economically motivated migration. Since the 1980s, inter-state migration rates have fallen by nearly 50 percent (Molloy et al., 2011). Recent evidence also suggests that people are unlikely to move away from poor performing labor markets (Monras, 2015), and have become less responsive to negative local economic conditions over time (Dao et al., 2017). This trend has drawn attention, as researchers consider the traditional role of migration as a mechanism through which labor markets adjust (Blanchard & Katz, 1992; Molloy et al., 2016). Recent work has exploited negative economic shocks associated with trade liberalization and the Great Recession to look at migration responses into and away from negative labor market conditions, to better understand economically motivated migration in the current context of low geographic mobility (Autor, Dorn, & Hanson, 2013; Autor et al., 2014; Hakobyan & McLaren, 2015; Cadena & Kovak, 2015; Foote et al., 2015; Monras, 2015; Foote, 2016). However, it is possible the response to positive labor market shocks might be asymmetric. In the next chapter, I document the migration response to positive local labor market shocks associated with fracking, and provide recent estimates of the short run elasticity of migration with respect to average earnings.

The geological and technological constraints associated with fracking have led to large labor market improvements in very well-defined local labor markets, that are likley not a function of pre-existing conditions or trends. Using these geological constraints and technological variation, I construct simulated measures of production to capture these plausibly exogenous labor market shifts and examine short run migration responses. First, these labor market impacts have been large. Average earnings have increased by 5-27 percent across the various fracking counties. Using the Internal Revenue Service (IRS) Statistics of Income (SOI) migration flows, in a reduced form setting, I estimate that these measures of simulated production are associated with large increases in both in-migration and out-migration rates. Using the American Community Survey (ACS), I find that both in-migrants and out-migrants have similar characteristics, suggesting fracking has led to increased churn as people move, stay for a few years, and then leave.

The data suggest that this migration response has also been heterogeneous across the different states involved in fracking. As fracking has impacted labor markets differently across the different states, I next impose more structure and estimate the short run migration response as a function of average earnings in an instrumental variables (IV) framework. This allows me to see if migration responded differently across states to a similar sized increase in average earnings. This estimation requires the strong identifying assumption that my measures of simulated production only affect migration through its impacts on local labor markets. In order to mitigate the effects of fracking on other markets that also interact with migration (e.g., housing markets), I focus on the short run analysis, while production is still expanding and markets have not fully adjusted or equilibrated.

When looking across regions there is still substantial heterogeneity in the migration response. For a ten percent increase in earnings the in-migration rate in North Dakota increased by 3.8 percentage points, but only by 2.4 percentage points in the West, 1.5 percentage points in the South, and 0.5 percentage points in the Northeast. These geographic disparities are significant, and robust to potential confounding factors such as changing house prices or local geographic spillovers.

I next explore several mechanisms to understand why migration has reacted differently across regions. These geographic disparities are only partially closed when accounting for additional commute behavior, differences in initial population characteristics (such as population size), or potential non-linearities in the relationship between earnings and migration. However, I do find a suggestive correlation: fracking counties that experienced more newspaper publicity saw more migration from the places where this information was published. Overall these findings would suggest that for some types of positive labor market shifts (such as those created by fracking) we still observe an increase in migration, despite the overall low levels of geographic mobility. However, the geographic heterogeneity would suggest that both labor market and non-market factors, such as a lack of information, can affect migration outcomes.

Given the positive correlation between news flows and migration flows documented in the previous chapter, I next attempt to estimate the causal effect of exposure to labor market information on migration decisions. Traditionally, migration is viewed as a human capital investment (Shultz, 1961), which allows individuals to encounter more favorable locations and labor market opportunities. However, if people lack information about labor market opportunities in other locations, this investment decisions becomes more uncertain, and might reduce people's willingness to invest and move. The role of information is frequently overlooked in the research, in part because it is difficult to measure and identify a causal relationship. By focusing on national news about fracking I am able to isolate one particular source of information transmission and estimate its impact on origin-destination specific migration flows. As explained before, the productivity of fracking was constrained by geological characteristics and aggregate technological innovation. This led to sudden, large, and highly publicized increases in both employment and earnings. Previous work has found these increases to be persistent and evident across industry (Feyrer et al., 2017). As such, fracking created plausibly exogenous, positive labor market shifts, that were talked about in the news. Because fracking was so novel – as a technology– it also introduced new words and vocabulary, making it easy to parse newspapers and TV news to see which sources were talking about fracking, what places they were talking about, what they were saying, and when they were saying it. This allows me to look across counties and determine how exposed residents were to news about fracking in a specific destination by measuring to what extent that news was circulated in their local area.

Although the labor market shifts were plausibly exogenous, the news and exposure to the news about fracking was not necessarily. Local newspapers make content decisions to cater toward their local geographic preferences (Gentzkow & Shapiro, 2010), and individuals can choose which news sources they consume. However, Gentzkow and Shapiro (2010) suggest that there are certain, large national newspapers that do not have well defined geographic markets but rather make content decisions to cater to the nation as a whole (such as the *USA TODAY*, *New York Times*, and *Wall Street Journal*). Also, since fracking began quite suddenly, it is possible to look at circulation rates of these national news papers during a "pre-period", before fracking began, to capture penetration measures that are not endogenous to preferences about fracking. By combining destination specific content from national news sources, with pre-fracking circulation rates I am able to isolate a plausibly exogenous component of exposure to news about fracking in a specific place.

Even though national newspapers are unlikely to adjust their content to each county's idiosyncratic trends and preferences, it does seem likely that they will produce more content about fracking places that are more impacted by fracking. This is potentially problematic for identification if these characteristics also directly affect migration outcomes. To overcome this challenge, I control for origin/destination pair specific characteristics (like distance) and then compare migration flows from different origins – with different levels of exposure– to the same destination, to hold constant any characteristic of the destination that might be changing and affect migration behavior. For example, in 2012 the *USA TODAY* published six articles about fracking in Pennsylvania. This baseline specification tests to see if origins that had historically higher circulation of the *USA TODAY* (and thus higher exposure) saw larger increases in migration to Pennsylvania fracking counties when this news is distributed. By exploiting within year, within destination variation this specification relies on variation across origins in historic circulation to identify the effect of news exposure on migration. I estimate that when counties are exposed to this national newspaper news about a particular destination state (i.e., North Dakota or Pennsylvania) the annual number of migrants to fracking counties in that state increase by 2.4 percent on average. This translates into approximately one to two additional people moving to the fracking destination from each origin on average. Although this response is small, it is economically significant given the scope of the "treatment" and the aggregate effects at the destination.

One remaining concern for identification is that the pre-fracking level differences in circulation – which generate the identifying variation – might be correlated with other origin level characteristics that are changing over time and affect migration. For example, counties with high readership of the *New York Times* might be more affluent and see larger income growth over time. If this additional growth in income affects migration decisions, the estimates would be biased. I find that areas with high circulation and low circulation follow similar trends in migration to fracking areas in the pre-period, but pre-fracking circulation rates do predict small, significant changes in other origin level characteristics (such as unemployment rates).

Because, there are 16 states involved in fracking, there are 16 potential destinations for each origin. This allows me to include an origin by year fixed effect and account for any observable or unobservable characteristics of the origin that are changing over time and affect migration. This specification controls for the characteristics of the destination that are changing over time and tests if origins saw more migration to the destinations that they got more news about, due to the fact that the different national news sources discuss different locations with varying frequency and that origins have pre-existing differences in the composition of newspaper readership. When making this comparison, the estimated effect of national news exposure is virtually the same, suggesting that when counties are exposed to this news there is an increase in migration to the destinations being discussed, and it is not driven by characteristics of the destination that are changing over time, or characteristics of the origin that are changing over time. The estimates are also robust to controls for local news exposure, various functional forms, sample restrictions, and estimation in levels.

Using this strategy, the remaining threats to validity are origin by destination specific characteristics that are (1) changing over time, (2) correlated with pre-fracking circulation rates, and (3) affect migration decisions. For example, perhaps counties with high circulation rates are more sophisticated or more tied to the oil and gas industry and therefore more likely to move to booming local economies precisely when that boom happens and gets talked about in the news. As the current strategy cannot control for these unobservable origin/destination links I employ a second identification strategy to determine the importance of these factors. Using the county circulation data from all domestic newspapers that published a story about fracking, I back out each newspapers geographic distribution market. I then identify counties on the border of this market, as well as their contiguous neighbors just outside the market (who do not get copies of the newspaper). I then compare migration flows to the destination being mentioned in the news between the counties that did get the newspaper and their neighboring counties that did not. When doing this I still find a significant role for information in the news affecting people's migration behavior, suggesting this is not just driven by changing pair specific characteristics.

Given this robust relationship, I next explore several aspects of the information/migration relationship to better understand how information provision affects migration. The news could be providing general information about the labor market impacts of fracking, but it might also provide location specific information about a certain labor market that is affected. I find that news about fracking in a specific state leads to more migration to fracking counties in that state than news about fracking elsewhere, suggesting this locational component is important. Also, people can take also advantage of distant labor market opportunities by commuting rather than moving. Using the Longitudinal Employer-Household Dynamics (LEHD) Origin/Destination Employment Statistics (LODES) I examine the effect of this news on commute behavior. Exposure to national news about fracking leads to increases in commuting to the destinations being discussed, with a larger effect than the effect on migration. The LODES also provides some broad demographic cuts, and the data suggest that workers between 30 and 54 are the most responsive, consistent with age differences in newspaper readership and the importance of print news as a source of information (Pew Research, 2013), suggesting the source of news will affect which populations respond.

Given the potential importance of the source, and the knowledge that many people rely on TV for news, I next test to see if TV news affects migration and commute behavior. Using the Vanderbilt Televisions News Archive (VTNA) archived nightly news broadcasts and local TV circulation rates from the Cable and Television Factbook I construct a similar measure of TV news exposure. I find that exposure to TV news significantly increases commuting, but only weak evidence that it has an effect, separate from newspaper news, on migration. Even though TV circulation rates are much higher than newspaper readership rates, the magnitude of these effects are similar. All of the TV news broadcasts were relatively short (one to five minutes) and provided less content than the newspapers, suggesting the impact of the information will depend on both the penetration of the news, as well as the quality of the content. Consistent with this I find that positive news (such as stories about jobs, growth, and booms) has a significantly larger effect on migration than negative news (stories about earthquakes and water contamination). However, both positive and negative news lead to increases in migration to the places being discussed, once again suggesting the importance of the locational signal.

To understand the mechanisms through which newspaper exposure might lead induce people to move and commute, I turn to Google search interest data. Using Google Trends, I find that search interest in the term fracking, and the names of the specific states mentioned in the news spikes directly after the news is broadcast, suggesting that people go online and search for more information. It is possible that the news provides an initial piece of information that then directs their future search.

Lastly, we know that migration from areas in decline is low, but we are not sure why or what –if anything – can be done about it. Given these results, I test to see if news exposure has heterogeneous effects across the different types of labor markets. The data suggest that the effect of news exposure is over twice as large in weak labor markets as it is in strong labor markets, even though they experience similar levels of exposure. This would suggest that providing labor market information can be a way of increasing geographic mobility, but that it might be particularly effective if targeted toward weak labor markets where the returns to migration are plausibly the largest and where we have also observed non-responsiveness in the past. This would suggest there is a potential policy lever of providing labor market information to increase geographic mobility. However, at this point it is unclear how this affects eventual labor market outcomes, what happens to the distribution or composition of workers, and how overall welfare is affected. As such, more work is needed to understand if this policy lever is an effective way of increasing geographic mobility and changing labor market fluidity and perhaps overall economic mobility.

Given this potential role for policy to improve economic mobility, in the final chapter I turn to a government intervention that has been designed to help support low-income households but also encourage labor force participation: the Earned Income Tax Credit (EITC). The EITC is credited with lifting millions of people out of poverty (Hoynes & Patel, 2016), and differs from other transfer programs in that it positively incentivizes work. A large, existing literature has found evidence that the EITC increases annual labor force participation rates. However, the labor force participation of one population that is highly exposed to the program, less-educated single women, is characterized by high levels of turnover and churn as these women cycle in and out of the labor force (Edin & Lein, 1997). At present it is unclear how the work incentives of the EITC affect less-educated single women's within year labor force attachment and exit from the labor force.

Using the short panel nature of the Current Population Survey (CPS) I link households across months and across years to determine how expansions in EITC generosity affect within year transitions and employment duration as well as transitions into and out of the labor force at an annual level. From the monthly data I am able to observe a four month employment history for each woman exactly one year apart. Importantly, this allows me to separately examine impacts by previous labor force attachment and avoid the compositional concerns that previous work has faced when examining similar within year intensive margin responses. From this panel, I estimate that increases in EITC generosity slightly increase the share of months worked, leading to stronger labor force attachment. As this increase is small, women are still likely to exit the labor force, but there is evidence that they exit less. This would suggest that increases in the EITC led to stronger labor force attachment and longer employment spells. However, this response is only observed among women who were employed during the first wave of the panel, suggesting the EITC increases attachment among those with some previous labor force attachment. These patterns are not observed among more educated women, or women who reported incomes above the EITC refund schedule.

Consistent with this increase in labor force attachment, these same women end up working more weeks during the year following the expansion. Using the annual short panel, I am also able to examine the impact of the EITC on annual labor market transitions. Increases in the EITC during the early 1990s have a noisy, insignificant effect on total participation rates (consistent with previous work), but there are stark patterns when conditioning on participation during the initial survey wave. On the one hand, Less-educated single women with previous labor force participation are more likely to be employed the year after an expansion, effectively reducing annual exit from the labor force. On the other hand, women who did not initially participate see no change in the probability of working after the expansion (i.e., no additional annual entry). This would suggest that the EITC increases labor force participation rates in part by delaying exits that would have happened, rather than by bringing new entrants into the labor force.

Across all specifications, these employment decisions are responding to changes in the maximum EITC a woman was eligible to *receive* in the current year, rather than the maximum EITC she was eligible to *earn*, which differ because the EITC is transferred as a tax credit with a one year lag. As only women who worked in the previous year (and would become eligible for the credit) respond, this pattern of results is consistent with women lacking information about the program and updating their employment decisions after filing their taxes and observing the additional benefits to work created by the EITC. Alternative explanations, such as the EITC relaxing liquidity constraints are less supported by the data.

Overall the data would suggest that EITC expansions increase labor force attachment and lead to longer employment spells. However, these impacts are only observed among women with previous attachment, suggesting the EITC operates by keeping individuals with previous labor force attachment in the labor force, rather than bringing new households into the labor force. This impact on labor market transitions and dynamics might in part be due to people's limited awareness of the EITC during the time period, and the learning that can occur with a lag when taxes are filed.

This dissertation provides evidence that people's labor market decisions respond to incentives created by both policy and natural economic shocks. However, it also highlights the various constraints and information frictions that might factor into these decisions and sheds light on potential interventions that can address these market frictions. Throughout the following chapters I analyze these responses to better understand why these frictions matter, and what effective policy would need to look like to improve individual's access to economic mobility.

Chapter 2: Moving to Economic Opportunity: The Migration Response to the Fracking Boom

2.1 Introduction

Migration provides an opportunity for individuals to encounter more favorable labor market conditions and improve their economic wellbeing. However, since the 1980s, geographic mobility within the US has fallen by nearly 50 percent with internal migration rates across all demographic groups at the lowest they have been in decades (Molloy et al., 2011; Molloy et al., 2016). This trend has led to a growing concern that people no longer move to better labor market opportunities.[1] Recent academic work has exploited negative economic shocks associated with trade liberalization and the Great Recession to look at migration responses into and away from negative labor market conditions (Autor, Dorn, & Hanson, 2013; Autor et al., 2014; Hakobyan & McLaren, 2015; Cadena & Kovak, 2015; Foote et al., 2015; Monras, 2015; Foote, 2016), but it is unclear if we should expect a symmetric response to positive economic shocks. In this paper I document the migration response to po-

[1] See, for example, newspaper articles in the New York Times by Brooks (2016) https://www.nytimes.com/2016/05/21/opinion/how-to-get-americans-moving-again.html and by Cohen (2016) https://www.nytimes.com/2016/05/25/business/economy/fewer-workers-choose-to-move-to-new-pastures.html and in the Washington Post by Fletcher (2010).

sitive local labor market shocks and provide estimates of the short run elasticity of migration with respect to average earnings, in the current context of low geographic mobility.

Identifying a causal relationship between labor market opportunities and migration outcomes requires variation in local labor market opportunities that is exogenous to migration preferences and non-labor market local characteristics that might affect migration decisions. Fracking led to large, positive, localized economic shocks that were largely defined by geological constraints and the introduction of technology over time, rather than initial labor market conditions. Using detailed well-level production data, I exploit these geological constraints and temporal variation to create a predicted measure of exogenous fracking production, similar to a simulated instrument, and then use this measure to identify the short run reduced form impacts of fracking on local labor markets and migration across regions. Then under more strict assumptions, I relate earnings to migration using an instrumental variables (IV) strategy to understand how migration responds to labor market improvements. To shed light on how market and non-market factors influence the migration decision, I also explore demographic and regional heterogeneity to characterize who moves to fracking and where they are moving.

Using data from the Quarterly Workforce Indicators (QWI), I reconfirm that fracking led to large gains in both potential earnings and employment. However, these labor market gains vary significantly across geography. Among high intensity fracking counties, fracking production increased earnings by over 27 percent in North Dakota, and between 5 and 22 percent in many other states highly involved in fracking by 2013.

I then use county level migration data from the Internal Revenue Service (IRS) Statistics of Income (SOI) to estimate the reduced form migration response to the localized fracking booms which caused these labor market improvements. In contrast to the recent literature exploiting negative shocks, the data suggest that there was significant migration to these positive labor market shocks. There is also distinct geographic heterogeneity, with migration increasing the population in North Dakota fracking counties by 12 percent on average, but by less than two percent on average in fracking counties in other states. These net impacts mask substantial churn. In-migration is concentrated in North Dakota, where between 2010 and 2013, a flood of in-migrants, nearly equal to 25 percent of the baseline county population, entered high intensity fracking counties in North Dakota. Migration to other fracking regions did occur, but to a lesser extent.

Fracking was also associated with an increase in out-migration suggest that either certain people were systematically sorting away from areas with fracking, or that the migration was short term, resulting in more churn. While systematic sorting might change the demographic and skill composition of the local population and labor force, short term migration might impose additional costs on firms demanding labor and affect local labor market dynamics for both natives and migrants. To separate these channels and characterize who is moving to and away from fracking, I use individual level migration decisions from the American Community Survey (ACS). The demographic groups that face larger labor market incentives or traditionally face fewer moving constraints are the most likely to migrate to fracking areas. They are more likely to be male, younger workers, unmarried, and either be a high school dropout or college graduate than the population as a whole and migrants more generally. The same groups are also more likely to move away from fracking, suggesting that fracking has led to increased churn and short term migration, with little evidence of systematic sorting away from fracking areas along observable characteristics. No previous academic work has characterized the types of people moving to fracking or documented the short-term nature of migration, which likely has broader impacts on labor market dynamics. The prevalence of short-term migration also suggests that the monetary fixed costs of moving (e.g., renting a moving truck) are not insurmountable.

To estimate how migration responds to positive labor market opportunities in the current context of low labor mobility, I impose more structure and estimate the relationship between in-migration and earnings in an IV framework where I use simulated production from new wells to instrument for average earnings. Rather than just capturing differences across regions in fracking intensity, this specification also allows me to test if in-migration rates responded to potential earnings differently across regions. I allow this relationship to vary by region and find that, for a ten percent increase in earnings, an additional 3.8 percent of the baseline population moved into North Dakota fracking counties, but only 2.4 percent in the West, 1.5 percent in the South, and 0.5 percent in the Northeastern fracking states.[2] I re-estimate these elasticities accounting for geographic spillovers and potential con-

[2]For comparison, Monras (2015) finds that a 10 percent decrease in GDP per capita reduced in-migration on the order of 2-3 percent of the baseline population. Foote et al. (2015) find that when 10 percent of the labor force is laid off, 0.6-0.8 percent of the population leaves.

founding changes in the housing market, as well as a range of other specifications and the patterns do not significantly change, suggesting that people did respond to these positive labor market shocks, but were more likely to move to earnings gains in North Dakota than elsewhere.

I explore several potential explanations for this geographic disparity in the migration response. Many workers respond to potential earning gains by commuting to nearby fracking locations rather than moving, but this only widens the gap between North Dakota and elsewhere. The gap is only partially explained by differences in initial population characteristics across regions. A non-linear relationship between earnings and in-migration might play a role, but the gap is still present when comparing counties that experienced similar earnings increases from fracking. There is, however, geographic heterogeneity in the amount of information about each localized boom, with fracking in North Dakota receiving a disproportionately large share of media attention per capita. I find that fracking counties that experienced more newspaper publicity saw more migration from the places where this information was published. This suggests that non-market factors, such as information, might play an important role in individuals' decisions to move to better labor markets, and should be explored further.

This paper makes several contributions. First, I characterize the migration response to some of the largest positive, local economic shocks in recent decades. In doing so, I am able to characterize which types of people move and where they move to, which has not been examined in the previous literature.[3] I also show that

[3]A contemporaneous working paper by Vachon (wp 2015) uses net migration flows and adjusted

the migration response to fracking is short-term in nature and that many workers take advantage of the potential earnings gain through commuting. In addition to shedding light on how various costs and factors might enter the migration decision, these findings also reveal compositional effects that are likely relevant to research looking at the impacts of fracking on other outcomes, such as local governance or educational attainment, where characteristics of the population might matter. This paper also highlights the role that both market and non-market factors can play in migration decisions. Understanding these factors will help identify potentially effective policy interventions aimed to increase economic mobility.

This chapter proceeds as follows. In Section 2.2 I outline a simple migration choice model and highlight the relevant empirical literature. In Section 2.3 I discuss the details of fracking and the small, recent literature exploring its effects. In Section 2.4 I explain my data, empirical strategy, and counterfactual similarity. Section 2.5 describes the reduced form and IV results, while Section 2.6 explores potential explanations for the geographic heterogeneity. Section 2.7 concludes.

gross incomes from the IRS for counties in North Dakota, South Dakota and Montana from 1999 to 2010 to estimate the elasticity of net migration with respect to income. She uses a difference in differences IV approach where the instrument is estimated oil reserves. She does not consider inflows or outflows, demographic differences, or potential differences across other regions. Another contemporaneous working paper by Bartik (wp 2016) is focused on the role of moving costs in migration decisions and exploits variation in local labor markets from shale play reserves in some specifications, although this is not emphasized. He only looks at differences by education and does not explore differences across geography.

2.2 Background: The Decision to Migrate

2.2.1 A Simplified Migration Choice Model

The economic literature exploring the role of potential earnings in migration decisions date back to Hicks (1932) and Sjaastad (1962). The simplest models of migration represent an individual's (i) decision to move (m_{iod}) between two locations, an origin (o) and a destination (d), as a static discrete choice comparison of indirect utilities (cf. Borjas, 1987,1999), as follows

$$m_{iod} = \begin{cases} 1 & \mathit{if}\ V_{id} - c_{iod} \geq V_{io} \\ 0 & \mathit{else} \end{cases} \tag{2.1}$$

where the indirect utility for region j , V_{ij}, depends on potential earnings ($w_{ij}(\mu_j, \varepsilon_{ij})$) which are a function of both a region-specific mean and idiosyncratic component, and the individuals' valuation of regional amenities ($\lambda_i'\theta_j$). Individuals also face moving costs, c_{iod}, which can be both monetary and psychic.[4] This indirect utility function is often modeled linearly, as $V_{id} = \mu_d + \varepsilon_{id} + \lambda_i'\theta_d$, so that an individual will find it optimal to move if

$$\varepsilon_{io} - \varepsilon_{id} \leq (\mu_d - \mu_o) + \lambda_i'(\theta_d - \theta_o) - c_{iod}. \tag{2.2}$$

The decision to move depends on earning differentials ($\mu_d - \mu_o$), the evaluation

[4] This simple model has been extended to allow agents to choose between multiple potential destinations (Borjas, Bronars, & Trejo, 1992; Dahl, 2002), and dynamic decisions (Kennan & Walker, 2011).

of regional amenity differences ($\lambda_i'(\theta_d - \theta_o)$), the individual's moving cost (c_{iod}), and individual selection ($\varepsilon_{io} - \varepsilon_{id}$) which is unobserved to the econometrician, but potentially observed by the individual. Given the distribution of $\varepsilon_{io} - \varepsilon_{id}$, the probability of individual i moving can be calculated as

$$Pr(m_{iod} = 1|\mu_o, \mu_d, \theta_o, \theta_d, \lambda_i, c_{iod}) = Pr(\varepsilon_o - \varepsilon_d \leq (\mu_d - \mu_o) + \lambda_i'(\theta_d - \theta_o) - c_{iod}). \quad (2.3)$$

This model is often used to conceptualize the issue of self-selection into moving, but is informative when considering regional shocks to labor markets. Suppose there is an exogenous labor market shock in region d (perhaps due to fracking) that increases μ_d. For all individuals, the probability of moving will increase, but the response will be heterogeneous. For example, demographic groups that face lower moving costs on average (such as young workers who do not own homes, or unmarried workers who do not need to move a family) should be more sensitive to shocks. These differences across demographic groups can be empirically verified.

In reality, the migration decision is likely more complicated: decisions could vary by initial location relative to the shock; individuals might choose across multiple locations; earnings and amenities might enter the decision non-linearly; a shock could differentially affect earnings across demographic groups; or even the spread of earnings could be affected by a shock like fracking all of which might affect who self-selects into moving and where they chose to move. For this reason it is important to understand heterogeneity across both demographics and regions as well as the separate decisions of moving in and moving out (Monras, 2015).[5]

2.2.2 Previous Empirical Studies

Empirically identifying the relationship between labor markets and migration requires variation in local labor markets that is exogenous to migration decisions and other local conditions. Previous work has relied on structural identification (Kaplan & Schulhofer-Wohl, 2017; Kennan & Walker, 2011), shift-share instruments (Bound & Holzer, 2000; Wozniak, 2010), or exogenous local economic shocks (Black et al., 2005; Carrington, 1996). The identifying variation I use most closely follows that exploited by Carrington (1996) looking at the Trans-Alaska pipeline in the 1970s and Black et al. (2005) looking at the Appalachian Coal Boom in the 1970s and 1980s. Both studies find that the for a one percent increase in earnings, the total population increased by approximately 0.16 percent. As both of these shocks occurred when migration levels were still relatively high, it is unclear how they relate to migration today. Previous work has highlighted demographic differences in migration to other labor demand shocks, mostly focusing on differences across education (Bound & Holzer, 2000; Dahl, 2002; Malamud & Wozniak, 2010; Wozniak, 2010) or the differential incidence of labor demand shocks (Notowidigdo, 2013). I examine demographic differences to characterize those that move to fracking, and I also explore differences across geography as fracking spans many areas. As stated before, only two working papers have considered migration to fracking in a much

[5] Local labor market adjustments to labor demand shocks can also occur through commuting (Monte, Redding, & Rossi-Hansberg, 2015), for this reason I also consider commute behavior.

more limited context and do not address important demographic and geographic differences (Bartik, wp 2016; Vachon, wp 2015).

A recent literature has developed exploring the migration response to negative shocks such as the Great Recession and trade liberalization. Work looking at the local labor market impacts of trade liberalization found that, in general, the population was not very responsive to negative shocks (Autor et al., 2013, 2014; Hakobyan & McLaren, 2015). In response to negative shocks from the Great Recession, out-migration increased and in-migration decreased (Foote et al., 2015; Monras, 2015). However, relative to earlier periods, labor market non-participation also increased suggesting the mobility response has become smaller (Foote et al., 2015). These migration responses have been found to vary with home ownership and home equity (Foote, 2016) as well as by nativity (domestic vs. Mexican-born) (Cadena & Kovak, 2015).

The existing literature has also considered the issue of short versus long term outcomes. The individual migration choice model predicts that an exogenous shock to earnings will increase migration ceteris paribus, but in a spatial equilibrium other markets (such as the housing market) might respond to increasing wages, or changes in migration (Roback, 1982; Rosen, 1974).[6] In any particular context, the degree to which other markets and amenities adjust and offset a positive earnings shock is an empirical question, and might differ in the short and long run. My analysis is

[6] An alternate conceptual framework, following Blanchard & Katz (1992) looks at migration as a mechanism by which labor markets adjust to shocks and converge to a new equilibrium. This model is more interested in the general equilibrium and dynamics than the individual specific decisions. For this reason I focus on the migration choice model, but draw on both models to inform my empirical analysis.

a short run analysis, and I return to a discussion of this issue when I present the results.

Migration responses to fracking should be placed in the context of current migration in the US. Since 2000, annual interstate migration rates have been about half the level observed in the 1980s (Molloy et al., 2011).[7] There is currently no consensus on what has driven this change. Some hypotheses highlight the role of frictions that lead to suboptimal migration levels. For example, more binding liquidity or credit constraints (Ludwig & Raphael, 2010), the rise of two-earner households (Molloy et al., 2011), and increased land-use regulation (Ganong & Shoag, 2017), might keep certain groups from moving or finding a high quality locational match. Other hypotheses suggest that the current low levels of migration are not necessarily suboptimal. The psychic costs of moving might have increased (Cooke, 2011; Fletcher, 2010; Kotkin, 2009; Partridge et al., 2012), or improvements to communication technology and falling geographic specialization might mean workers no longer have to move to take advantage of wage gains (Kaplan & Schulhofer-Wohl, 2015; Molloy et al., 2011).[8]

[7] The decrease described by Molloy et al. (2011), accounts for the methodological change in imputation in the CPS (Kaplan & Shulhofer-Wohl, 2011).

[8] There are two other strands of economic literature looking at migration that are related to the present paper only tangentially. The first, is the evaluation of the Moving to Opportunity (MTO) experiment (cf. Kling, Liebman, & Katz, 2007). Rather than examining why low-income and low education households do not migrate, the MTO experiment informs us on what might change when someone does migrate. The other literature examines welfare migration (Gelbach, 2004; Goodman, 2016; McKinnish, 2005; Moffitt, 1992). This literature is relevant, in that it examines individual's migration decisions when monetary incentives change, but is interested in a population with different skills and labor market attachment.

2.3 Background: Fracking in the United States

Throughout the United States, there are several regions where layers of low permeability shale rock have trapped natural gas and oil molecules. These shale rock formations lie miles below the Earth's surface and are referred to as shale plays (outlined in black in Figure 2.1). Prior to the 2000s, oil and gas extraction from shale plays was technologically infeasible because conventional vertical drilling without fracking could not extract gas or oil at the molecular level. At a fracking well, a mixture of water, sand, and chemicals is pumped into the well at extremely high pressure, causing the rock to fracture and relieve pressure.[9]The water is removed leaving the sand to prop open the fractures, and the gas (shale gas) or oil (tight oil) escapes into the well due to the pressure release. By combining fracking with horizontal drilling, wells can be constructed that run parallel to the horizontal layers of shale, allowing for more extractable area from the same well opening. In essence, these combined technologies made extraction from shale both feasible and profitable. These technological innovations, combined with high prices, fueled localized fracking booms. Prior to 2005, shale gas and tight oil production was almost non-existent (see Figure 2.2). However, by 2014, there was over $80 billion (2010$) of tight oil production and nearly $50 billion of shale gas nationwide. Fracking has been particularly intensive in ten states, each with over a thousand wells drilled and fracked and over two billion dollars of oil and gas extracted.

Although the presence of some of these plays was known, they were not belie-

[9]The concept of well fracturing has been used for nearly 50 years. However, advances in the process around the turn of the 21st century made it more effective and less costly (Gold, 2014).

ved to hold extractable resources and had no economic value attached to them. The rapid innovations in resource extraction directly affected the production function of gas and oil in these shale plays, creating quasi-experimental variation in fracking potential that is not driven by preexisting population and labor market characteristics which might enter migration decisions.

As fracking rapidly expanded, local labor demand shifted out and created large and significant increases in both employment and earnings (Allcot & Keniston, 2014; Eliason, 2014; Fetzer, 2014; Feyrer et al., 2015; Maniloff & Mastromanaco, 2014). These increases spread across county borders and to other industries, suggesting fracking created a shock to the local labor market, rather than just the industry (Feyrer et al., 2017; Maniloff & Mastromanaco, 2010). These labor market impacts suggest migration incentives might exist.

If people expect the boom to be short lived, they might not move even if labor market gains are large.[10] Although there is not much more than anecdotal evidence on workers expectations, industry executives, market professionals, and political figures viewed fracking as a long run shock to regional economic activity. For example, executives at Chesapeake Energy, one of the largest natural gas extraction companies, expected prices to remain high for many years as demand shifted away from coal to natural gas (Gold, 2014). Current predictions from both the Energy Information Administration (EIA) (2015) and independent researchers (Lasky, 2016) suggest

[10]Work looking at oil booms in the 1970s and 1980s finds that although labor markets improve substantially during the boom, the negative effects are even larger during the bust (Jacobsen & Parker, 2014). This has raised concerns about fracking leading to a "natural resource curse" and Dutch Disease; multiple authors have not found evidence of this (Allcott & Keniston, 2015; Maniloff & Mastromonaco, 2014).

long run expansion and only temporary slowing from falling prices. Although falling prices and well depletion rates have caused some to question the sustainability in recent years (Hughes, 2013), this was initially viewed as a long run shift in economic activity.[11]

Importantly, recent working papers have also found that fracking impacts high school students' graduation decisions (Cascio & Narayan, 2015) and local public finance (Bartik et al., 2017; Newell & Raimi, 2015), and provides mixed evidence that crime rates have adjusted (Bartik et al., 2017; Feyrer et al., 2017; James & Smith, 2014). Perhaps the most relevant to migration is the effect on local housing markets. For data reasons, most of this work has focused on housing markets in Pennsylvania and New York, where shale gas development has positively affected home values, although homes very close to fracking or dependent on private wells saw a drop in prices (Gopalakrishnan & Klaiber, 2014; Muehlenbachs et al., 2015; Boslett, Guilfoos, & Lang, 2016). Looking across the US, Bartik et al. (2017) find that housing values increased by about 6 percent. To understand the relationship between fracking and local labor markets, it will be important to econometrically control for these potentially confounding factors.

[11] In his 2012 State of the Union Address, President Obama suggested that domestic natural gas supplies found in shale plays would last 100 years and support over 600,000 jobs by the end of the decade (State of the Union, 2012).

2.4 Data and Empirical Approach

2.4.1 Data

Estimating the effect of fracking on local earnings and migration requires local labor market level data on earnings, migration, and fracking. I briefly describe my key data sources and provide a full explanation in the Data Appendix (Section 2.11 Appendix B). I use the QWI to construct annual county-level measures of employment and average earnings for all workers in the county which I can separate by industry, gender, and education (U.S. Census Bureau, 2014). To measure migration I use the county migration flows provided by the IRS SOI. The IRS only provides the number of households and individuals that moved into or out of a county, without demographic identifiers. This data only captures internal migration and might miss foreign immigrants and low income households that are not required to file taxes. To explore differences across demographics I use the public-use microdata from the 2005-2011 ACS to look at individuals who move (Ruggles et al., 2015). The lowest geographic level of migration available in the public-use ACS is the migration public use microdata area (MIGPUMA), which often encompasses several counties.[12] This data provides a rich set of demographics and allows me to identify individuals who moved into and away from fracking regions. One weakness of migration data in the United States, is that it does not fully capture temporary relocations. By looking at both in- and out-migration, individual-level data, and commuting data, I can make

[12]In 2012, the MIGPUMA delineations were updated and no longer correspond to the same geographic regions. For this reason I only use the years 2005-2011 when the geographies were consistent.

some inferences about short term migration.

This data is then combined with well level production data obtained through a restricted-use agreement with the private company, DrillingInfo. This data provides detailed information including the exact location, drilling date, well type, and quarterly oil and gas production. As in Feyrer et al. (2017) and Cascio and Narayan (2015), I identify non-vertical wells as fracking wells. I then combine this data with county boundary shapefiles (provided by the Census) and shale play boundary shapefiles (provided by the EIA) to determine if counties and shale plays intersect, which is used to identify variation in fracking potential due to exogenous geological constraints.[13]

2.4.2 Identifying Exogenous Variation in Production

One could exploit variation in oil and gas production from new wells as a local shock to estimate the reduced form impact of fracking on labor markets and migration. However, oil and gas extraction firms might choose to drill more in counties with more favorable labor market or legal conditions. As such, using the actual drilling intensity to compare counties might introduce omitted variables bias if the same characteristics that attract firms also affect individual earnings and migration decisions. Anecdotally, decisions about drilling were largely a function of estimated reserves, and how quickly firms could gain access to mineral rights, not characteristics of the local population (Gold, 2014). Once a potentially productive

[13]A special thanks to Lisa Boland and Michael Bender of the Geography Department at the University of Maryland for their help calculating areas in ArcGIS.

shale play was confirmed, extraction firms would quickly send out "landmen" to sign leases with local mineral rights owners before the competition did. Once enough acreage was leased, the firm would begin the drilling and fracking process (Gold, 2014). Even so, some of the decision might be endogenous to migration.

Fracking production at both the extensive and intensive margin strongly depends on exogenous geological characteristics and the current levels of technology and prices. To isolate exogenous variation in fracking production I follow the method of Feyrer et al. (2017) and simulate the annual county-level production from new wells as a function of exogenous geological characteristics (to capture differences in feasibility and inherent productivity) and time variation (to capture variation in aggregate technology and prices). Specifically, I take the sample of counties with shale play and estimate

$$ln(new\ prod._{ct}+1) = \alpha_c + \sum_{\tau}\sum_{j=1}^{J} \theta_{\tau j} I\{county\ c\ over\ shale\ play\ j\} * I\{year = \tau\} + \nu_{ct} \tag{2.4}$$

where *new prod.*$_{ct}$ represents the total dollar amount of oil and gas production in county c from wells that started producing in year t, and is constructed from well level production data and annual prices from the EIA. Using the log of one plus production as the outcome in equation (2.4), allows me to include non-producing counties in the simulation and isolate exogenous variation along both the extensive and intensive margin of production.

The vector of coefficients $\theta_{\tau j}$ traces out the average effect of being in shale play j in each year. This is done by interacting an indicator for intersecting a shale play, as constructed from the county and shale play boundary shapefiles, with year indicators, to account for year to year changes in world prices and technology. Although there are 48 individual shale play boundaries, I allow counties to be in multiple plays and combine small plays that cover less than nine counties into an "other" category so that total play production is not driven by any one county. I also include a county fixed effect to capture time invariant county specific differences in reserve intensity.

I then exponentiate the predicted values from equation (2.4), subtract one, and call this transformed prediction, simulated new production. This transformed variable captures exogenous variation in new production associated with the geological and time constraints. Simulated and actual production are highly correlated (p=0.68), and the F-statistic on the joint test of the interactions in equation (2.4) is over 61, suggesting that considerable variation in drilling is in fact due to exogenous geology and time, as suggested by the anecdotal evidence.[14] As seen in Figure 2.1, county level simulated new production is the highest in plays that are conventionally viewed as inherently more productive.

I can now estimate the causal impact of fracking on labor market and migration outcomes by comparing counties with simulated production to similar untreated counties. Because economic conditions and policies, moratoriums, and attitudes toward both fracking and migration varied by state, counties might not be compa-

[14]In Table 2.6 I re-estimate my IV estimates using actual new production rather than simulated production as the instrument and find similar results.

rable across states. To construct a counterfactual I will compare fracking counties to non-fracking counties in the same state as these counties are likely more similar along unobservable characteristics. In practice, I do this by including state by year fixed effects, which removes state specific shocks resulting in a within state and year comparison.

This comparison, however, does not account for cross-county spillovers that might arise from fracking. Previous work has suggested that the labor market impacts of fracking propagate beyond county borders, leading to large earnings and employment spillovers (Feyrer et al., 2017), which could bias these estimates. For this reason, I will also consider specifications that account for these potential spillovers. First, I adopt a method similar to Feyrer et al. (2017) by considering the total amount of simulated new production in the county and each of its neighbors. As such, production in neighboring counties can affect earnings and migration. I also estimate specifications which exclude non-fracking counties within 100 miles of counties with simulated new production.

2.4.3 Counterfactual Similarity

If fracking feasibility and new production is exogenously determined, we would expect fracking and non-fracking counties to be similar on average prior to fracking. In Table 2.1, I present county level descriptive statistics from 2000 (before fracking) for both non-fracking and fracking counties. Both groups are similar on average along most population dimensions, and especially so when comparing counties within the same state. Although fracking counties were slightly more white and less educated, the data suggest that simulated new production is not driven by initial county conditions.

I next explore changes over time to see if fracking and non-fracking counties followed similar trends prior to fracking. This also provides initial reduced form estimates of the impact of fracking on migration. To do this I calculate each county's total simulated new production between 2000 and 2013, and divide it by the within state mean total simulated new production among fracking counties. As such a one unit increase will represent the effect for the average fracking county in the state. This is done to better compare the average effects across states. I then interact this measure with a set of state and year indicators and regress in-migration rates between 2000 and 2013, on this set of interactions along with a set of county fixed effects and state by year fixed effects, omitting 2003, just prior to the start of the fracking boom, as the reference year. This allows me to trace out changes in migration in the average fracking county, relative to untreated counties in the same state. To show these trends, I plot the percentage point difference for in-migration rates in Figure 2.3. The in-migration rate is calculated as the number of in-migrants as a percent of the county's baseline population in 2000. A one percentage point increase in the in-migration rate means that an additional one percent of the baseline population moved into the county. The vertical gray bars in 2004 and 2008 indicate the early transition years of fracking.

Before 2003, the differences between fracking and non-fracking in the same state are flat and insignificant, suggesting counties that would later be affected by fracking were not on different in-migration trends. After 2003, there is a massive increase in in-migration in North Dakota; between 2010 and 2013, a flood of migrants, equivalent to nearly 23 percent of the baseline population, entered the average fracking county in North Dakota. There is also small but significant migration in a few other states, but in-migration is never more than 1.1 percent of the baseline population. This geographic disparity might reflect heterogeneous treatments (labor demand shocks) or heterogeneous responses (differences in propensities to move).

2.5 Estimation and Results

2.5.1 Reduced Form Impact of Fracking on Labor Markets

The previous figure could simply reflect differences across counties in production intensity, not necessarily heterogeneous migration behavior. I next regress earnings and migration on simulated new production to estimate the marginal effect of production. This relies on a less restrictive identifying assumption, as I will now be comparing different levels of production within a given state. To show that there are potential migration incentives, I first estimate the reduced form impact of simulated production on various labor market measures as follows

$$Y_{ct-1} = \alpha_c + \beta_1 Sim.NewProd._{ct-1} + \phi_{st} + \varepsilon_{ct} \tag{2.5}$$

where the dependent variable, Y_{ct-1}, is the labor market measure in logs and *Sim. New Prod.* is the simulate production from new wells in tens of millions of dollars. It seems likely that individuals would observe earnings or employment in $t-1$ when making migration decisions in period t. When looking at migration responses I will look at the impact of lagged production, or lagged earnings on current migration. As such, I lag both the outcome and simulated new production in this specification, to correspond to that first stage relationship.[15]

Because migration data is not separated by demographic characteristics, I estimate the impacts on average labor market measures. Although fracking does require some workers with advanced training (such as petroleum engineers), the tasks associated with most fracking jobs are manual in nature (e.g., hauling pipe, operating heavy machinery, driving) and the few technical tasks, such as monitoring equipment, do not required advanced degrees. I also examine impacts by gender and education in the appendix, as these groups might be affected differently by the shock. I include a county fixed effect, to account for time-invariant characteristics that affect labor markets, as well as state-by-year fixed effects to account for state-specific shocks and compare counties in the same state. The idiosyncratic ε_{ct} component might be correlated within a county over time, so I adjust the standard errors to account for clustering at the county level.[16] In all of my estimation, I only include

[15]The relationship does not qualitatively change when using contemporaneous production and earnings.

[16]Standard errors are similar if I correct for clustering at the commuting zone. However, because there are few commuting zones in North Dakota the standard errors for North Dakota estimates are slightly smaller when clustering at this level. I have also estimated Conley (1999) standard errors that account for correlations across different combinations of space and time. These standard errors are smaller, so I report the more conservative standard errors that account for clustering at the county level.

states that have any shale play and restrict my sample to counties with over 1,000 people in 2000, to limit the influence of very small counties.[17]

The reduced form impact of simulated production on earnings is reported in Table 2.2. For reference, the average simulated production from new wells in 2013 was $13 million (2010$). I estimate that for an additional ten million dollars of production, average earnings increased by one percent. In 2013, the average county with simulated new fracking production saw a 1.3 percent increase in earnings from fracking. However, the distribution of simulated production is heavily skewed; among counties with over 10 million dollars of production, average earnings increased by 6.6 percent, while among the top 50 counties the increase was 13.2 percent. Earnings outside of oil and gas extraction also increase, suggesting the shock to labor demand in oil and gas extraction had a ripple effect on other industries (Feyrer et al., 2017). Next, I follow the method of Ganong and Shoag (2017), and subtract five percent of the average house price from average earnings to construct a measure of consumption earnings that adjusts for the cost of living (Blanchard & Katz, 1992). This measure of "real earnings" also significantly increased, suggesting that there are potential net benefits to moving. An additional ten million dollars of production also increased the county jobs to population ratio by one percent, suggesting there were more employment opportunities in addition to higher earnings. The final column of Table 2.2 combines the effects on earnings and employment and looks at average earnings per capita. Ten million dollars of production increased average earnings

[17] I also exclude Broomfield County CO which was created during the sample period, Pitkin County CO for missing housing data, and to remove outliers I trim the data to exclude counties with over $1 billion of simulated production in a year, which excludes the county with the highest simulated production, Webb County TX.

per capita by two percent. In Appendix Table A2.1 we see that men without a college degree saw the largest labor market improvements.

I next explore differential labor market impacts across geography. To do this I interact my measure of simulated production with indicator variables for each of the four Census regions. Because the reduced form migration behavior in North Dakota is so different, I include North Dakota as a separate fifth group and will test for differences across regions. I then estimate

$$Y_{ct-1} = \alpha_c + \sum_{r}^{R} \beta_r Sim.\ New\ Prod._{ct-1} * I\{region_c = r\} + \phi_{st} + \varepsilon_{ct} \tag{2.6}$$

where r equals *North Dakota*, *West*, *South*, *Northeast*, or *Midwest*. Through 2013, very little fracking had occurred in the Midwest outside of North Dakota, I include this region for completeness, although it often lacks variation to identify meaningful relationships. By excluding the direct effect of simulated new production and looking within state, β_r will be the marginal effect of simulated new production in that region. These results are also reported in Table 2.2.

The labor market impacts vary considerably across regions, with ten million dollars of simulated new production increasing average earnings by 2.5 percent in North Dakota, 0.9 percent in the West, 0.4 percent in the South, and 10.3 percent in the Northeast, and an insignificant 10.3 percent in the Midwest. Across all measures the marginal impact of production is largest in the Northeast, with large effects in North Dakota, smaller effects in the West and South, and insignificant impacts in the Midwest. These short run labor market improvements suggest net benefits to moving and migration incentives might exist.

2.5.2 Reduced Form Impact of Fracking on Migration

I next explore the reduced form impacts of simulated new production on migration. I re-estimate equations (2.5) and (2.6) where the outcome of interest is the migration rate (not lagged). Because the decisions to move in and move out are affected differently by fracking, I will separately look at net migration (to capture total population growth due to migration), in-migration, and out-migration. I measure migration as the number of migrants in the county, scaled by the baseline county population in 2000, and multiplied by 100, to reflect the percent of the baseline population that each migration flow represents. Defined this way, a one percentage point increase in the net migration rate implies the population grew by one percent, while a one percentage point increase in the in-migration rate would mean that an additional flow of migrants, equal to one percent of the initial population, arrived in the county.[18]

Migration impacts are reported in Table 2.3. On average, the population grew in response to the labor demand shocks associated with fracking. An additional 10 million dollars of simulated new production increased the baseline population by 0.11 percent. However, there is stark regional heterogeneity, significant population growth only occurred in fracking counties in North Dakota and the Northeast. An additional 10 million dollars of simulated production increased the baseline popu-

[18]The number of migrants could also be measured in logs, so that $beta_r$ would approximate the percent change relative to baseline migration in region r. This is difficult to compare across regions as the scale will depend on initial migration levels. In Table 2.8 I show that regional differences are robust to differences in initial population.

lation by 0.42 percent in North Dakota and 0.29 percent in the Northeast, with an insignificant 0.05 percent increase in the West and negative point estimates in the South and Midwest. Although the marginal impacts in North Dakota and the Northeast are not statistically different, the total impacts are vastly different. Between 2000 and 2013, the average fracking county in North Dakota had over 290 million dollars of simulated new production, suggesting that the baseline population grew by over 12 percent on average. The implied total population growth from fracking in the most productive counties in North Dakota was nearly 25 percent. In contrast, the implied average county population growth from fracking in the Northeast was only 0.26 percent as new production was substantially lower during this period. Even among the most productive counties in the Northeast the implied impact would only be around one percent.[19]

An additional ten million dollars of simulated new production increased the number of in-migrants (as a percent of the 2000 population) by 0.95 percentage points in North Dakota, 0.21 percentage points in the West, 0.06 percentage points in South states, 0.48 percentage points in the Northeast, with an imprecise 0.38 percentage point increase in the Midwest. This would suggest that during this period an additional 28 percent of the baseline population moved into the average fracking county in North Dakota, whereas the inflow in fracking counties in other states increased by less than four percent. Perhaps surprisingly, simulated new production also led to higher rates of out-migration. This is not a prediction that would arise

[19]The implied average county population growth would be an insignificant 1.1 percent in the West and -0.07 percent in the South.

from the static migration choice model, unless fracking induced certain individuals to systematically sort away from fracking. However, as many migration decisions are eventually reversed by a second move, or return migration (Kennan & Walker, 2011), higher outflows could also arise if migrants only stay for a short period of time (long enough to file taxes). Understanding the role of these two channels also has implications for future population and labor market dynamics. On the one hand, certain groups systematically sort away from fracking (such as the wealthy, more educated, or politically progressive) might have real effects on local governance and public good provision. On the other hand, short-term migration might propagate the labor demand shock (as the stock of workers does not increase), require firms to spend more resources finding new workers, or result in more of the gains from fracking moving out of the local labor market.

To better understand if fracking led to sorting or short-term migration, I next turn to the 2005-2011 ACS microdata. These data help characterize the types of people that move to or away from fracking areas. Unfortunately, the ACS only provides migration information at the state and MIGPUMA level. In many of the rural areas involved in fracking, a MIGPUMA will cover multiple counties. As such, I simply construct an indicator for whether or not the MIGPUMA contains a county with any simulated new production. I restrict my sample to adults (25+); collapse the data to unique cells defined by migration status and destination, original location, year, and a set of demographic characteristics X_i; and then run the following regression at the cell (j) level

$$Y_j = \alpha_{s_{-1}} + X_j' \Gamma + \phi_t + \varepsilon_j. \tag{2.7}$$

Where X_j is a set of cell specific demographic characteristics including indicators for gender, marital status, gender by marital status, race, age bins, and educational attainment. I also include year fixed effects (ϕ_t), to account for year specific shocks, and fixed effects for the state (or country) of residence in the previous year ($\alpha_{s_{-1}}$), to remove time invariant differences across geography in individuals' initial circumstances. In this regression the coefficients in the vector Γ indicate how likely individuals with certain demographic characteristics were to migration. Cells are weighted by the summed individual weights provided by the ACS to be population representative. These demographic results are provided in Table 2.4.

I first look at the outcome of moving to a fracking region. In column (1), I include the full sample, to understand how migrants to fracking areas are different from the population as a whole. I multiply the binary outcome by 100 to scale the coefficients to represent percentage point changes. Unmarried individuals were over 50 percent (1.18/2.256) more likely to move, men were 11-19 percent more likely to move than women, and the migration response was almost entirely driven by 25 to 44 year olds.[20] High school dropouts were also the education group most likely to move to fracking, which is surprising given the general result that migration increases with education. Overall these characteristics match the predictions of the model as young and unmarried individuals face potentially lower costs on average and men and the

[20] Marriage decisions could potentially adjust to fracking, although this does not seem to be the case (Kearney & Wilson, 2017).

less educated faced the largest earnings gains. I next restrict the sample to migrants in column (2), to see how people moving to fracking are different from other migrants in general. Migrants to fracking are selected differently than other migrants and are more likely to be male, unmarried, and high school dropouts, and less likely to be 65 or older or black. In column (3) I look only at individuals who moved to fracking and regress this on the binary outcome of moving to fracking in the Bakken Play (in North Dakota), to see if these migrants were selected differently. Along most dimensions, the people that moved to North Dakota were similar to other people moving to fracking, although they were more likely to be non-Hispanic white and less likely to have a college degree.

I next look at moving away from fracking over the same samples to examine sorting. The same demographics that characterized individuals moving to fracking, also characterize those moving away from fracking. The inflows and outflows were composed of the same types of people, which would be consistent with short term migration rather than sorting along observable characteristics.[21] Such prevalent short term migration would suggest that monetary costs associated with moving (such as renting a truck) do not create binding constraints for many individuals. This phenomenon of short term migration to positive labor demand shocks has only started to be examined in the literature (Monte et al., 2016), and warrants further exploration in the future.

[21]As further evidence of short term migration, if I regress county level inflows from fracking counties on lagged outflows to those same counties, the coefficient is positive and significant and becomes larger when simulated production at the fracking destination is higher, suggesting that return migration increased with fracking production.

2.5.3 IV Estimated Impact of Earnings on In-Migration

To understand peoples' decisions to move to labor market improvements we must relate a measure of labor market strength (such as earnings) and migration. To do this I am interested in estimating an equation similar to

$$In\ migration\ rate_{ct} = \alpha_c + \gamma_1 ln(Ave.\ Earnings_{ct-1}) + \phi_{st} + \varepsilon_{ct} \tag{2.8}$$

where average earnings is a proxy for labor market opportunity and captures the earnings potential associated with moving. OLS estimation of equation (2.8) will be biased if cross-sectional variation in average earnings is correlated with unobserved county characteristics that affect migration decisions. To estimate this relationship I will use lagged simulated new production to instrument for lagged log average earnings as described by the following first and second stage equations

$$\begin{aligned} ln(Ave.\, Earnings_{ct-1}) &= \alpha_c + \beta_1 Sim.\, New\, Prod._{ct-1} + \phi_{st} + \varepsilon_{ct} \\ In\, migration\, rate_{ct} &= \alpha_c + \gamma_1 ln Ave.\, \widehat{Earnings}_{ct-1}) + \phi_{st} + \nu_{ct} \end{aligned} \tag{2.9}$$

Simulated new production is highly predictive of average earnings, with an F-statistic over 29 (see Table 2.2). To identify a causal relationship between in-migration and labor market strength, I must assume that simulated new production only affects the number of in-migrants through its effect on local labor markets, as proxied by average earnings. This assumption might seem strong, as other markets might adjust to fracking and enter migration decisions as well.

In particular, if the economic shocks generated by fracking are interpreted in a Rosen (1974) and Roback (1982) spatial equilibrium framework, then one would expect prices in the housing market to eventually endogenously respond to fracking and migration. The extent to which housing markets have adjusted across regions in the short run is an empirical question. As seen in Appendix Table A2.2, ten million dollars of new production leads to a significant 0.4 percent increase in the housing price in the North Dakota and a 3 percent increase in the Northeast.[22] Given this response, I must consider the possibility that housing prices also enter the migration decision in the short run and violate the exclusion restriction. To understand the role of housing prices, I will estimate the migration relationship under the baseline assumption, that housing markets do not affect migration, then use two separate approaches to account for changing housing markets.

Ideally, I would like to instrument for housing prices. However, as also seen in Appendix Table A2.2, many of the measures that could be used to identify exogenous variation in housing supply or price (e.g., housing market slackness in the pre-period, geographic elasticity constraints, the share of well water dependent households) are only weakly related, and often go in an unexpected direction. This weak relationship is not entirely unexpected as many fracking areas are rural and sprawling with elastic housing supplies. Rather than use these weak instruments, I will first directly control for housing prices in the equation. It should be noted

[22]This measure is constructed from the Federal Housing Finance Agency housing price index and converted to real dollars as explained in the data appendix. Other local measures of housing markets and rental rates are available through American Fact Finder from the 2000 Census and the 5-year ACS. However, none of these are available for the entire sample period.

that in this specification, housing prices are potentially endogenous and should not be given a causal interpretation. Directly controlling for housing prices absorbs the variation in migration correlated with housing markets, and allows me to determine if average earnings has a separate effect. If the coefficients on earnings are insensitive to this control, then the instrumental variation is not driven by changes in the housing market as a result of increased production. My second method of addressing changes in the housing market uses the measure of consumption earnings reported in Table 2.2 to account for the cost of living. In both of these specification I am interested in seeing if the coefficient on log earnings is sensitive to controlling for housing prices, which would suggest the exclusion restriction is invalid.

Although housing markets seem like the most likely threat to validity, the complexity of the migration decision make it impossible to account for the universe of potential confounding factors. To some degree, other potential confounders, such as crime levels or pollution, will be capitalized into housing values, and accounted for. However, implicitly I must assume no other factor violates the exclusion restriction. In an attempt to mitigate any bias due to equilibrium adjustment responses to production or migration that occur in the long run, I only look at early years of production and restrict my analysis to the short run.[23] For robustness I also consider an even shorter period, and find similar patterns.

In practice, I estimate a variation of equation (2.9) by interacting both simulated new production and average earnings with the set of region indicators, to

[23]This also limits the effect of long run equilibrium adjustments in earnings. Because average earnings are lagged, they are not directly affected by current migration.

estimate the regions specific relationship between earnings and in-migration rates. These estimates are reported in Table 2.5. The baseline model estimates that a 10 percent increase in average earnings in North Dakota led to an inflow of migrants equal to 3.8 percent of the baseline population. Similar increases in earnings increased in-migration rates by 2.4 percent in the West, 1.6 percent in the South, 0.5 percent in the Northeast, with no impact in the Midwest. The impact in North Dakota is nearly twice as large as in all other regions and statistically different. When controlling for housing markets, the coefficients on log earnings are remarkably similar and the geographic differences persist, suggesting the variation captured by earnings is not driven by responses to housing prices. I also run specifications accounting for potential cross-county spillovers. In Column (4) I use the total simulated new production in each county and its adjacent neighbors as the instrument, to allow nearby production to affect earnings. In Column (5) I exclude non-fracking counties within 100 miles of the nearest fracking county. In both cases the estimated elasticities are similar, suggesting that cross-county spillovers affect earnings and migration in a similar way across regions. Even in the current context of low mobility, I find significant migration to positive labor market shocks, however, for similarly sized increase in earnings, the migration response is quite varied across regions, with a particularly large response in North Dakota. As seen in Table 2.6, the point estimates and regional disparity is robust to weighting by population, shortening the sample to 2011, using actual production, the play by year interactions from equation (2.4), or simulated new wells as the instrument, and measuring migration in terms of current population levels.

Grouping states by region might mask variation across states so I also allow the relationship to vary for each state, rather than by region (see Appendix Table A2.3). In this specification I also explore the relationship between in-migration and two other proxies for labor market opportunity: employment and per capita earnings. Across all five specifications only two of the 50 point estimates are larger (although not statistically different). For most states the effect is substantially and statistically smaller than the relationship in North Dakota. The states where the migration response is the most similar are Montana, Colorado, and Texas. Overall the data indicate a positive causal effect of earnings on migration, but the treatment effect varies by region, with the largest response in North Dakota. In the next section, I explore four potential explanations for this geographic disparity in an attempt to unpack individuals' migration decisions.

2.6 Explaining Geographic Heterogeneity

2.6.1 Commuting as a Response to Potential Earnings Gains

It is possible that workers in nearby counties could respond to potential earnings gains by commuting rather than moving to fracking areas. This might be a more relevant alternative in fracking counties that are surrounded by larger populations (e.g., in Pennsylvania or Texas), rather than in fracking counties in North Dakota that are far from existing populations. If people respond by commuting in other fracking states we might not observe migration, but we would see the number of long distance commuters and workers living in other counties rise in these areas.

To test this I use the Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES) provided by the Census to construct the distance between the home Census Block Group and the work Census Block Group population centroids for all jobs within a county (U.S. Census, 2015). I then count the number of jobs in each county that are held by a long distance commuter (¿50 miles to the home Census Block Group) or by workers living in other counties. In Table 2.7 I estimate the impact of log earnings on the number of long distance commuters and workers living in other counties, as a percent of the 2000 population, similar to my migration specifications. The number of long distance commuters and workers from other counties increase with earnings across all regions, but the response is by far the largest in North Dakota. This response is also larger than the migration response, suggesting that many more workers responded to earnings gains by commuting rather than moving. To see if the total response (migration plus commuting) to labor market gains is the same across regions I estimate the combined impact on the number of workers living in other counties plus the number of in-migrants in column (3). The impacts in North Dakota are two to eight times as large as elsewhere, suggesting that, although many workers across the country responded by commuting, both movers and commuters were more responsive to earnings gains in North Dakota than elsewhere.[24]

[24]Differences in state policies might make it harder or easier to relocate. Anecdotal evidence suggests many individuals moving to North Dakota lived in cars or trailers in grocery store parking lots, when this might not be legal in other states (NYT Davey, 2010). These restrictions might have created a barrier to migration, but should not affect commuting behavior. Because the commute response and total response was larger in North Dakota, state temporary residency policies do not seem to explain the difference either.

2.6.2 Differences in Initial Population and Labor Market Characteristics

One possibility is that there were not enough people in North Dakota to meet the large labor demand increase from fracking and people had to move (or be moved) to meet demand. This could be due to either a sparse population, or a tight labor market with no additional labor supply. However, in other parts of the country, there were similarly rural counties that experienced fracking. To test this hypothesis, I re-weight counties in the other regions to resemble the distribution (mean and variance) of several population characteristics in 2000 for North Dakota counties as presented in Table 2.7.[25] When re-weighting to resemble the baseline population of North Dakota counties, the elasticity estimates rise, suggesting that some of the regional disparity can in fact be explained by differences in the initial population. However, there is still a gap between North Dakota and the other regions that is significantly different for the Northeast and Midwest and has a p-value of 0.14 for the West and 0.17 for the South. The pattern is similar if I instead re-weight to resemble the 2000 population of men ages 16 and older, which might be more relevant. Re-weighting to resemble the employment to population ratio of men 16 or older is similar to the baseline results. In the final column I re-weight counties to resemble the 16 and older male population density. In this case the point estimates in the West, South, and Northeast all rise to 26-28, still 10 points less than the North Dakota estimates, but are imprecisely estimated for the South and Northeast. This imprecision is likely

[25] In most cases this results in overweighting rural counties with low populations.

because counties in the South and Northwest are smaller, and there is less common support across regions in population density. Nonetheless, among similarly rural counties, the point estimate in North Dakota is still 40 percent larger. Although initial population characteristics explain some of the regional gap, there still appear to be regional differences in responsiveness that are unexplained by initial population size or density.[26]

2.6.3 Non-linear Relationship between Earnings and Migration

Another alternative explanation for the heterogeneous migration estimates is that the relationship between earnings and migration is non-linear, perhaps due to the fixed costs of moving. If people face a fixed cost, they will only move if the increase in earnings is sufficiently large. Perhaps fracking counties in North Dakota experienced large enough earnings gains that justify moving, while other regions did not. Non-linearities could also arise if fracking counties in North Dakota uniformly experienced the largest earnings gains, leading individuals to choose North Dakota over an alternative potential destination in their choice set. To see if the regional difference is due to non-linearities, I compare fracking counties in North Dakota and other regions that experienced similar gains in earnings from fracking. To do this I estimate the first stage relationship between simulated production and average earnings in equation (2.6), and then predict the annual earnings gains associated with

[26] Interacting log average earnings with these initial population characteristics produces the same patterns. The elasticities only change slightly with the initial population, and cannot predict the impacts in North Dakota. I have estimated these re-weighted specifications using the number of migrants (in levels), and although the estimates are less precise, the point estimate for North Dakota is in general 30 to 40 percent larger, suggesting this is not solely a mechanical result due to differences in initial population size.

simulated new production. I then truncate my sample to county/year observations below the maximum of predicted earnings increases excluding North Dakota. This limits my sample to counties in North Dakota and elsewhere that experienced similar earnings increases. These predicted earnings increases are then plotted against residual in-migration rates (after removing county and state by year fixed effects) to see if the relationship varies by region among similarly treated counties (see Appendix Figure A2.1). For reference I also plot the OLS linear relationship between residual in-migration rates and predicted earnings increases for each region and report the coefficients. Even when restricting the sample to counties that experienced a similar labor market treatment the relationship in North Dakota is three times as large, and statistically different than elsewhere. Although fixed costs or choice sets with multiple potential destinations might produce non-linearities among the most productive fracking counties, the data suggest that even for similar earnings gains migrants were more likely to select North Dakota.

2.6.4 Geographic Heterogeneity in Information

A fourth potential driver of the heterogeneous migration response is geographic variation in the flow of information about localized fracking booms. Fracking in North Dakota has received national attention and an outsized amount of media coverage per capita.[27] In Figure 2.5 I plot the number of domestic newspapers articles from LexisNexis which reference both fracking and the states name, divided

[27] See for example, Edwin Dobbs National Geographic article (2013), Konigsbergs New Yorker article (2011), or Daveys NYT article (2010) http://ngm.nationalgeographic.com/2013/03/Bakken-shale-oil/dobb-text,http://www.newyorker.com/magazine/2011/04/25/kuwait-on-the-prairie, or http://www.nytimes.com/2010/04/21/us/21ndakota.html?pagewanted=all.

by the state population, to account for the fact that more populous states have more newspapers and to scale it similar to migration rates. Starting in 2011, North Dakota has been disproportionately represented, being mentioned over three times per resident more often than other states by 2013.

In the context of the migration choice model, information could affect individuals' expectations about local average earnings (μ_d), the cost of moving (c_{iod}), or even their idiosyncratic component of earnings (ε_{id}) if it is not perfectly observed by the individual. This can shift the individual's threshold, changing their propensity to move. Information can also adjust the individual's choice set. The simple model only allows for two alternatives: stay or move, when in reality individuals might face many alternative destinations. The high level of information about North Dakota might induce people to add it to their choice set, while the large labor market gains experienced in other states such as New Mexico or West Virginia are not as publicized, so these states might not be considered. Information could also help explain the differential commuting response. If the labor market gains from fracking in nearby areas remain unknown, the commute response will be attenuated because individuals are not aware of the potential gains.

To see how information relates to migration, I construct an annual measure of newspaper publications that cite both fracking and a state name, by state of publication. Using the IRS county to county flows, I identify the migration inflow from each state to each county. In column (1) of Table 2.9 the data suggest that an addition billion dollars of simulated production increased these state-specific inflows by 0.12 percentage points. I next interact the state by state specific measure of newspapers with simulated production, to see if counties that received more publicity or information exposure, experienced more migration from the places this information was disseminated. The direct effect of news articles is small (0.04 percentage points for 100 news articles) but highly significant, suggesting that even when controlling for the shock (simulated production) newspapers publicity is correlated with migration. The interaction between production and articles is a significant 0.02 percentage points, and the migration response to production is larger from areas that received more news coverage about that specific fracking state. Meanwhile, the direct effect of simulated production falls to half the size and is insignificant, suggesting a large portion of the response to production is correlated with news coverage. This relationship is significant, although smaller, when we exclude North Dakota or include a state of origin fixed effect to control for changing characteristics at the origin.

This measure of information is potentially endogenous to migration, as the media might report more about fracking in areas that have a higher propensity to move to fracking. These coefficients do not have a purely causal interpretation, but the data do suggest that places that get more information about the economic shocks from fracking in certain areas also send more people to those areas. In a companion paper, I exploit differences in national news content and pre-fracking readership to explore the causality of this relationship (Wilson, 2017). I exploit variation in national news coverage and pre-fracking newspaper circulation rates to mitigate concerns about endogenous news producer and consumer decisions, and find that increased exposure to news about potential labor market opportunities leads to more migration to the places being talked about.

2.7 Conclusion

Internal migration rates in the US are historically low (Malloy et al., 2011), and evidence from the trade liberalization and the Great Recession suggests that people have become less likely to move away from negatively affected areas (Cadena & Kovak, 2015; Foote et al., 2015). Using recent economic shocks associated with localized fracking booms, this paper documents a sizable migration response to positive labor market shocks and highlights substantial heterogeneity in the migration response across both demographic groups and regions of the country.

The reduced form analysis suggests that both in- and out-migration positively respond to fracking production. However, the magnitude of this response varies significantly across regions. The population increased by 12-25 percent between 2000 and 2013 in North Dakota fracking counties, but by less than two percent in fracking counties in the West, South, Northeast, and Midwest. The ACS microdata show that this in-migration response is driven largely by the groups that face the largest earnings gains and potentially lowest moving costs: the young, unmarried, males, high school dropouts and college graduates. Migrants to fracking counties are also more likely to be high school dropouts than movers more generally, which contrasts with the general result that less educated workers are less likely to move. I also find that the same types of people move away from fracking, which suggests that fracking has led to high levels of short term migration and churn, but not necessarily selective sorting away from fracking. This has important implications for the labor market dynamics in these regions.

This paper also documents geographic heterogeneity in migration elasticities. The data imply that a 10 percent increase in average earnings was associated with an additional 3.8 percent of the baseline population moving into North Dakota, as compared to only 2.4 percent in the West, 1.6 percent in the South, and 0.5 percent in the Northeast. Previous work looking at negative shocks from the Great Recession find estimates comparable to the response in the West and South. This geographic disparity in in-migration is significant and robust to changes in the housing market, geographic spillovers, and a range of other specifications. Only a small part of this gap can be explained by commuting behavior or differences in initial population characteristics, or non-linear effects of earnings on migration, suggesting that potential migrants might view North Dakota differently than other areas.

The last alternative I propose is the potential role of information. Information can change individual expectations and migration choice sets. In particular, fracking in North Dakota has received a tremendous amount of news coverage. People move more to the fracking counties they get information about, suggesting non-market factors, such as information might influence migration decisions in addition to the traditional market factors, like earnings. Understanding the role of information could help understand differences across demographics and geography as well as explain potential mismatch and provide important policy implications. Further work is needed to understand why people do or do not move to better economic opportunities, and if policy measures can be taken to address potential market failures and increase social welfare.

2.8 Tables

Table 2.1: Pre-fracking 2000 County Population and Labor Market Summary Statistics

	Mean Values		Within State
	Non-Fracking Counties	Fracking Counties	Differences
	(1)	(2)	(3)
Total Population	80,972	102,189	14,232
Percent Male	49.45	49.50	-0.09
Percent White	82.90	87.32	2.54***
Percent Less than College (18+)	84.16	85.09	1.26***
Median Age	37.10	37.33	0.04
Percent Under 20	28.53	28.39	-0.02
Percent 20-34	18.36	18.17	-0.01
Percent 35-64	38.47	38.91	0.15
Percent 65 and older	14.63	14.52	-0.11
Male Average Earnings (2010$)	40,307	42,444	429
Male Employment Probability	0.55	0.55	-0.01
Female Average Earnings (2010$)	24,359	24,976	-123.17
Female Employment Probability	0.56	0.54	-0.01
Number of Counties	1587	742	-

Notes: County characteristics measured in 2000, prior to fracking and obtained from the 2000 Census and QWI. Sample restricted to counties in states over shale plays. Monetary values reported in dollars deflated to 2010 values using the personal consumption expenditures price index. Columns (1) and (2) report mean values, while column (3) report within state differences between non-fracking and fracking counties. Stars indicate values statistically different from zero. p¡0.01 ***, p¡0.05 **, p¡0.1 *.

Table 2.2: Reduced Form Impact of Simulated Production on Local Labor Market Measures

	County Labor Market Measure in t-1				
	Log Average Earnings	Log Average Non-O&G Earnings	Log Earnings Adjusted for Housing Price	Log Jobs to Pop. Ratio	Log Average Earnings per capita
	(1)	(2)	(3)	(4)	(5)
Sim. New Prod. Value in Cty_{t-1}	0.010***	0.006***	0.011***	0.010***	0.020***
(10 Millions 2010$)	(0.002)	(0.001)	(0.002)	(0.003)	(0.005)
	Regional Heterogeneity				
Sim. New Prod. Value in Cty_{t-1}	0.025***	0.016***	0.027***	0.029***	0.054***
*(10 Millions 2010$)*North Dakota*	(0.003)	(0.001)	(0.003)	(0.004)	(0.006)
Sim. New Prod. Value in Cty_{t-1}	0.009***	0.005***	0.010***	0.006***	0.015***
*(10 Millions 2010$)*West*	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Sim. New Prod. Value in Cty_{t-1}	0.004**	0.002	0.004**	0.003	0.007*
*(10 Millions 2010$)*South*	(0.002)	(0.001)	(0.002)	(0.003)	(0.004)
Sim. New Prod. Value in Cty_{t-1}	0.103***	0.068***	0.105***	0.101***	0.205***
(10 Millions 2010$) Northeast*	(0.015)	(0.022)	(0.018)	(0.024)	(0.035)
Sim. New Prod. Value in Cty_{t-1}	0.103	0.051	0.083	0.046	0.15
(10 Millions 2010$) Midwest*	(0.081)	(0.056)	(0.081)	(0.099)	(0.15)
F-statistic	29.42	33.44	27.78	18.01	25.47
Dependent Mean	34,247	33,848	28,450	0.538	19,208
Observations	31,157	31,157	31,155	31,143	31,143

Notes: Earnings data from QWI and simulated production constructed from DrillingInfo. Each column in each panel is a separate regression. Observation at the county by year level from 2000-2013. Average earnings are annual job level earnings and exclude the non-employed. Non-O&G excludes earnings from oil and gas extraction. Average earnings per capita divides total earnings by the working age population to account for non-employment. All regressions include county and state by year fixed effects, making this a comparison between counties in the same state. Standard errors are corrected for clustering at the county level. p<0.01 ***, p<0.05 **, p<0.1 *.

Table 2.3: Reduced Form Impact of Simulated Production on Internal Migration

	Number of Migrants, as Percent of 2000 Population		
	Net-Migrants	In-Migrants	Out-Migrants
	(1)	(2)	(3)
Sim. New Prod. Value in Cty_{t-1}	0.107**	0.300***	0.193***
(10 Millions 2010$)	(0.048)	(0.087)	(0.044)
	Regional Heterogeneity		
Sim. New Prod. Value in Cty_{t-1}	0.418***	0.952***	0.534***
*(10 Millions 2010$)*North Dakota*	(0.080)	(0.057)	(0.047)
Sim. New Prod. Value in Cty_{t-1}	0.054	0.207***	0.153***
*(10 Millions 2010$)*West*	(0.038)	(0.053)	(0.035)
Sim. New Prod. Value in Cty_{t-1}	-0.002	0.062***	0.064***
*(10 Millions 2010$)*South*	(0.013)	(0.014)	(0.012)
Sim. New Prod. Value in Cty_{t-1}	0.290**	0.483***	0.193
(10 Millions 2010$) Northeast*	(0.146)	(0.125)	(0.122)
Sim. New Prod. Value in Cty_{t-1}	-0.098	0.377	0.474
(10 Millions 2010$) Midwest*	(0.510)	(0.640)	(0.564)
Dependent Mean	0.0779	5.167	5.089
P-value North Dakota equals West	<0.01	<0.01	<0.01
P-value North Dakota equals South	<0.01	<0.01	<0.01
P-value North Dakota equals Northeast	0.44	<0.01	<0.01
P-value North Dakota equals Midwest	0.32	0.37	0.91
Observations	31,157	31,157	31,157

Notes: Migration data from IRS SOI, and simulated production constructed from DrillingInfo. Analysis at the county by year level. In the bottom panel, simulated production is interacted with a binary indicator for each of the five regions: North Dakota, West, South, Northeast, and the Midwest. The impact across regions are estimated jointly, and p-values testing for differential impacts between North Dakota and the other regions are reported. All regressions include county and state by year fixed effects, which make this a comparison between counties in the same state. Standard errors are corrected for clustering at the county level. p<0.01 ***, p<0.05 **, p<0.1 *.

Table 2.4: Characteristics of People who Move to and away from Regions Involved in Fracking

	To Fracking Regions			Away from Fracking Regions		
	Move to Fracking*100		Move to Bakken*100	Move to Fracking*100		Move to Bakken*100
Sample	Full Adult Pop. (1)	All Migrants (2)	Migrants to Fracking (3)	Full Adult Pop. (4)	All Migrants (5)	Migrants to Fracking (6)
Male	0.25***	0.36***	-0.02	0.11***	0.18***	0.003
	(0.05)	(0.12)	(0.02)	(0.03)	(0.06)	(0.01)
Unmarried	1.18***	1.66**	-0.09**	0.37***	-0.97***	0.002
	(0.24)	(0.72)	(0.04)	(0.10)	(0.36)	(0.01)
*Male*Unmarried*	0.18***	-0.17	0.01	0.19***	0.98***	-0.004
	(0.07)	(0.29)	(0.04)	(0.05)	(0.26)	(0.01)
34 and Under	2.66***	0.36	0.00	1.04***	-0.06	-0.04
	(0.50)	(0.44)	(0.03)	(0.28)	(0.16)	(0.03)
Age 35-44	0.90***	0.46**	-0.01	0.33***	-0.17	-0.03
	(0.19)	(0.21)	(0.04)	(0.09)	(0.17)	(0.03)
65 and Over	-0.55***	-1.23***	-0.03	-0.18***	0.26	-0.04
	(0.13)	(0.42)	(0.05)	(0.05)	(0.19)	(0.03)
Black-NH	0.11	-4.57***	-0.10***	-0.24***	-4.81***	0.01
	(0.30)	(1.33)	(0.03)	(0.05)	(1.31)	(0.01)
Hispanic	-0.16	0.78	-0.16***	-0.59*	-4.54	0.02
	(0.48)	(3.62)	(0.04)	(0.34)	(2.81)	(0.01)
Other-NH	0.09	-0.05	-0.00	-0.02	-1.55***	0.10
	(0.12)	(1.87)	(0.07)	(0.07)	(0.40)	(0.09)
Less than HS	0.28***	1.21**	-0.06	0.08**	-0.15	-0.0001
	(0.09)	(0.53)	(0.04)	(0.03)	(0.26)	(0.002)
Some College	0.07	-0.39	-0.06	0.03	-0.75***	-0.01
	(0.05)	(0.25)	(0.05)	(0.02)	(0.18)	(0.02)
College Degree	0.16***	-1.04	-0.15*	0.08*	-1.30**	-0.01
	(0.06)	(1.39)	(0.08)	(0.04)	(0.53)	(0.01)
Dependent Mean	2.258	31.04	0.280	0.807	11.10	0.109
Observations	427,593	330,362	93,799	427,593	330,362	93,799

Notes: Sample constructed from the 2005-2011 ACS microdata, and collapsed to unique cells by geography, migration status, and demographic characteristics as explained on page 23. Observations are then weighted by the summed population weights to be population representative. The dependent variable for moving to fracking and moving to the Bakken region are multiplied by 100 such that a coefficient of one represents a one percentage point increase. Only people who move across MIGPUMA boundaries are labeled as migrants. All regressions include fixed effects for the year and the state of residence in the previous year. Standard errors are corrected for clustering at the state of residence in the previous year level. p<0.01 ***, p<0.05 **, p<0.1 *.

Table 2.5: Impact of Average Earnings on the Number of In-migrants by Region, 2SLS

	Outcome: Number of In-migrants as a Percent of 2000 Population				
	Baseline	Adjustments in Housing Markets		Neighboring County Spillovers	
		Control for Housing Price	Adjust Earnings for Housing Price	Own + Neighbors' Prod. as Instrument	Exclude Neighbors <100 Miles
	(1)	(2)	(3)	(4)	(5)
Log Average Earnings$_{t-1}$	38.02***	40.35***	35.03***	36.47***	38.40***
**North Dakota*	(5.82)	(6.32)	(5.25)	(5.68)	(6.11)
Log Average Earnings$_{t-1}$	24.20***	24.53***	20.59***	25.55***	24.93***
**West*	(3.81)	(3.72)	(3.29)	(4.41)	(3.71)
Log Average Earnings$_{t-1}$	15.67**	16.15**	14.47**	12.53	13.77*
**South*	(7.14)	(7.60)	(6.63)	(9.79)	(8.27)
Log Average Earnings$_{t-1}$	4.71***	4.69***	4.60***	5.56***	5.09**
**Northeast*	(1.61)	(1.65)	(1.68)	(1.97)	(2.01)
Log Average Earnings$_{t-1}$	3.65	3.96	4.52	-1.17	9.52
**Midwest*	(7.04)	(7.49)	(9.26)	(1.77)	(23.24)
P-value North Dakota equals West	0.05	0.03	0.02	0.13	0.06
P-value North Dakota equals South	0.02	0.01	0.02	0.03	0.02
P-value North Dakota equals Northeast	<0.01	<0.01	<0.01	<0.01	<0.01
P-value North Dakota equals Midwest	<0.01	<0.01	<0.01	<0.01	0.23
Observations	31,157	31,157	31,155	31,157	16,854

Notes: Data compiled from the IRS SOI, QWI, Federal Housing Finance Agency (FHFA), and DrillingInfo. The impact across regions are estimated jointly to test for differences. The p-values provided are from the test of equality across the regions. Columns (2) and (3) account for potential changes in the housing market in response to fracking production. Column (2) directly controls for log median housing prices. In column (3) earnings are adjusted to account for differences in housing prices following the method of Ganong & Shoag (2015). Columns (4) and (5) account for potential spillovers into nearby counties. Column (4) includes simulated new production from bordering counties in the instrument, to capture potential changes in earnings in non-producing counties. Column (5) excludes non-producing counties within 100 miles of a fracking county. All regressions include county and state by year fixed effects, which make this a comparison between counties in the same state. Standard errors are corrected for clustering at the county level. p<0.01 ***, p<0.05 **, p<0.1 *.

Table 2.6: Robustness of Regional Migration Elasticities

	Outcome: Number of In-migrants as a Percent of 2000 Population						Outcome:
Specification:	Baseline	Weighted by 2000 Population	Shorter Sample (≤2011)	Actual Prod. as Instrument	Play by Year Interacts as Instruments	Sim. New Wells as Instrument	In-migrants as a Percent of Current Population
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Average Earnings$_{t-1}$	38.02***	37.14***	28.69***	40.81***	36.51***	35.45***	24.46***
**North Dakota*	(5.82)	(3.09)	(1.78)	(7.14)	(6.37)	(5.85)	(2.80)
Log Average Earnings$_{t-1}$	24.20***	20.19	20.14***	21.44***	0.74	19.05**	16.02***
**West*	(3.81)	(30.82)	(4.08)	(4.48)	(2.15)	(7.76)	(3.99)
Log Average Earnings$_{t-1}$	15.67**	8.77	17.17	10.83*	2.86	14.47**	13.38**
**South*	(7.14)	(14.00)	(15.35)	(6.40)	(1.88)	(7.24)	(6.20)
Log Average Earnings$_{t-1}$	4.71***	3.17*	17.01**	5.31***	5.03	4.33***	6.46***
**Northeast*	(1.61)	(1.63)	(7.59)	(1.78)	(3.62)	(1.58)	(1.78)
Log Average Earnings$_{t-1}$	3.65	-6.99	28.60	-5.89	3.62**	17.76	10.12
**Midwest*	(7.04)	(22.85)	(44.64)	(6.00)	(1.81)	(33.54)	(10.91)
P-values:							
North Dakota equals West	0.05	0.58	0.05	0.02	<0.01	0.09	0.08
North Dakota equals South	0.02	0.05	0.46	<0.01	<0.01	0.02	0.10
North Dakota equals Northeast	<0.01	<0.01	0.13	<0.01	<0.01	<0.01	<0.01
North Dakota equals Midwest	<0.01	0.06	0.99	<0.01	<0.01	0.60	0.20
Observations	31,157	31,157	26,533	31,157	31,157	31,157	31,143

Notes: Data compiled from the IRS SOI, QWI, and DrillingInfo. Each column is modified as specified. All regressions include county fixed effects. All regressions include state by year fixed effects, to control for time invariant county characteristics as well as state specific shocks, making this a comparison between counties in the same state. Standard errors are corrected for clustering at the county level. p<0.01 ***, p<0.05 **, p<0.1 *.

Table 2.7: Impact of Average Earnings on Long Distance Commuters and Out of County Workers

	Long Distance Commuters (>50 Miles)	Workers living in Other County	Workers living in Other County + In-Migrants
	As Percent of 2000 Population		
	(1)	(2)	(3)
*Log Average Earnings$_{t-1}$*North Dakota*	113.18***	134.91***	177.06***
	(19.71)	(21.03)	(28.00)
*Log Average Earnings$_{t-1}$*West*	53.65***	67.62***	99.48***
	(11.99)	(13.14)	(20.51)
*Log Average Earnings$_{t-1}$*South*	50.26***	73.34***	89.66***
	(17.93)	(26.41)	(30.48)
*Log Average Earnings$_{t-1}$*Northeast*	8.65*	12.73	17.88
	(5.13)	(11.22)	(11.35)
*Log Average Earnings$_{t-1}$*Midwest*	-6.08	-30.57	-26.27
	(10.17)	(39.36)	(36.47)
Dependent Mean (in Levels)	5.7	16.0	21.2
P-value North Dakota equals West	<0.01	<0.01	<0.01
P-value North Dakota equals South	<0.01	<0.01	<0.01
P-value North Dakota equals Northeast	<0.01	<0.01	<0.01
P-value North Dakota equals Midwest	<0.01	<0.01	<0.01
Observations	23,038	23,038	23,038

Notes: Data on long distance commuters and out of county workers come from the LEHD Origin-Destination Employment Statistics (LODES) and is combined with QWI and DrillingInfo data. Each column is a separate regression. In Column (1) the dependent variable is the number of jobs held by workers (as a percent of the 2000 population) where the distance between the home and work Census Block centroid is over 50 miles (regardless of county). In Column (2) the dependent variable is the number of jobs in the county held by workers living in a different county, as a percent of the 2000 population. In Column (3) I combine the number of jobs held by workers living in different counties with the number of in-migrants from the IRS SOI data to estimate the total mobility response by region. The p-values provided are from the test of equality across the regions. All regressions include county and state by year fixed effects, which make this a comparison between counties in the same state. Standard errors are corrected for clustering at the county level. p<0.01 ***, p<0.05 **, p<0.1 *.

Table 2.8: Role of Initial Characteristics: Re-weighting regions to Resemble North Dakota Counties

		Outcome: Number of In-migrants as a Percent of 2000 Population Re-weighting Characteristic in 2000			
	Baseline (1)	Total Population (2)	16+ Male Population (3)	16+ Male Emp/Pop Ratio (4)	16+ Male Population Density (5)
Log Average Earnings$_{t-1}$	38.02***	38.017***	38.017***	38.017***	38.017***
**North Dakota*	(5.82)	(5.822)	(5.822)	(5.822)	(5.822)
Log Average Earnings$_{t-1}$	24.20***	28.426***	28.444***	22.734***	26.792***
**West*	(3.81)	(2.748)	(2.766)	(4.761)	(2.983)
Log Average Earnings$_{t-1}$	15.67**	21.621**	20.856**	15.276**	26.748
**South*	(7.14)	(10.379)	(10.029)	(6.062)	(19.940)
Log Average Earnings$_{t-1}$	4.71***	9.375	10.194	5.069***	28.318
**Northeast*	(1.61)	(5.988)	(6.590)	(1.624)	(46.819)
Log Average Earnings$_{t-1}$	3.65	5.532	4.785	0.730	16.460
**Midwest*	(7.04)	(8.490)	(7.920)	(5.903)	(18.018)
P-value North Dakota equals West	0.05	0.14	0.14	0.04	0.09
P-value North Dakota equals South	0.02	0.17	0.14	<0.01	0.59
P-value North Dakota equals Northeast	<0.01	<0.01	<0.01	<0.01	0.84
P-value North Dakota equals Midwest	<0.01	<0.01	<0.01	<0.01	0.26
Observations	31,157	31,157	31,157	31,157	31,157

Notes: Data compiled from the IRS SOI, QWI, 2000 Census, and DrillingInfo. The impact across regions are estimated jointly to test for differences. The p-values provided are from the test of equality across the regions. Column (1) provides the baseline results from Table 5. Columns (2) through (5) re-weight counties in other regions to resemble the distribution of the specified population characteristic in 2000 among North Dakota counties. Weights are selected to match both the mean and variance. All regressions include county and state by year fixed effects, which make this a comparison between counties in the same state. Standard errors are corrected for clustering at the county level. p<0.01 ***, p<0.05 **, p<0.1 *.

Table 2.9: Potential Mediating Rule of Information

	Number of In-migrants from State of Publication as Percent of 2000 Population					
	Include North Dakota			Exclude North Dakota		
	(1)	(2)	(3)	(4)	(5)	(6)
Sim. New Prod. Value in Cty_{t-1}	0.120***	0.056	0.057	0.076***	0.035	0.036
(In Billions of 2010$)	(0.031)	(0.038)	(0.038)	(0.017)	(0.029)	(0.028)
Articles by state of $publication_{t-1}$		0.0004***	0.0004***		0.0003***	0.0004***
		(0.0001)	(0.0001)		(0.0001)	(0.0001)
Sim. New Prod. Value in Cty_{t-1} *		0.020**	0.020**		0.013*	0.013*
Articles by state of $publication_{t-1}$		(0.010)	(0.010)		(0.008)	(0.008)
State of Origin by Year Fixed Effects			X			X
Observations	815,388	815,388	815,388	778,974	778,974	778,974

Notes: Articles were collected from LexisNexis and combined with data from the IRS SOI and DrillingInfo. Observation at the county by year by state of origin level, and capture the annual county migration inflow from each state. “Articles” is the number of news articles that reference the fracking county’s state and were published in the state of origin. All regressions include origin state by destination county and state by year fixed effects, to control for time invariant pair specific characteristics as well as state specific shocks, making this a comparison between counties in the same state. In columns (3) and (5) state of origin by year fixed effects are also included to account for potential unobserved origin characteristics that are changing over time and affecting migration decisions. Standard errors are corrected for clustering at the county level. p<0.01 ***, p<0.05 **, p<0.1 *.

2.9 Figures

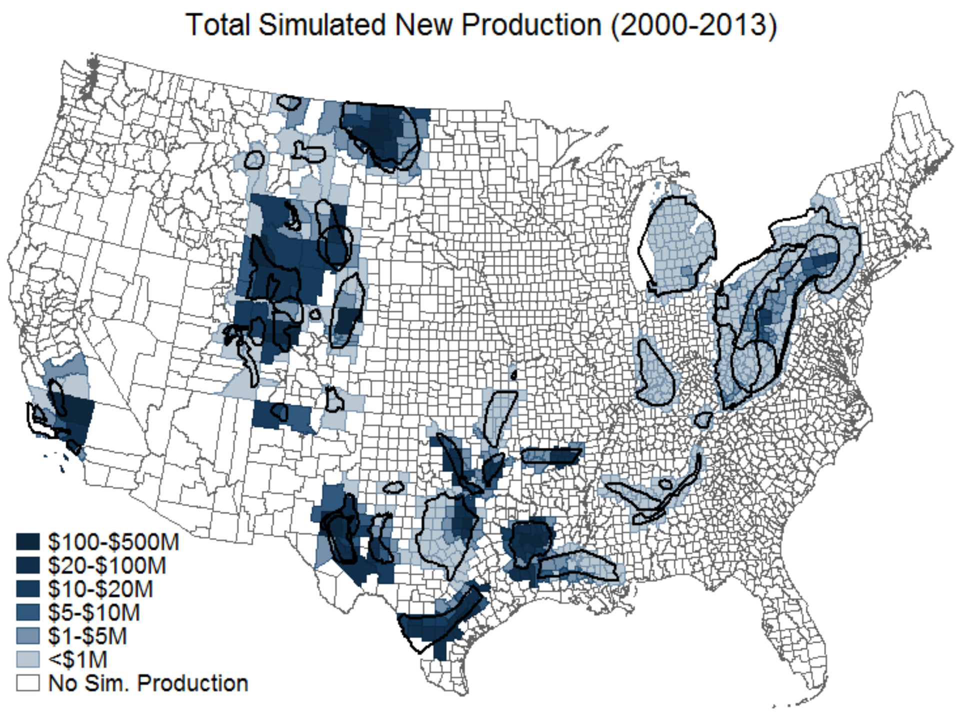

Figure 2.1: Geographic Variation in Fracking Feasibility and Simulated Production

Notes: Black outlines indicate the location of shale plays. Simulated new production estimates the production value from new wells in each county as a function geology and time (see equation (2.4)).

Source: Author's calculations from DrillingInfo well level reports. Shale play boundaries obtained from the Energy Information Administration.

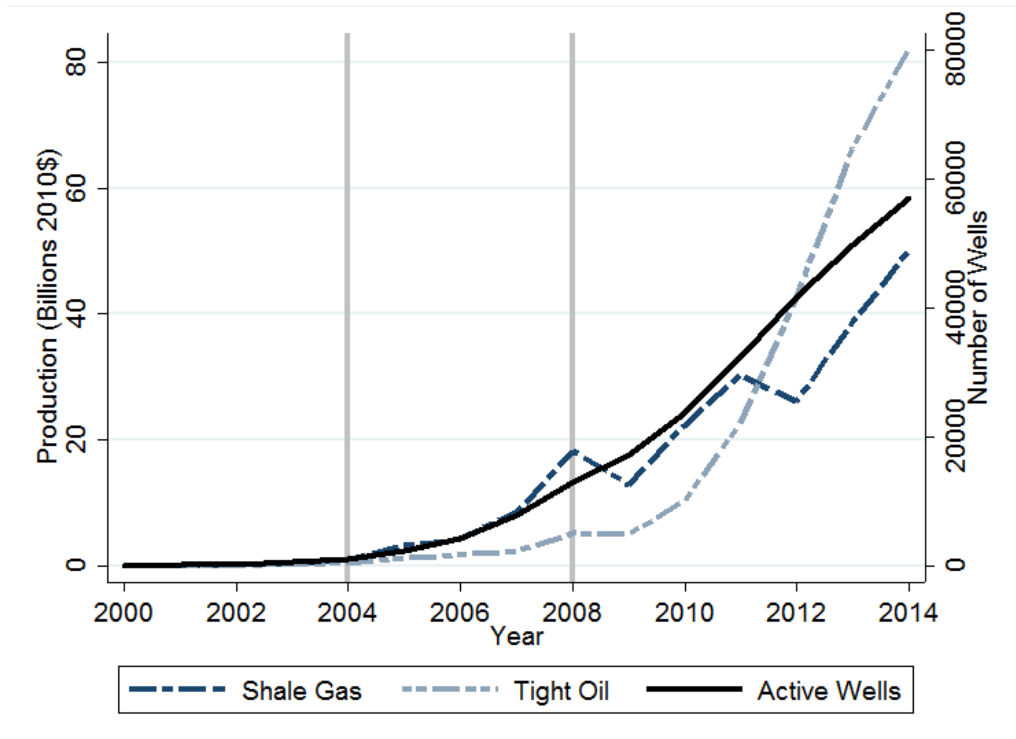

Figure 2.2: Oil and Gas Production from the Fracking Boom

Notes: The vertical, gray lines in 2004 and 2008 indicate the early transition years of the fracking boom. Oil and gas production is converted to 2010 dollar values using national oil and gas prices from the EIA.

Source: Author's calculation from DrillingInfo well level reports.

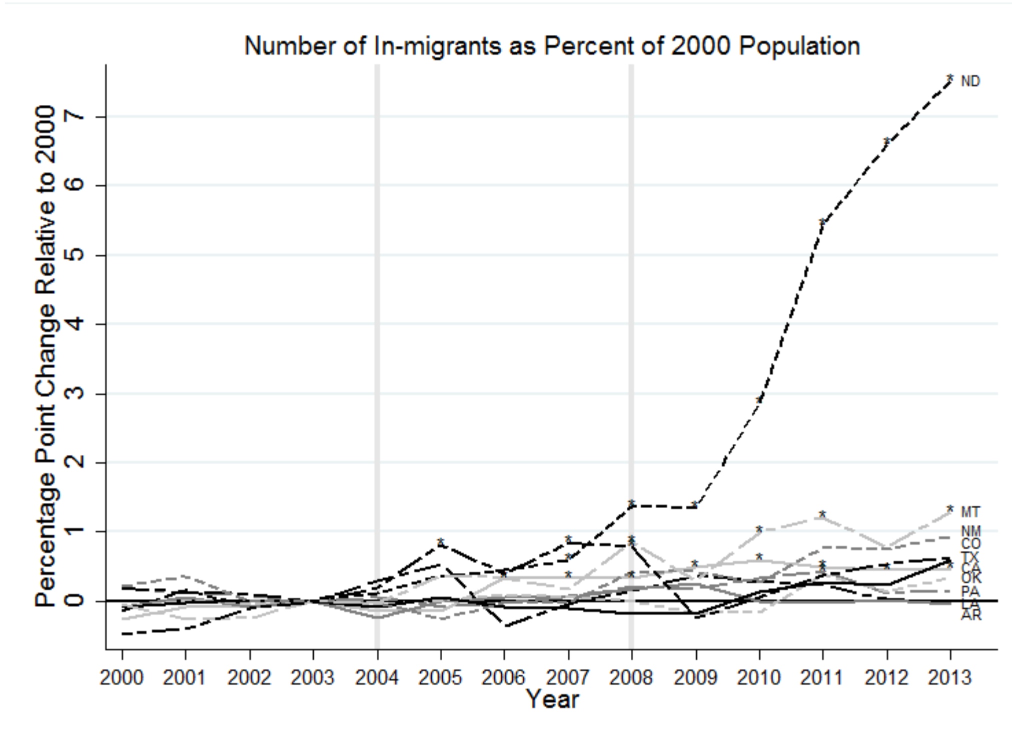

Figure 2.3: Trends in In-migration by State

Notes: Notes: The change in the in-migration rate for average total simulated new production in each state and year is plotted. Point estimates are obtained by regressing the in-migration rate on a set of interactions between total simulated new production between 2000 and 2013 with year indicators with county and state by year fixed effects. The indicator for the year 2003 is omitted as the reference year. Total simulated production is divided by the within state average among fracking counties, so that the estimated effects represent the average effect for fracking counties in that state. The vertical, gray line in 2004 and 2008 indicate the early transition years of the fracking boom. Asterisk indicates a statistically significant value at the 5 percent level.

Source: Author's calculation from DrillingInfo, QWI, and IRS SOI.

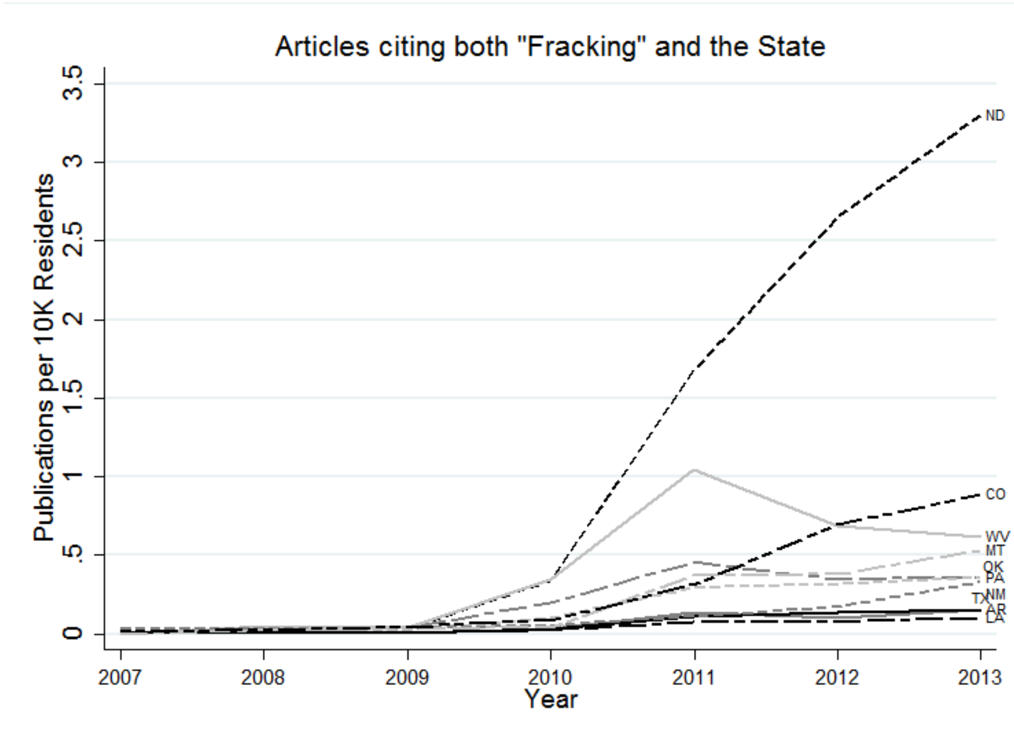

Figure 2.4: Per Capita Publication Count Mentioning Both "Fracking" and the State's Name

Source: Author's calculation from LexisNexis. Only US based publications are included. The number of publications is standardized by the state population to account for potential variation in the number of news outlets.

2.10 Appendix A. Additional Tables and Figures

Table A2.1: Reduced Form Impact of Simulated Production on Labor Market Measures by Gender and Education

	Log Average Earnings$_{t-1}$				Log Jobs to Pop. Ratio$_{t-1}$			
	Men		Women		Men		Women	
	No College Degree	College Degree	No College Degree	College Degree	No College Degree	College Degree	No College Degree	College Degree
	(1)	(2)	(3)	(4)	(5)			
Sim. New Prod. Value in Cty$_{t-1}$	0.010***	0.006***	0.005***	0.004***	0.020***	0.017***	0.003**	0.003
(10 Millions 2010$)	(0.002)	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.001)	(0.002)
Regional Heterogeneity								
Sim. New Prod. Value in Cty$_{t-1}$	0.023***	0.014***	0.013***	0.009***	0.050***	0.042***	0.010***	0.007**
*(10 Millions 2010$)*North Dakota*	(0.003)	(0.002)	(0.001)	(0.001)	(0.006)	(0.007)	(0.002)	(0.003)
Sim. New Prod. Value in Cty$_{t-1}$	0.010***	0.007***	0.005***	0.003***	0.011***	0.010***	0.004**	0.003
*(10 Millions 2010$)*West*	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.004)	(0.002)	(0.003)
Sim. New Prod. Value in Cty$_{t-1}$	0.005***	0.002*	0.001	0.002**	0.010**	0.010*	-0.001	0.002
*(10 Millions 2010$)*South*	(0.002)	(0.001)	(0.001)	(0.001)	(0.004)	(0.005)	(0.001)	(0.004)
Sim. New Prod. Value in Cty$_{t-1}$	0.147***	0.062***	0.036***	0.003	0.174***	0.130***	0.012	-0.005
*(10 Millions 2010$)*Northeast*	(0.018)	(0.022)	(0.013)	(0.018)	(0.035)	(0.027)	(0.032)	(0.033)
Sim. New Prod. Value in Cty$_{t-1}$	0.210**	0.196	-0.013	-0.135**	-0.093	0.125	0.067	0.160
*(10 Millions 2010$)*Midwest*	(0.106)	(0.131)	(0.045)	(0.060)	(0.123)	(0.261)	(0.108)	(0.188)
Dependent Mean	37,055	60,556	23,300	37,065	0.544	0.667	0.553	0.649
Observations	31,094	31,157	31,062	31,157	31,094	31,157	31,062	31,157

Notes: Data compiled from the QWI, ACS, and DrillingInfo. Each column in each panel is a separate regression. Observation at the county by year level. All regressions include county and state by year fixed effects, which make this a comparison between counties in the same state. Standard errors are corrected for clustering at the county level. p<0.01 ***, p<0.05 **, p<0.1 *.

Table A2.2: Reduced Form Effect of Simulated Production on Housing Prices

	Baseline (1)	Share Vacant in 2000 (2)	Characteristic Geography Constraint (3)	Share Own Water in 2000 (4)
Sim. Prod. Value in Cty_{t-1}	0.004***	0.002**	0.005***	0.003***
*(10 Millions 2010$)*North Dakota*	(0.001)	(0.001)	(0.001)	(0.001)
Sim. Prod. Value in Cty_{t-1}	0.001	0.002	0.002	0.002
*(10 Millions 2010$)*West*	(0.001)	(0.003)	(0.003)	(0.003)
Sim. Prod. Value in Cty_{t-1}	0.001	-0.005**	-0.001	0.001
*(10 Millions 2010$)*South*	(0.001)	(0.002)	(0.002)	(0.001)
Sim. Prod. Value in Cty_{t-1}	0.030**	0.027*	0.013	0.027*
*(10 Millions 2010$)*Northeast*	(0.015)	(0.015)	(0.017)	(0.015)
Sim. Prod. Value in Cty_{t-1}	0.036	-0.003	-0.033	0.033
*(10 Millions 2010$)*Midwest*	(0.065)	(0.072)	(0.084)	(0.066)
Sim. Prod. Value in Cty_{t-1}		0.007***	0.001	0.002**
**North Dakota*Characteristic*		(0.002)	(0.001)	(0.001)
Sim. Prod. Value in Cty_{t-1}		-0.004	0.001	-0.003
**West*Characteristic*		(0.005)	(0.001)	(0.004)
Sim. Prod. Value in Cty_{t-1}		0.066***	0.007	0.005
**South*Characteristic*		(0.021)	(0.006)	(0.010)
Sim. Prod. Value in Cty_{t-1}		0.127*	-0.090*	2.665**
**Northeast*Characteristic*		(0.074)	(0.050)	(1.277)
Sim. Prod. Value in Cty_{t-1}		0.546	-0.175	0.041*
**Midwest*Characteristic*		(0.397)	(0.110)	(0.023)
F-statistic	4.764	3.972	3.345	3.625
Observations	31,157	31,155	31,155	31,155

Notes: Housing price constructed from the housing price index provided by the Federal Housing Finance Agency and converted to dollars using county median house prices in 2000. Simulated production is interacted with a binary indicator for each of the five regions. The impact across regions are estimated jointly, to test for differences. In columns (2) through (4) region specific production is then interacted with various characteristics prior to the boom that could possibly affect pricing but otherwise be exogenous to migration. All regressions include county and state by year fixed effects, to control for time invariant county characteristics as well as state specific shocks, making this a comparison between counties in the same state. Standard errors are corrected for clustering at the county level. p<0.01 ***, p<0.05 **, p<0.1 *.

Table A2.3: State Specific Migration Responses to Earnings

	Outcome: Number of In-migrants as a Percent of 2000 Population				
Labor Market Measure	Average Earnings (1)	Average Earnings Controlling for Housing Price (2)	Housing Adjusted Earnings (3)	Jobs to Population Ratio (4)	Average Earnings per capita (5)
Log Measure$_{t-1}$	38.02***	40.35***	35.03***	32.96***	17.54***
	(5.82)	(6.32)	(5.25)	(5.40)	(2.69)
			Western States		
Log Measure$_{t-1}$*MT*	-8.53	-10.76	-9.11	7.95	-0.61
	(6.21)	(6.68)	(5.62)	(7.76)	(3.04)
Log Measure$_{t-1}$*NM*	-27.27***	-29.01***	-25.59***	-17.51***	-11.17***
	(5.97)	(6.47)	(5.38)	(6.68)	(2.86)
Log Measure$_{t-1}$*CO*	-7.73	-3.66	-11.65	-362.4	15.77
	(24.74)	(20.80)	(14.69)	(3,469)	(60.96)
Log Measure$_{t-1}$*CA*	-22.63	-25.13	-47.95***	-19.73*	-10.46
	(20.19)	(20.46)	(18.36)	(10.53)	(6.58)
			Southern States		
Log Measure$_{t-1}$*TX*	-17.96	-18.64	-16.30	-15.91	-8.41
	(12.88)	(14.67)	(11.97)	(16.33)	(7.02)
Log Measure$_{t-1}$*OK*	-31.91***	-33.88***	-29.38***	-26.52*	-14.43**
	(11.60)	(12.87)	(10.64)	(15.87)	(6.34)
Log Measure$_{t-1}$*AR*	-34.08***	-36.50***	-31.71***	-17.83	-14.42***
	(6.12)	(6.59)	(5.49)	(18.64)	(2.87)
Log Measure$_{t-1}$*LA*	-26.33	-28.70	-23.37	-43.99**	-1,047
	(25.62)	(27.85)	(27.79)	(21.96)	(143,500)
			Northeastern States		
Log Measure$_{t-1}$*PA*	-32.81***	-35.01***	-30.25***	-27.86***	-14.98***
	(6.06)	(6.57)	(5.50)	(5.74)	(2.83)
			Other States		
Log Measure$_{t-1}$*Other*	-12.53*	-14.58**	-16.60***	-3.73	-3.96
	(6.40)	(6.80)	(6.05)	(7.42)	(3.23)
Independent Mean	34,516	34,516	28,688	0.538	19,363
Observations	31,143	31,143	31,141	31,143	31,143

Notes: Data compiled from the IRS SOI, QWI, Federal Housing Finance Agency, and DrillingInfo. Each column is a separate regression. The direct effect of log average earnings represent the impact for North Dakota, and all interactions are deviations from this base. In column (2), I directly control for log housing prices. In column (3) earnings are adjusted to account for differences in housing prices following the method of Ganong & Shoag (2015). All regressions include county and state by year fixed effects, which make this a comparison between counties in the same state. Standard errors are corrected for clustering at the county level. p<0.01 ***, p<0.05 **, p<0.1 *.

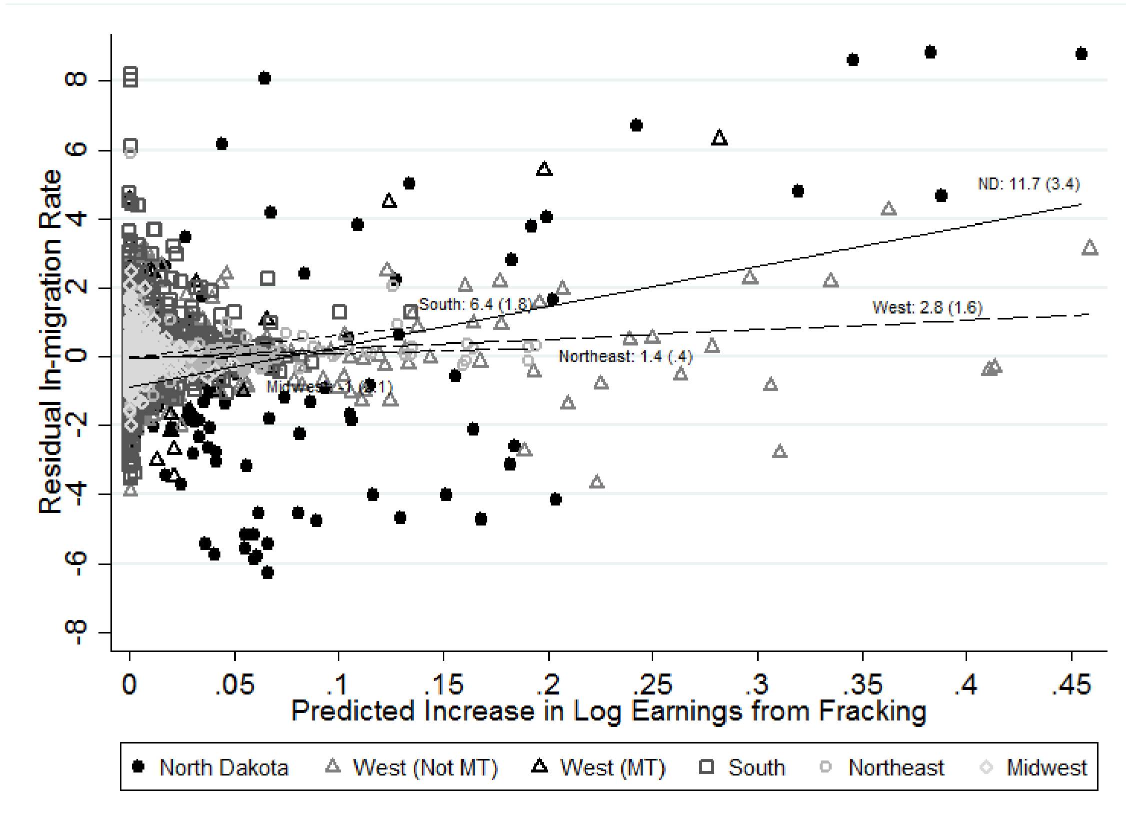

Figure A2.1: Elasticity Estimates by Regions for Counties that Experienced Similar Earnings Increases

Notes: Predicted increases in log average earnings from simulated new production are obtained from the first stage regression in equation (2.6) and then plotted along the x-axis. Residual in-migration rates that account for county and state by year fixed effects are plotted along the y-axis. The sample is then truncated from above at the highest predicted earnings increase *among fracking counties outside of North Dakota* so as to compare North Dakota fracking counties to fracking counties in other states that experienced similar earnings increases. OLS fits for each region are plotted with the estimated coefficient and standard error in parentheses. The OLS relationship in North Dakota is significantly larger than that in all other regions. As many fracking counties in Montana also lay over the Bakken shale play, observations from Montana are indicated with a black triangle. The largest predicted increases in log earnings outside of North Dakota are in the West, but not in Montana.

Source: Author's calculations using IRS SOI migration data and QWI earnings.

2.11 Appendix B. Data Appendix

Below I describe each of the key datasets used in my analysis, as well as important characteristics of data construction.

Internal Revenue Service Statistics of Income County Flows

The Internal Revenue Service (IRS) Statistics of Income (SOI) division provides annual counts of county-to-county flows. This provides the raw number of tax returns and exemptions that were filed in one county in year $t-1$ and in another county in year t. Each year, the IRS provides county-to-county flows of exemptions in a file with two years (e.g., 2002to2003). This represents exemptions that were in one county when filing in 2002 and in another county when filing in 2003. As most people file in the beginning of the year before April, I assign this flow to the year 2002.

Using exemptions to approximate people in a household, I collapse each county, year to a single observation of the total number of exemptions.[28] The in-migration rate can be constructed by dividing the number of exemptions by the county population. Throughout my analysis, I divide exemptions by the baseline county population in 2000, in order to provide a common base across all years. Unfortunately, the IRS county to county flows only provide aggregate numbers, and do not break up the migration levels by demographic characteristics (gender, marital status, education). As such, I am unable to use the IRS measures to look at differences across demographics. The only measure provided is the total adjusted gross income for all of the moved- returns. This is the adjusted gross income in the earlier year, but only the average for all movers in the county pair is provided.

The IRS data does not capture every move from one county to another. Low income individuals and households are not required to file a tax return, and thus might be under represented in

[28]The IRS censors county pairs that have fewer than ten returns move in each year. However, all of these returns are listed in a separate category as “from same state” or “from different state”. As such, when I collapse to the county level, I will capture the total number of returns, regardless of where they originated.

the data. It is likely that individuals that move to fracking areas will earn well beyond the filing threshold after moving, but they might not have been required to file in the previous year. If there are individuals that did not file in the first year, but moved in response to fracking and filed in the second year, my estimates would be attenuated. In order for the gap across geography to be biased upward, these individuals would have to be sorting into North Dakota. This systematic sorting would provide further evidence that people responded differently to the fracking boom in North Dakota.

The IRS data also does not capture temporary moves. Individuals who moved after filing in year t, but move back before filing in t+1 will not be counted as a move. Anecdotal evidence suggests that there was also large scale short-term relocation in North Dakota. My estimates will not fully capture this, but rather capture long-term adjustments. This measure likely seems more relevant when considering economic mobility, although it would be useful to test and see if individuals are responding by short term relocation rather than long term moving.

American Community Survey

To explore demographic differences and understand who moves, I use the American Community Survey (ACS) between 2005 and 2011. The ACS is an annual survey ran by the Census Bureau of approximately a one percent sample of households and has replaced the Census long for. All participants are asked where they lived one year ago, and both the previous state and local migration public use microdata area (MIGPUMA) are recorded. These MIGPUMA usually correspond to PUMA, but are enlarged to encompass the entire county. For rural areas MIGPUMA can often cover multiple counties or large portions of the state. When looking at fracking regions this can be problematic, as the MIGPUMA covering fracking areas also cover many surrounding counties. I identify the fracking status of a MIGPUMA, by simply indicating if it has any county with simulated production in it. I also do this separately for different plays (Bakken region) to look at heterogeneity within fracking. To the extent that I am capturing untreated areas as well, this will attenuate my estimates toward zero. Unfortunately, the boundaries for MIGPUMA changed in 2012. In many of the states the uniquely identifiable areas between 2005 and 2013 encompasses most of the state. For this reason I choose to focus on the ACS from 2005 to 2011. As such, I am not able to capture demographic characteristics in the later years, which might be important given the steep rise in North Dakota.

In all of my estimation using the ACS microdata, I collapse my observations from the individual level to unique cells. These cells are defined by demographics (e.g., gender, marital status, race, age group, education), migration status, fracking destination, and state of previous residence. When collapsing to these cells, I sum the individual weights provided by the Census Bureau and then use these weights in my regression analysis. These estimates are population representative and are identical to estimates obtained using weights at the individual level. Unfortunately, the migration questions from the 2000 Census ask about migration in the previous 5 years, and are thus not comparable to migration in the ACS.

U.S. Census Quarterly Workforce Indicators

The Quarterly Workforce Indicators (QWI) are constructed by the Census from the Longitudinal-Employer Household Dynamics (LEHD) Program and use firm level employment to construct aggregate employment and earnings reports. The QWI is aggregated from the Longitudinal Employer-Household Dynamics micro-level data collected from unemployment insurance earnings data from participating states and several other sources.[29] The QWI is aggregated to the county level, and can be tabulated by firm characteristics (industry, size) or worker characteristics (gender, age, education).[30] When tabulating by worker characteristics, only two levels of tabulation are feasible (gender by age or gender by education). Because I cannot separate migration from the IRS by demographics, I focus on earnings for men, as they seem to be the population responding. The QWI data is constructed through a state sharing process, and as such, only states that have made agreements with the Census have reported data. Many of the states began participating in 2000

[29]Most states began participating prior to 2000. However, during the years of the fracking boom South Dakota and Massachusetts did not participate in the data submission.

[30]I take the implied average annual wage across all four quarters weighting by the quarter specific employment to construct the group specific average wage for each year.

with most participating by 2003. As such, some states and counties are missing wage information in the early years. Most of these were not involved in fracking.

The main measure I use is the beginning of quarter earnings for all jobs. This measures the quarterly earnings for all jobs that existed at the beginning of the quarter. I choose this measure rather than stable jobs (spanning multiple quarters) and total jobs (employed at any time during quarter). I take the implied average annual earnings across all four quarters weighting by the quarter specific employment to construct the group specific average earnings for each year.

Because the QWI is constructed from firm employment, all measures are constructed for the job count. This means that average quarterly earnings are the average earnings of all jobs in a given quarter. Individuals who are unemployed are not considered, and individuals who hold two jobs will be treated as two separate individuals. In general, average earnings levels in the QWI are higher than those calculated elsewhere, as it records average earnings conditional on working. Also, because some workers might hold jobs for less than the full year, the average annual earnings constructed from the QWI will be higher, because my construction implicitly assumes the job lasts the entire year. This measure of earnings can be interpreted as the potential earnings if an individual was to move to the region.

DrillingInfo Well Database

Well level information on drilling date, lease agreements, location, direction, and geological formation as well as other characteristics are provided through a restricted use data agreement from DrillingInfo. This data is proprietary, and obtained through an academic use agreement with DrillingInfo, available through their academic outreach initiative. These well level characteristics are then merged to well level quarterly oil and gas production reports also provided by drilling info. Oil and gas production are reported in barrels and thousands of cubic feet respectively. Using the annual West Texas Intermediate crude oil price and the Henry Hub Natural Gas national prices provided by the Energy Information Administration (EIA), I convert these into dollar amounts and deflate to 2010 dollars.

DrillingInfo does not indicate if a well is a fracking well, as fracking is a means of stimulating production. To infer wells that are affected by the technological innovation associated with fracking, I use details on drilling direction and well location. Localized fracking booms occurred in part because of the combination of horizontal (directional) drilling and hydraulic fracturing. The DrillingInfo data reports whether a well is horizontally or vertically drilled. In addition, fracking was particularly impactful over shale plays, as these resources were not extractable previously. For this reason I assign non-vertical wells drilled in counties that intersect with shale plays as fracking wells.

Shale Play Boundary Shapefiles

Shale play boundary shapefiles are provided by the EIA in order to map the estimated boundaries of shale formations. These shapefiles have been updated over the years as new formations and reserves have been discovered. Prior to the shale boom, these formations had not be systematically mapped because they did not have economic value. I use the latest shapefile available at the time from 2015 to map shale play boundaries. These shapefiles are then overlaid by county shapefiles provided by the U.S. Census Bureau, and with the help of two research assistants I calculate the area of shale play and county intersections. This intersection measure is used when simulating production.

Housing Price Index

The Housing Price Index is constructed by the Federal Housing Finance Agency at the three digit zip code. Three digit zip codes span the entire country, allowing me to construct a measure for rural counties. To construct the county level measure I assign each county the average housing price index of all three digit zip codes that intersect the county, weighted by the share of the county in that zip code. For some three digit zip codes there is insufficient data, so the zip code is assigned the index from a larger geographic unit (such as the MSA or the state). I then adjust the housing price index baseline to be equal to 100 in 2000. Using the county level median house value from the 2000 Census, I convert the housing price index to dollars. A similar developmental index is available at the county level but does not include all counties. I find that both indices follow similar patterns for the available counties.

Chapter 3: Moving to Jobs: The Role of Information in Migration Decisions

3.1 Introduction

Migration is a human capital investment that provides access to more favorable labor market opportunities (Shultz, 1961; Sjastaad, 1962). However, people are unlikely to move away when areas experience adverse labor demand shifts (Monras, 2015), even though there is significant geographic heterogeneity in employment prospects and many could encounter more favorable labor markets by moving to a different state or county (BLS, 2015). These patterns have led to a growing concern and puzzle, as to why populations in weak labor markets appear unlikely to move to labor market opportunities.[1] As the previous literature recognizes, low migration might be the outcome of optimal decision-making, but might also be the result of market frictions such as credit constraints or incomplete information.

One factor that has been difficult to evaluate empirically is the role of information in migration decisions. While incomplete information introduces uncertainty and increases the risk associated with the migration "investment", information provision may reduce perceived risk and change migration outcomes. In this paper, I evaluate the role of information in migration decisions by asking, does information in the news about potential, lucrative employment opportunities in other labor markets induce people to move to those markets?

The role of information is frequently overlooked because it is difficult to measure and identify a causal relationship. I overcome these challenges by exploiting a unique setting that allows me

[1]Not only is out-migration from negatively affected areas low, but it has decreased over time (Dao, Furceri, & Loungani, 2017). This concern is evident in the news, such as the *New York Times'* "Fewer Americans Strike Out for New Jobs, Crimping the Recovery" by Patricia Cohen (2016) and "How to Get Americans Moving Again" by Arthur Brooks (2016).

to isolate one source of information transmission. Over the last 10 to 15 years, the combination of two technologies, horizontal drilling and hydraulic fracturing, has led to localized "fracking booms" and large local labor market shocks. These booms have not only affected workers in oil and gas extraction, but have created large, persistent increases in employment and earnings across industries (Feyrer, Mansur, & Sacerdote, 2017; Maniloff & Mastromonaco, 2014), resulting in perceived net benefits (Bartik et al., 2017). From the beginning, fracking has been associated with "boom towns", local economic growth, and a flurry of economic activity. The "gold rush" style approach to fracking, and later the environmental and safety concerns, have led to numerous newspaper articles and television news broadcasts touting the economic impacts or debating the adverse side effects. Because fracking is so novel (not only as a technology but also as a term), it is easy to track news coverage of fracking and the areas it has affected – for better or for worse.

Using the universe of newspaper articles from large national newspapers (*USA TODAY*, *New York Times*, *Wall Street Journal*) from LexisNexis, and national TV news broadcasts from the Vanderbilt Television Archive, I am able to identify the extent to which the fracking booms across 16 states are reported in the news. I then combine this with proprietary pre-fracking newspaper specific circulation rates from the Alliance for Audited Media (AAM) and TV channel viewership rates from the Television and Cable Factbook to construct a measure of news exposure. I investigate whether migration flows to a given fracking area increase more from origin counties that experience higher levels of exposure to news about that area. Combining information on stories from national news sources (whose content decisions are plausibly exogenous to local migration trends) with information on pre-fracking circulation and viewership (to eliminate endogenous changes in penetration) allows me to capture exogenous variation in news exposure. By comparing migration flows from different origin counties to a particular fracking destination state, I am able to isolate the effect of news and estimate the causal impact of news exposure on migration flows.

The data suggest that for an origin county with a five percent circulation rate, one additional newspaper article about fracking in a specific state increases migration flows to fracking counties in that state by 2.4 percent. This relationship controls for characteristics of the destination that are changing over time. The estimated relationship does not change when I control for changing characteristics of the origin or exposure to fracking news through local newspapers. The data also suggest that one additional newspaper article about fracking in a specific state increased cross-county commuting flows to fracking counties in that state by 6.6 percent. These patterns are robust to functional form, sample restrictions, placebo tests, and an alternative strategy comparing neighboring counties around local newspaper market boundaries. My estimates suggest that for the average origin county, the 36 articles about fracking in Pennsylvania in the *New York Times* in 2011 increased migration flows to Pennsylvania fracking counties by 8.2 percent and increased commute flows to these counties by as much as 22.5 percent. If there had been no news about local fracking booms in 2012, migration flows to fracking counties would have been 4.2 percent lower and commute flows would have been 11.7 percent lower on average.

I conduct additional analyses to better understand how the news influences migration behavior. Looking simultaneously at both newspaper exposure and TV news exposure, I find responses to both sources of information. The percentage effects on migration are similarly sized, despite the fact that typical TV viewership is significantly higher than typical national newspaper circulation rates. For commute flows the estimated impact of TV news exposure is larger than newspaper exposure, although not statistically different. Using linguistic techniques, I find that news articles that are more "positive" discussing things like, jobs, booms, or growth, have a larger positive effect on migration than "negative" news articles, discussing contamination, pollution or earthquakes. The effect of negative news is still positive, suggesting "any news is good news" as it might provide information about where fracking is occurring. Positive and negative news affect cross-county commuting similarly, consistent with the theory that non-resident workers mostly experience the benefits of fracking while not incurring many of the costs (e.g., potential home water contamination). Migration significantly responds to content in *USA TODAY*, less so to the *New York Times*, with no significant response to the *Wall Street Journal*. The effect of news exposure on both migration and commuting also varies with distance, peaking for counties 400 to 1,000 miles away from the potential fracking destination, consistent with people being aware of nearby opportunities, but lacking information about distant labor market opportunities.

The marginal impact of newspaper exposure is largest in origin counties with low employment to population ratios, suggesting that providing information about potential labor market opportunities has a larger impact in poor performing labor markets. Policy simulations suggest the level of news that maximizes the migration response is 20 percent higher in poor performing labor markets than in strong labor markets and the corresponding migration response is twice larger. These heterogeneous impacts are not due to differences in news exposure. All regions would benefit from more information, but information provision has the largest impact in weaker than average labor markets where the gains to moving are also plausibly the largest. Targeted information provision in poor performing labor markets can increase geographic mobility and potentially result in more beneficial labor market transitions (Molloy, Smith, Trezzi, & Wozniak, 2016) and higher economic mobility (Chetty & Hendren, 2016).

As further evidence that information in the news matters, I look to Google search data. In the days following a broadcast about fracking, search interest in "fracking" increases and there are more searches for the specific states mentioned in the news broadcast. Although this does not confirm that individuals are interested in moving, it does suggest that the news spurs them to seek more information on the internet, which might inform their decision.

The paper is organized as follows. Section II discusses the related empirical evidence and formalizes a conceptual model of information in migration decisions. Section III briefly summarizes the data. Section IV explains the empirical strategy. Section V presents the main results. Section VI explores an alternative strategy using newspaper market borders. Section VII presents additional explorations. Section VIII explores internet searches as a potential mechanism. Section IX concludes.

3.2 Information in Migration Decisions

3.2.1 Related Empirical Evidence

The propensity to migrate varies significantly by demographics, educational attainment, and geographic region (Molloy, Smith, & Wozniak, 2011), and some demographic groups are more likely to move in response to local labor market conditions than other groups (Bound & Holzer, 2000; Wozniak, 2010). Differential migration responses might be optimal if, for example, individuals are differentially affected by labor market shocks (Notowidigdo, 2013), or if strong labor markets also have high costs of living, resulting in a small or negative net return to migration for low wage workers (Ganong & Shoag, 2017). However, there is also credible evidence that liquidity constraints, credit constraints, and other market frictions impact the migration decision (Kling, Liebman, & Katz, 2007; Bryan, Chowdhury, & Mobarak, 2014). One potential frictions is a lack of information.

It has long been recognized that information will affect migration decisions, but most of the empirical work has been limited to focusing on the role of networks or linguistic and cultural enclaves.[2] Although there is not much work that speaks directly to the impact of labor market information on migration decisions, several recent studies have explored somewhat related topics. The Moving to Opportunity (MTO) experiment, which provided guidelines and information about local neighborhood poverty along with housing vouchers, induced treated households to move to more affluent neighborhoods, suggesting this type of information can change migration behavior (Kling et al., 2007). Although this did not improve economic outcomes for treated adults, recent work has found positive long-run effects for the young children who were treated (Chetty, Hendren, & Katz, 2015). Malamud and Wozniak (2012) exploit variation in the Vietnam draft and find that college attendance causally increased the incidence of migration. They suggest that exposure to other areas, and peers from other areas, provide information about alternative labor market opportunities. Using a structural model, Kaplan and Schulhofer-Wohl (2017) propose a framework,

[2]See for example Greenwood (1975), Winters, de Janvry, & Sadoulet (2001), Munshi (2003), McKenzie & Rapoport (2007, 2010), and Hanson & McIntosh (2010).

where information helps people learn about amenities in different locations.

In a randomized controlled trial in Bangladesh, households that received information about potential labor market opportunities and a conditional cash transfer were more likely to migrate, while households that were only given the information were not (Bryan et al., 2014). This suggests that relaxing credit constraints and information barriers together could increase migration. Farre and Fasani (2013), show that as villagers in Indonesia gain access to more TV stations, they become less likely to move. The authors propose that this is because media access corrects overly optimistic expectations of the return to migration. However, it is difficult to generalize the results from Bangladesh and Indonesia to the United States.[3] The question of how to conceptualize the role of labor market information in migration decisions still remains.[4]

3.2.2 Conceptual Framework: Information and Migration

In this section I present a conceptual framework to show how information frictions affect migration behavior. Although I do not estimate the structural parameters of this model, the intuition helps inform the empirical strategy. In the canonical migration choice model, the decision to move is an investment in human capital (Sjaastad, 1962). An individual will move if the lifetime utility derived from moving minus the fixed costs of moving exceeds the utility of staying at the original location. The individual observes the fixed utility cost c_{od} associated with migrating from o to d as well as the real returns $y_d(t)$ and $y_o(t)$ for each period in each location, which are defined to account for earnings, cost of living, local amenities, and idiosyncratic fit. The location specific returns can vary over time, and are discounted by β. Assume the individual is risk averse and has monotone preferences (the utility function is strictly increasing and concave). The decision to

[3] For example, the conditional round-trip transfer in Bangladesh was only equal to \$8.50 (about one weeks work), suggesting these people are highly credit constrained (Bryan et al,. 2014).

[4] An early related literature explores how things like the risk of unemployment (Todaro, 1969) and uncertainty about the future affect migration and human capital investments more generally (see Becker, 1962; Greenwood 1975, 1985; Langley, 1974; O'Connell, 1997). Under uncertainty, different states of the world occur with some known probability. Under incomplete information, potential destinations, possible states of the world, and the true probabilities are potentially unobserved.

move from o to d (m_{od}) is characterized as follows

$$m_{od} = \begin{cases} 1 & if \quad \sum\limits_{t=0}^{T} \beta^t u(y_d(t)) - c_{od} \geq \sum\limits_{t=0}^{T} \beta^t u(y_o(t)) \\ 0 & \qquad\qquad else \end{cases} \tag{3.1}$$

But individuals likely face uncertainty about conditions in the potential destination such that $y_d(t)$ is a random variable, where $y_d(t) \sim G(y;\theta)$.[5] The individual will thus decide to migrate if

$$\sum_{t=0}^{T} \beta^t \big(Eu(y_d(t)) - u(y_o(t))\big) - c_{od} \geq 0 \tag{3.2}$$

where the E operator is the expected value at time zero. Changes in the parameters θ will affect the outcome of this decision. For example, define

$$c^*_{od} = \sum_{t=0}^{T} \beta^t \big(Eu(y_d(t)) - u(y_o(t))\big). \tag{3.3}$$

The value c^*_{od} is the threshold moving cost at which the individual is indifferent between staying and moving. If $y_d(t)$ is distributed normally with a mean (μ_d) and variance (σ^2), the nature of u implies that

$$\frac{\partial c^*_{od}}{\partial \mu_d} = \sum_{t=0}^{T} \beta^t \frac{\partial Eu(y_d(t))}{\partial \mu_d} > 0 \text{ and } \frac{\partial c^*_{od}}{\partial \sigma^2} = \sum_{t=0}^{T} \beta^t \frac{\partial Eu(y_d(t))}{\partial \sigma^2} \leq 0. \tag{3.4}$$

Intuitively, as the mean increases, less weight is placed on low values of y_d and expected utility rises. This increases the threshold moving cost, and the individual will be willing to pay a larger cost to move. An increase in the variance, holding all else equal, represents a mean preserving spread which results in weakly lower expected utility because the individual is risk averse (Rothschild & Stiglitz, 1970).[6] The increase in variance increases risk, and the individual's moving cost threshold becomes smaller, as she must be compensated by a lower cost to move.

People might have incomplete information about the parameters that govern the distribution

[5]The model implications are similar if the individual is also uncertain about conditions at the origin.

[6]If the distribution of $y_d(t)$ is governed by more than just locational parameters this is not necessarily true (Tobin, 1965; Dionne & Harrington, 1991). More generally, if $\hat{\sigma}^2$ is a mean preserving spread of σ^2, then Rothschild and Stiglitz (1970) prove that $Eu(y_d(t);\sigma^2) \geq Eu(y_d(t);\hat{\sigma}^2)$. If instead the utility is linear and individual is risk neutral, changes in the dispersion that preserve the mean will not affect the cost threshold.

of $y_d(t)$, and this additional uncertainty will also affect migration decisions.[7] For example, if the individual's prior belief is that the return to migration is low, she will be less willing to move. Similarly, if her prior is diffuse and the investment in migration appears more risky, she will also be less willing to move. Receiving additional information can change migration outcomes as individuals update their beliefs about these parameters. Specifically, exposure to news stories that credit fracking with creating local booms, fueling local economic growth, or raising wages in potential destinations might change people's perceptions of the distribution of the returns in the fracking destinations mentioned; even negative news can provide information about where fracking is occuring and change people's beliefs.[8] For example, individuals exposed to numerous newspaper articles and TV news broadcasts touting the local economic benefits of fracking in Texas might adjust their original mean or dispersion beliefs about the returns to migrating to a Texas fracking county. This news information does not necessarily need to be correct, as long as the individual believes it contains truthful information.

When an individual receives information in the news about fracking in a specific destination, she can update her prior beliefs following a process like Bayes' Rule. Under Bayesian updating, sample moments from the new information are used to update posterior beliefs. Although individuals might not perform exact Bayesian updating to incorporate new information, using sample moments from information in the news seems reasonable at an intuitive level.[9] Observing a large sample mean (news that the return is high) will increase the posterior belief about the mean, but the magnitude of this increase will depend on how precise or diffuse the prior belief is. Observing more information provides a larger sample and reduces uncertainty about the parameters. Howe-

[7]This type of uncertainty is prevalent. Even among highly educated medical students in the residency match process there is substantial heterogeneity in their ability to accurately predict the expected cost of living and earnings rank in their top two ranked locations (Bottan & Perez-Truglia, 2017).

[8]Up through 2012, the last year of my sample, about 60 percent of adults were familiar with fracking, and over half of this population was in favor of fracking (Pew Research, 2013a). For someone that views fracking favorably, even a negative news story could provide information about where fracking is occurring, and result in updated beliefs.

[9]Wiswall and Zafar (2015) show that when college students receive information about the distribution of earnings, they update their beliefs, but often do not strictly follow a Bayesian updating process.

ver, the marginal impact of information becomes smaller as she gets more information and her beliefs converge to the true distribution.

As the individual incorporates new information about the parameters, she better understands the distribution of $y_d(t)$ and can compute the likelihood of observing the return y given her information set. The effect of additional information on the perceived mean of $y_d(t)$ will depend on her prior beliefs. If she initially believed the average return in a potential destination was lower than the news suggested (a likely case given the coverage about fracking jobs and booms), the information will increase her perception of μ_d. This in turn increases c^*_{od}, meaning she is more willing to move (see equation (3.4)). The converse is also true. Similarly, receiving more information will reduce uncertainty, which makes moving less risky all else equal. If the individual is uncertain about the distribution of potential returns in a destination, receiving information in the news that portrays $y_d(t)$ as larger than she initially believed will increase her propensity to migrate.[10] Although I do not estimate the structural parameters of this individual level model, the predictions provide motivation for how geographic differences in news exposure might affect migration behavior.

3.3 Data Sources

Fracking provides a unique setting to explore the impact of news exposure on migration outcomes. Fracking began quite suddenly in the mid-2000s and by 2012 had affected oil and gas production in 252 counties in 16 states. These local fracking booms increased economic activity and

[10] Appendix Figure A3.1 presents results from a simulation that illustrates how new information provision can change migration decisions. Two scenarios are presented for three types of people. Individual 1 has a diffuse prior over the expected return (μ_d) and incorrectly believes μ_d is lower than the true mean. Individual 2 has a precise belief that μ_d is low. Individual 3 has a diffuse prior, but correctly predicts μ_d. In both scenarios the true parameters are the same, the only difference is that individual are exposed to more information in scenario 1 than in scenario 2. The perceived distribution of both μ_d and y_d are plotted for each individual in each scenario over two iterations of receiving more "news". If initial beliefs about the expected return are low, new information shifts up the beliefs about μ_d and y_d. Additional information also reduces uncertainty about μ_d and y_d, which increases expected utility and the probability of migrating. Updating is more drastic when there is more information, and changes in the probability of moving will be the largest among people or areas that are exposed to more new information. Initial draws of information are very beneficial, but the marginal value of additional information becomes smaller.

improved labor markets in those counties (Feyrer et al., 2017). States in all four census regions have been affected and many people were unaware of exactly where these fracking booms were occurring. Both positive and negative aspects of fracking have been highly publicized through newspapers and TV news, and many of these news stories reference specific locations affected by fracking. Because fracking is a novel term, I am able to parse news content to identify which sources discuss fracking, which places they talk about, and what aspects of fracking were discussed. By linking this with measures of news penetration, I am able to estimate how geographic differences in exposure to news about fracking affect migration flows. This estimation requires detailed data on fracking production, migration flows, news content, and news circulation. In this section I briefly describe each data source and highlight key strengths and limitations with a full description in the online data appendix.

Fracking Production Data. Oil and gas extraction data is obtained through a proprietary data agreement with the private company DrillingInfo. DrillingInfo provides well-level information on drilling and quarterly production. The two technologies that characterize fracking, horizontal drilling and hydraulic fracturing, have drastically increased the productivity of thick layers of dense shale rock, known as shale plays. Using the well characteristics, I define a fracking county as any county with positive oil or gas extraction from a non-vertical well in a drilling formation that corresponds to a shale play. This amounts to fracking regions in 16 states: Arkansas, California, Colorado, Louisiana, Michigan, Mississippi, Montana, New Mexico, North Dakota, Ohio, Oklahoma, Pennsylvania, Texas, Utah, West Virginia, and Wyoming as seen in Figure 3.1.

Migration Data. Migration data is obtained from the Internal Revenue Service (IRS) Statistics of Income (SOI). Using tax documentation, such as Tax Form 1040, the IRS tracks the number of households that filed their taxes in one county in one year and in a different county the next year. Most filing occurs between February and April, so annual migration flows capture moves from approximately March or April from one year to the next.[11] This measure is then aggregated

[11]For example, migrants who moved between March/April of 2011 and March/April 2012 will be assigned the year 2011. This introduces a slight lag relative to the measurement of news (from January to December).

up to the county to determine the approximate flow of households (returns) and individuals (tax exemptions) from one county to another. These two numbers are provided for pairs of counties in the United States, but is censored for county pairs with fewer than 10 returns for privacy purposes. In 2013, the censoring threshold increased from 10 to 20 returns, leading to much higher levels of suppression. For this reason I restrict my analysis to migration between 2000 and 2012. The IRS data only provide a raw count and do not provide information about demographic characteristics. Although this is perhaps the most comprehensive data on internal migration in the United States, it might under-represent a subset of the extremely poor (who fall below mandatory tax filing thresholds and do not file for other benefits such as the Earned Income Tax Credit) as well as a small subset of the extremely wealthy (who are more likely to be granted filing extensions for complex returns).

Newspaper Circulation Data. Proprietary newspaper readership data is obtained from the Alliance for Audited Media (AAM). The AAM conducts regular newspaper circulation audits for national, regional, and most local newspapers in the United States. This includes the number of copies sold on the audit date and the number of copies as a percent of households for each county with over 25 copies. Counties with fewer than 25 copies sold are assigned a zero value. For some newspapers, these measures are only available at the Designated Market Area (DMA) level. Historic circulation rates from 2005 through 2008 are scraped from pdf files.

TV Viewership Data. TV viewership data is calculated from the Television and Cable Factbook using Nielsen viewership data. For my analysis, I use viewership rates from both the 2008 and 2016 Factbook. Between 2007 and 2009, TV stations were transitioning from analog to digitally transmitted broadcasts on a market-by-market basis. When a market transitioned, viewers were required to obtain digital reception equipment, and it is unclear how this affected viewership and if 2008 viewership is representative of later years.[12] For this reason I also examine the most recent viewership rates from 2016. TV viewership is reported at the DMA level for each

[12] A special thanks to Matt Long from Warren Communication News for finding out how the viewership rates for the 2008 Factbook were constructed, and to Colin Wick for transcribing viewership rates from the 2008 Factbook.

TV station and includes viewership from both cable and non-cable households. These data are available at the station-level and are not specific to news programming. The viewership rate is constructed by dividing total weekly viewership by the total number of households in the DMA.

Newspaper Content Data. Newspaper content is obtained through the LexisNexis database, which provides access to articles from over 2,600 news sources, including *USA TODAY*, the *New York Times*, and the *Wall Street Journal.* First, I preserve all articles since 1999 that include any of the search terms "frack∼", "fracing", or "hydraulic fractur∼" anywhere in the text. I then linguistically parse each article to exclude spurious keyword references such as "frick and frack", unrelated acronyms, and the last names of people. Most of my analysis is restricted to three national news sources: *USA TODAY*, the *New York Times*, and the *Wall Street Journal.*[13] The in depth news coverage of fracking begins in 2009, and dramatically increases each year. In these three newspapers there were 562 news articles related to fracking between 1999 and 2012. The first two articles in the national news were in 2002 and 2003 in the *New York Times*, which briefly reference court cases about patents related to hydraulic fracturing. There was then one article in 2006, five in 2008, 20 in 2009, 48 in 2010, 198 in 2011, and 288 in 2012. Next, I linguistically parse the entire text of each of these articles to determine which of the 16 fracking states listed above each article discusses.[14] I also parse each article for specific keywords such as "growth", "boom", "contaminat∼", and "earthquak∼" to determine the positive and negative content of each article (discussed in detail later). These statistics are reported in Appendix Table A3.1. Articles that mention specific states are more likely to refer to things like jobs, booms, and growth and there is also heterogeneity across states in how frequently these "positive" effects of fracking as well as "negative" effects such as pollution, danger, and earthquakes are cited.

TV News Content Data. TV news content is obtained from the Vanderbilt Television News Archive (VTNA). The VTNA database contains TV news recordings and transcript abstracts for

[13]When including local news coverage, I restrict the sample to news articles from domestic US newspapers.

[14]Not every article mentions a specific state. I have also parsed each article for city names from the U.S. Postal Service's registry of city names, but find that local jurisdictions are referenced far less frequently.

nightly news broadcasts from the three major news networks (ABC, CBS, and NBC) and the cable news channels CNN and Fox News. The database only includes one hour of programming each day for both cable news outlets. Because the available content of cable news is limited, and viewership rates are only available for the TV networks, I restrict the sample to TV broadcasts from the three major news networks. I parse the transcript abstracts for search terms such as “fracking” and “shale” as well as which state is being discussed. Between 1999 and 2012 there is far less coverage of fracking on the nightly news than in the newspaper. The VTNA database only records 17 news broadcasts, with one in 2006, two in 2008, three in 2010, four in 2011, and seven in 2012.

Cross-County Commute Data. In addition to migration, I also explore impacts on workers who live in one county but work in another. This captures both long distance commuting and temporary relocation, such as moving to the job site for several weeks at a time but maintain the same permanent address. This data is available through the Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES). This data provides statistics on the number of jobs for each home and work census block pair. I aggregate up these pairs to the county level to determine how many workers live in one county but work in another. This data is available for all years since 2002, and also provides statistics by broad age groups (under 30, 30-54, over 54), monthly earnings (under $1,250, $1,250-3,333, over $3,333), and industry (goods, trade/transportation, other). This allows me to explore heterogeneous commute responses across different groups.

County Characteristics Data. County level economic and population characteristics are obtained from a range of sources. Economic outcomes such as employment to population ratios, unemployment rates, and average earnings are obtained from the Bureau of Labor Statistics (BLS) Quarterly Census of Employment and Wages (QCEW). County-level age and racial demographics are obtained from the National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) Program population data and are constructed from the U.S. Census Bureau’s Population Estimates Program. Other county level characteristics are obtained from the 2000 Census and ACS through the American Factfinder.

3.4 Empirical Strategy

To estimate the impact of labor market news on migration behavior, I exploit county-level differences in exposure to news about fracking in a specific destination state. To capture this variation, consider the following measure

$$newspaper\ exposure_{oSt} = \sum_{n \in N} \Big(\text{total articles in } n \text{ about fracking in } S\Big)_t * circ.\ rate_{ont} \tag{3.5}$$

where N is the set of all domestic newspapers. *Newspaper exposure*$_{oSt}$ is the weighted sum of news articles that mention fracking in destination state S in year t, where each newspaper is weighted by its circulation rate in the origin county o (ranging from zero to one). News exposure varies at both the origin and destination level, and increases when there are more news articles, but the magnitude of this increase is weighted by circulation rates. To identify the causal impact of labor market news exposure on migration, consider the hypothetical relationship between newspaper exposure and Y_{oSt}, a measure of migration flows from origin county o to fracking counties in destination state S in year t:

$$Y_{oSt} = f(newspaper\ exposure_{oSt}; \theta) + \phi_{oS} + \eta_{oSt}. \tag{3.6}$$

Migration flows are composed of three parts. The first part is a flexible function of newspaper exposure. The second part is an origin/destination pair specific level effect that accounts for time invariant origin/destination specific characteristics, such as distance or industry ties. The third part is an error term, η_{oSt}, composed of two systematic components that vary over time: destination state characteristics (ψ_{St}) and origin county characteristics (λ_{ot}), as well as an idiosyncratic origin-by-destination pair specific error term (ν_{oSt}) as follows

$$\eta_{oSt} = \psi_{St} + \lambda_{ot} + \nu_{oSt}. \tag{3.7}$$

If *newspaper exposure*$_{oSt}$ is correlated with destination specific, origin specific, or pair specific characteristics that are changing over time, the estimated effect of *newspaper exposure*$_{oSt}$ will be a biased estimate of the causal effect of labor market news on migration.

Consider the following thought experiment: if a set of origin counties are randomly assigned different levels of exposure to news about fracking in Texas (or any other fracking state), do counties that were more exposed to this information see larger increases in migration to fracking counties in Texas? By random assignment, news exposure will be uncorrelated with the unobserved origin component (λ_{ot}) and the origin/destination idiosyncratic error (ν_{oSt}). By comparing migration flows to the same destination, everything about fracking counties in the destination state is held constant, allowing the effect of news exposure on migration to be isolated. The relationship over multiple destinations can be explored in a regression framework by stacking the estimation as follows

$$Y_{oSt} = f(newspaper\ exposure_{oSt}; \theta) + \phi_{oS} + \psi_{St} + \varepsilon_{oSt}. \tag{3.8}$$

The level of observation is the annual migration flow from an origin county to all fracking counties in a destination state. An origin/destination pair fixed effect, ϕ_{oS}, is included to control for time invariant characteristics of the pair that affect migration, like distance. A destination state-by-year fixed effect, ψ_{St}, is also included to control for destination specific characteristics that are changing over time, and makes this a comparison of migration flows to the same destination state from origin counties that have different levels of news exposure. Importantly, this fixed effect captures changes in fracking production, labor market characteristics, and amenities which might directly affect migration behavior and lead to higher news exposure.

Unlike the thought experiment, actual exposure to news about fracking in a specific region is not exogenously assigned and endogenous to decisions of both news providers and consumers. This endogenous variation in news exposure might be correlated with both origin (λ_{ot}) and origin/destination (ν_{oSt}) characteristics and could lead to biased estimates of the causal impact of news exposure on migration in equation (3.8). To obtain exogenous variation in news exposure, two major concerns about endogeneity must be addressed.

The first concern is the endogeneity of news content arising from news providers' decisions. For example, if people from Franklin County, Ohio start moving to Alleghany County, Pennsylvania, the local *Columbus Dispatch* might produce more content about fracking in Pennsylvania, raising concerns about reverse causality. This concern seems particularly relevant for local newspapers, where content decisions strongly respond to consumer preferences in local geographic markets (Gentzkow & Shapiro, 2010). However, large, national newspapers such as *USA TODAY*, the *New York Times*, and the *Wall Street Journal* do not have a well-defined geographic market and operate differently (Gentzkow & Shapiro, 2010).[15] These newspapers are read throughout the country and make content decisions to cater to the nation as a whole. While content decisions of the *Dearborn County Register* might be endogenous to the number of people moving from Dearborn County, Indiana to fracking counties in Ohio, content decisions of the *New York Times* are likely driven by aggregate trends rather than idiosyncratic patterns. Counties across the country are exposed to the same national news, regardless of their idiosyncratic deviation from the national trend.[16]

Although the same national news is available across the country, counties will be differentially exposed to newspaper articles about fracking because they vary in their reading habits (circulation rates). This raises the second concern of endogenous news consumers' decisions. Over time, residents of Franklin County, Ohio might come to view fracking more positively (or negatively), which might affect both their readership of the *New York Times* and migration to fracking destinations, leading to omitted variable bias. However, because fracking began quite suddenly there is a clear pre-period when circulation was not a function of preferences toward fracking. The largest expansions in oil and gas production due to fracking started after 2008 and national news about fracking only began in earnest in 2009. By using pre-2009 circulation rates, I can isolate variation in exposure due to pre-fracking differences in circulation rather than changes in exposure that arise from changing preferences toward fracking.

Exploiting variation in news exposure due to national news content and pre-fracking circula-

[15] Gentzkow and Shapiro (2010) also list the *Christian Science Monitor* as a national newspaper, however, circulation for this newspaper is only available at the state level, so it is excluded from all analysis.

[16] National newspapers might report more about destinations that see large changes in labor markets or migration (nationwide trends). However, including destination by year specific effects compares migration flows from different origins to the same destination, eliminating destination specific differences that might drive news coverage. It could be argued that readers in and around New York City have a large effect on the content decisions of the *New York Times*. As a precaution, I exclude counties in the New York City DMA from the analysis. In Column (1) of Appendix Table A3.5 I show that the migration response is still significant if the New York City DMA is included.

tion can mitigate concerns about endogenous decisions of both news producers and news consumers in the modified definition of $newspaper\ exposure_{oSt}$ originally presented in equation (3.5)

$$newspaper\ exposure_{oSt} = \sum_{n \in N'} \Big(\text{total articles in } n \text{ about fracking in } S\Big)_t * 09\ circ.\ rate_{on}. \quad (3.9)$$

The set of newspapers is now restricted to national news sources, $N' = \{$*USA TODAY, New York Times, Wall Street Journal*$\}$, and the number of articles is weighted by the pre-2009 circulation rate, which is the average circulation between 2005 and 2008. As such, an additional news article in a national newspaper will increase newspaper exposure, but this increase will be largest in counties that had high readership prior to the fracking boom. This strategy is similar to previous work using variation in circulation exposure to explore the impact of media and news on other outcomes.[17]

When using this new definition of newspaper exposure to estimate equation (3.8), the effect of national news exposure is identified by variation across origin counties in pre-fracking circulation rates. This variation is potentially problematic if pre-fracking circulation is correlated with changes over time in other local characteristics that affect preferences to move to fracking, captured in λ_{ot}. For example, if baseline readership of the *New York Times* was higher in more affluent counties and the distribution of income became more dispersed over time, this differential increase in income would introduce omitted variables bias if it is correlated with both baseline circulation and residents' propensity to migrate to fracking. Because fracking was such a new, unknown technology, it is not always clear what direction these characteristics might bias the estimates.[18] As seen in Figure 3.2, there is no obvious, strong geographic correlations in pre-2009 circulation

[17]For example, Gentzkow (2006) examines TV introduction on voter turnout, DellaVigna and Kaplan (2007) examine Fox News introduction on Republican vote shares, Jensen and Oster (2009) examine Indian cable introduction on women's status, Chong and La Ferrara (2007) and La Ferrara et al. (2012) examine Brazilian soap opera introduction on divorce and fertility, Garthwaite and Moore (2012) examine exposure to Oprah Winfrey content on votes for Barack Obama after her endorsement, Kearney and Levine (2015a;2015b) examine exposure to the MTV series "16 and Pregnant" on teen births, and exposure to Sesame Street on grade completion.

[18]For example, even the prominent environmental organization, the Sierra Club, went from supporting fracking and natural gas extraction (as a cleaner alternative to coal) to later condemning it (Gold, 2014).

of the *USA TODAY* and even within close regions there is significant variation. As seen in Table 3.1, areas with low and high circulation of the *USA TODAY* were similar in 2000 on average.

Although counties with below median circulation of the *USA TODAY* had slightly lower employment, lower median income, higher poverty, and an older population, these level differences will be controlled for by the origin county by destination state fixed effects. Of more concern to causal identification are changes over time that are correlated with pre-fracking circulation. Columns (3) and (4) in Table 3.1 suggest that between 2000 and 2010 both low and high circulation counties followed parallel trends in migration. Other origin county characteristics evolved similarly in low and high circulation counties, although high circulation counties saw relatively larger decreases in employment and the percent white, and larger increases in unemployment, median income, and poverty, slightly closing the gap that existed in 2000 between low and high circulation counties. However, these differences are never more than one or two percentage points.

Column (5) formally tests if pre-2009 *USA TODAY* circulation rates predict differential changes in origin county characteristics between 2000 and 2010. Pre-2009 circulation rates do not predict changes in migration or the employment to population ratio, but some of the other local characteristics are statistically different. However, these differences are quite small: an increase in readership from the 25th to the 75th percentile of *USA TODAY* circulation (1.88 percentage points) predict a 0.09 percentage point reduction in the unemployment rate, a $290 increase in median income, a 0.51 percentage point increase in the poverty rate, a 0.17, 0.29, and 0.41 percentage point increase in the percent black, Hispanic and other race respectively, and a 0.27 percentage point decrease in the population 35 to 64. *New York Times* and *Wall Street Journal* pre-2009 circulation rates predict similarly small changes in county characteristics (Appendix Table A3.2).[19] These differential trends in origin county characteristics are small and unlikely to have large effects, but it is still possible that they might bias the estimated effect of news exposure.

To evaluate if these potential threats to identification are valid concerns, I estimate the effect

[19]Readership of the *New York Times* and the *Wall Street Journal* are highly correlated, and the predicted effects are similar. The one characteristic that varies the most across newspapers is median household income. The *New York Times* and *Wall Street Journal* have higher readership in large urban areas that saw larger increases in earnings.

of national news exposure on migration in a series of progressively more conservative regressions. I first estimate the model corresponding to the thought experiment as follows

$$Y_{oSt} = \beta_1 newspaper\ exposure_{oSt} + \beta_2 newspaper\ exposure^2_{oSt} + \phi_{oS} + \psi_{St} + \varepsilon_{oSt}. \tag{3.10}$$

The main outcome of interest is the inverse hyperbolic sine of the number of migrants from origin county o to fracking counties in state S in year t. The inverse hyperbolic sine approximates a natural log transformation but is defined for flows with zero migrants, allowing me to measure the percent effect of news exposure. Origin county by destination state fixed effects are included to control for time-invariant pair specific characteristics, and destination state by year fixed effects are included to account for changing characteristics of the fracking destinations. To account for correlated shocks across geography, the standard errors are adjusted for clustering at the origin DMA level (203 clusters), a geographic measure meant to capture media markets. As suggested by the theoretical framework, I include news exposure quadratically to capture decreasing marginal returns to information, although the relationship is robust to different functional forms. I begin with the specification in equation (3.10) because the identifying variation is highly transparent: origin counties experience different exposure to news about a specific destination because they have different pre-fracking circulation of national newspapers.

I then progressively adjust this baseline specification to address potential concerns associated with this identifying variation. First, I include a vector of time varying origin county labor market controls, including the employment to population ratio, the unemployment rate, and average earnings (in 2010\$). These controls account for observable changes in the origin labor market that might be correlated with news exposure and affect migration. Second, I include origin county by year fixed effects which account for both observed and unobserved components of λ_{ot}. This is possible because I observe migration flows to 16 different fracking states from each origin county/year pair. Including origin county by year fixed effects will control for changing characteristics of the origin county that affect all of these migration flows. For example, if counties with higher readership of the *New York Times*, and thus higher newspaper exposure, become more opposed to fracking over time, this omitted variable might affect decisions to move to fracking in general. Origin county by year fixed effects will absorb these and other changes over time and exploit variation in news exposure across potential destinations from the same origin. This specification tests to see if, for a given origin county, destination states that had more news exposure also experienced larger increases in migration flows. In this specification, confounding omitted variables must be origin/destination pair specific and vary over time (contained in ν_{oSt}).

For example, if local and national news exposure are strongly correlated and local news is endogenous to migration preferences, omitting local news will bias the coefficient on national newspaper exposure.[20] I combine the content of all domestic newspapers available through LexisNexis with circulation rates, and construct an analogous measure of local newspaper exposure to test and see if local newspaper exposure changes the estimated effect of national newspaper exposure.[21] I also conduct placebo tests and alternative estimation strategies designed to test if the observed relationship is actually driven by unobserved origin/destination specific changes over time.

In these specifications, the sample is restricted to exclude origin counties involved in fracking as information in the news might effect the decisions of people originally living in fracking counties differently. For example, residents of Bradford County, a fracking county in Pennsylvania, are likely affected differently by news about fracking in Pennsylvania than residents of non-fracking Adams County, Pennsylvania. However non-fracking origin counties in states with fracking are still included.[22]

As $\textit{newspaper exposure}_{oSt}$ is a weighted sum, it is not immediate how to interpret a one unit increase in this measure. From equation (3.9), if the pre-2009 county circulation rate is one (meaning every household receives the newspaper) an additional news article will increase

[20] The actual correlation between national and local newspaper exposure is 0.12.

[21] Many local news sources provide free access to content online, which is not captured by this measure of local news exposure. National and regional news sources often provided limited free access, but ultimately require a paid subscription. The AAM circulation data includes digital replica newspapers, but not necessarily individual browsing behavior. To the extent that online exposure is positively correlated with print exposure, the estimates will simply represent the response to total news exposure (where print exposure is used as a proxy).

[22] As noted earlier, origin counties in the New York City DMA are also excluded. Relaxing these sample restrictions do not significantly impact the results (see Appendix Table A3.5).

newspaper exposure by one unit. In reality, newspaper circulation rates are significantly lower than one hundred percent. To make exposure more readily interpretable, I divide $newspaper\ exposure_{oSt}$ by 0.05, such that a one unit increase is equivalent to one additional news article in a newspaper with a five percent circulation rate.[23] This level of circulation is comparable to a county with high readership of *USA TODAY*.[24] Conveniently, when using this scaling average news exposure among treated observations is 0.99, suggesting a one unit increase also approximates the mean effect.

3.5 Main Results

3.5.1 Graphical Analysis

Before estimating the regression in equation (3.10), I explore pre-trends and present event study graphical evidence of the impact of national newspaper exposure on migration. This specification can verify that origin counties that will eventually be highly exposed to news do not have higher migration flows in years prior to actual exposure to this news, relative to origins that will be less exposed. There are various ways to measure this exposure "treatment", but I will focus on differences in exposure due to initial differences in circulation rates of national newspapers that will eventually report on fracking.[25] This tests to see if origin/destination specific news exposure is correlated with other unobserved characteristics that evolve over time and affect migration (ν_{oSt}). For each origin county I collapse the pre-fracking circulation rates of the *USA TODAY*, *New York Times*, and *Wall Street Journal* to a single weighted average, where the weights are the share of the total national news articles about fracking in destination S in each newspaper. This measure captures the extent to which an origin will eventually be exposed to news about fracking in the destination state. I interact this measure with year indicators between 2001 and 2012, and then

[23]This scaling is similar to a continuous versions of the persuasion rate (DellaVigna & Kaplan, 2007). The effect of one additional news article is scaled by the exposed population and in this case, the population available to persuade is approximately one. However, as I do not observe individual exposure, I cannot construct a direct comparison.

[24]For reference, *USA TODAY* circulation ranges from 0 to 27.8 percent, with a mean of 1.2 percent; *New York Times* circulation ranges from 0 to 3.3 percent, with a mean of 0.51 percent; and the *Wall Street Journal* circulation ranges from 0 to 6.4 percent, with a mean of 1.2 percent.

[25]This measure is ideal for testing that different levels of initial circulation do not follow differential trends. The figure is almost identical when looking at alternative measures of treatment, such as the total newspaper exposure summed over all years.

regress the inverse hyperbolic sine of the number of migrants on this set of interactions to trace out the impact of pre-fracking circulation on migration over time. I include origin-destination pair fixed effects as well as destination-by-year fixed effects to exploit the same variation used in the main analysis. The coefficients on these year interactions are interpreted as the marginal effect of a one percentage point increase in the pre-fracking circulation rate on migration flows in that given year, and are plotted with 95 percent confidence intervals in Figure 3.3. For reference, a bar graph of the average number of articles about fracking in a specific state is superimposed, to show when news content about fracking was published.

Between 2000 and 2007, specific destination states were only mentioned in the one 2006 *New York Times* article, otherwise, there were no national newspaper articles referencing fracking destinations. There are small increases in the number of articles about fracking between 2008 and 2010, with a large jump in 2011 and 2012. Prior to 2010, migration fluctuates around zero, with only one statistically significant, negative estimate in 2003. Starting in 2006 there is a slight, statistically insignificant increase, but overall it appears that origins that would eventually be highly exposed to news about fracking followed similar trends in migration. In 2010 the effect on migration becomes significant, and discontinuously jumps in 2011, when news content increased dramatically. The data suggest that a one percentage point increase in the pre-fracking circulation rate did not increase migration prior to news exposure, but was associated with a 2.5 percent increase in migration in 2011 and 2012, precisely when there was intense news coverage of fracking.[26]

3.5.2 Impact of Newspaper Exposure on Migration

I now formalize this relationship by estimating the regression in equation (3.10). These estimates are provided in Column (1) of Table 3.2. Given the absence of news in early years, I interpret effects as changes from zero to one. For an origin county with a five percent circulation rate, one additional newspaper article about fracking in a specific state increased migration flows

[26] As seen in Appendix Figure A3.2, commuting responds similarly, although the increase is larger (8-12 percent) and begins earlier in 2009.

to fracking counties in that state by 2.4 percent on average $(0.025 * 1 - 0.001 * 1)$.[27] As average news exposure is also approximately one, this would suggest the mean effect of news exposure on migration was 2.4 percent as well. The most news was published in 2012 (average news exposure was 1.8) suggesting that in 2012, news about fracking increased migration flows to fracking counties by 4.2 percent on average.

I next adjust the baseline specification as outlined above to determine if changing characteristics of the origin bias the estimates. In Column (2) I include the annual origin county-level labor market measures, and the estimated impact of national newspaper exposure is virtually unchanged at 2.4 percent. In Column (3) I include origin by year fixed effects. This absorbs the labor market measures included in Column (2) as well as any other unobserved characteristic of the origin that is changing over time and affects migration behavior. In this specification the effect of one additional newspaper article is 2.5 percent, and not statistically different from the baseline estimates. Finally, in Column (4) I include the origin by year fixed effects and control for local newspaper exposure. The effect of one additional national newspaper article remains 2.4 percent. For completeness, I repeat the same estimation using the number of migrants in levels as the outcome. In each of these specifications the marginal impact ranges from 1.4 to 1.7 and is not statistically distinguishable.[28] For the remainder of the paper, I will estimate the model corresponding to Column (2), which includes controls for labor market conditions at the origin, although the results are not sensitive to this choice of specification.

The data suggest that increased national news coverage of fracking increased migration to the fracking counties in states publicized. Although these estimated impacts are small, they are both statistically and economically significant. These estimates imply that news about fracking increased migration flows to fracking counties by 2.4 percent on average, and that the 36 articles in the *New York Times* in 2011 that discussed fracking in Pennsylvania led to an 8.2 percent

[27] These estimates are not just statistically significant due to a large sample. As seen later, the significance remains when estimated over much smaller subsamples.

[28] An increase of 1.4 migrants represents a much larger effect at the mean than captured by the inverse hyperbolic sine specification. This appears to be driven by origin counties with large migrant flows. If the sample is restricted to origin/destination pairs with non-zero flows, the two specifications yield similar percent effects at the mean.

average increase in migration flows to Pennsylvania fracking counties. This would suggest that, at the margin, relaxing information constraints and providing information about potentially lucrative labor market opportunities elsewhere will increase migration to those destinations.[29]

3.5.3 Impact of Newspaper Exposure on Cross-County Commuting

To avoid the monetary, psychic, and amenity costs that might accompany a move, an individual can choose to commute rather than migrate. In a companion paper, I show that many people took advantage of the earnings gains associated with fracking by taking up jobs in fracking counties and commuting, rather than migrating (Wilson, 2017). Information revealed through newspaper exposure might also affect aggregate behavior at this margin. In Table 3.3 I report the impact of newspaper exposure on the total number of workers who live in one county but work in a fracking county in the state mentioned in the newspaper article. In addition, I report the number of jobs for three pre-defined age groups: workers under 30, workers 30 to 54, and workers 55 and older. These data are obtained from the LEHD Origin-Destination Employment Statistics (LODES). For an origin county with a five percent circulation rate, one additional news article about fracking in a specific state increased the number of workers commuting to fracking counties in that state by approximately 6.6 percent. The impact on commuting is nearly three times as large as the migration response, which is not surprising as commuters avoid many of the fixed costs associated with moving.

When looking across age groups, the response for one additional news article for 30 to 54 year olds is 5.2 percent and statistically larger than the response of both younger workers (3.1 percent) and older workers (3.6 percent).[30] These differential responses among workers under 30

[29]Unfortunately the IRS migration data does not provide demographic information. In theory, the microdata from the American Community Survey could be used (Ruggles et al., 2015). However, because the annual ACS is only a one percent sample of households, the probability of observing a mover to a specific state is very low, leading to measurement error and attenuation. Often these flows are constructed from less than five people. Using the 2005 through 2011 ACS, I have estimated the corresponding models at the PUMA level by gender, age, race, education, and family status. Consistent with attenuation bias due to measurement error, the estimates are about one tenth the size and insignificant. The largest group effects are for men (0.23 percent, p=0.13), 18-34 year olds (0.16 percent, p=0.19), non-Hispanic other (0.16 percent, p=0.15), and college graduates (0.15 percent, p=0.16).

[30]The reader will notice that the percentage effect is larger for all workers than for any of the

and between 30 and 54 also correspond to patterns of newspaper readership. Throughout the 2000s, 30 to 49 year olds were more likely to report the newspaper as a main source for national news than 18 to 29 year olds (Pew Research, 2013b). In Appendix Table A3.4, I also report differences by job level earnings and broad industry. Consistent with people commuting for fracking jobs that pay well, workers in higher paying jobs are more responsive, but there is also an increase in commuting among low paying jobs. Non-production or trade workers in the "other" industry, which includes oil and gas extraction, are the most responsive. However, workers in goods production and trade and transportation also significantly respond, suggesting that increased news exposure not only induced people to commute for oil and gas extraction, but for other jobs that were also affected by the labor market shock.

The effects of newspaper exposure on both migration and commuting are robust to different ordered polynomials of newspaper exposure (see Appendix Table A3.4) and insensitive to sample restrictions (Appendix Table A3.5).[31] The estimated migration relationship is also robust when I account for censoring in the IRS migration data (Appendix Table A3.6).[32]

3.5.4 News about Fracking in Another State

News about fracking in a specific state can affect migration decisions by providing general information about the labor market shifts associated with fracking, or by providing specific information about where these labor market shifts are occurring.[33] To separate these channels of

three subgroups. This is in part because the pooled specification constrains the controls and fixed effects to be the same for each group. If this specification is run in levels, the effect for all workers is the sum of the effects for each subgroup, as expected.

[31] There are only two years in the end of the sample with high levels of news exposure, making dynamic effects less relevant. In specifications including one and two year lagged exposure, the effect of concurrent newspaper exposure on migration is nearly identical and the first lagged effect is small and marginally significant. For commuting the effect of concurrent exposure is slightly smaller, but not statistically different, and the lags have small, significant effects.

[32] The effect of newspaper exposure is similar in the extreme lower bound case where all zero county-to-county flows are changed to 9 (the highest possible censored value). Newspaper exposure increases the probability of not being censored by over 15 percent (0.005/0.03) at the mean, as well as the number of migrants for the severely limited subset of origin destination pairs that always report positive flows.

[33] This is related to the advertising literature, suggesting information could either lead to more migration overall (expansion) or shift people from other destinations (share stealing). In Appendix Table A3.7 I include the total exposure to news about any of the 16 destination states within an

effects, I evaluate how migration flows to fracking counties in a particular destination state respond to news about fracking in a different state. For example, observing that the migration response to news about fracking in a different state is smaller than the response to news about fracking in the destination state, would suggest the news provides some locational signal.

In practice, I randomly assign all observations indexed by S the fracking news exposure of S', one of the other 15 fracking states. For example, all observations for the destination Arkansas might be randomly assigned the news exposure of North Dakota, while the observations of North Dakota might be randomly assigned the news exposure of Pennsylvania. I then estimate the regression similar to equation (3.10), but replace $\textit{News Exposure}_{oSt}$ with the randomly assigned $\textit{News Exposure}_{oS't}$, and calculate the marginal impact of a one unit increase in $\textit{News Exposure}_{oS't}$ (i.e., one news article in a county with a five percent circulation rate). I repeat this regression 200 times to plot the distribution of potential impacts. This histogram of potential impacts is plotted in Panel A of Figure 3.4 and the estimated effect using actual news exposure from Table 3.2 is indicated for reference.

The estimated effect from Table 3.2 is larger than the estimates from all but 5 of the repetitions (2.5 percent), suggesting that at least part of the migration response is due to locational signaling. The distribution of effects using randomly assigned news exposure is centered around 0.013, suggesting news about fracking in a different state has some positive predictive power. However, this 1.3 percent effect cannot be strictly attributed to general information about fracking. National newspapers report about multiple destinations, and news exposure is positively correlated across potential destination states. Among the 200 regressions, the average correlation coefficient between actual news exposure and randomly assigned news exposure was 0.44. To some degree,

origin and year, to determine if news about fracking in general has an effect. This has no effect on migration, but a small, significant effect on commuting (0.5 percent). I also estimate a separate specification including the news exposure for the state that received the highest exposure within an origin and year. This effect is positive and areas with the most news did not reduce flows to other fracking destinations, suggesting news exposure led to expansion, rather than shifting. See for example, Garthwaite (2014) examining book sales after celebrity endorsements. It is also possible that migrants to fracking areas simply shifted from moving to other, non-fracking areas. However, when looking at the impact of total news exposure about fracking on migration to non-fracking areas, the coefficient is small, but positive.

randomly assigned news exposure will proxy for actual news exposure, which might drive the estimated 1.3 percent effect.

For this reason I adjust the regression specification to include origin county by year fixed effects, and repeat the process 200 times. This specification absorbs average changes in news exposure at the origin county level and exploits destination specific deviations from the origin average. This specification looks to see if, for example, an origin that had unusually high exposure to news about fracking in North Dakota saw larger increases in migration to fracking counties in Arkansas. Randomly assigned news exposure no longer proxies for actual news exposure, but provides a placebo test to determine if destination specific fluctuations in news exposure impact the corresponding migration flows. If the effect size from these placebo regressions were comparable to the estimates in Table 3.2, then we would be concerned that the results are driven by things correlated with news exposure and pre-fracking circulation rates in general, not a causal effect of news content. This histogram of potential impacts is plotted in Panel B of Figure 3.4, along with the estimated effect from the origin county by year fixed effects specification from Table 3.2. The distribution of effects are centered around zero and are all smaller than the estimated effect in Table 3.2, suggesting that destination specific deviations in news exposure only affect migration flows to the destination that is being mentioned. The information about fracking conveyed in the news affects migration decisions primarily by providing information about where fracking is occurring.

3.6 Alternative Strategy: Newspaper Market Border Comparison

Although origin level characteristics and local news exposure do not appear to introduce bias into the estimates in Table 3.2, it is possible that news exposure is correlated with other unobserved characteristics captured by ν_{oSt}. For example, counties that have higher circulation of national news, and thus higher news exposure might, for unobserved reasons, also have a higher propensity to migrate to fracking areas when a boom hits. This could happen if, for example, origin counties more tied to the oil and gas industry also had higher readership of national newspapers, and thus higher exposure when these booms happened. To address this potential concern, I employ an alternative strategy that exploits variation in news exposure among neighboring counties. Using all domestic newspapers that had at least one article about fracking between 1999 and 2012 and had circulation data available, I construct newspaper geographic markets. This is the set of counties in the newspaper's distribution network. For many local newspapers, distribution costs inhibit broad distribution and these markets are composed of a small group of adjacent counties around a central hub.[34] I then identify counties on the border of this distribution network as well as contiguous counties that do not receive the newspaper and compare the effect of news articles, specific to that newspaper, on migration and commuting for counties on both side of the market border. This is done in a stacked regression as follows

$$\begin{aligned} Y_{oSt} = \gamma_1 Articles_{nSt} * InMarket_{on} + \gamma_2 Articles^2_{nSt} * InMarket_{on} \\ + \gamma_3 InMarket_{on} + X'_{ot}\Gamma + \phi_{oS} + \psi_{nSt} + \varepsilon_{oSt}. \end{aligned} \tag{3.11}$$

The outcome in equation (3.11) is the same as in previous specifications. The variable *Articles* is the number of articles in newspaper n published in year t about fracking in state S (in units of ten), while *InMarket* is an indicator variable that equals one if the origin county o is in the market for newspaper n. To be specific n uniquely identifies each newspaper and the corresponding market border. So counties that do not receive newspaper n but are on the market border will also be assigned to n. Time-varying origin controls and an origin-destination pair fixed effect are included. A newspaper-by-destination state-by-year fixed effect is also included, making this a comparison of flows to the same destination among counties along the same market border. The identifying assumption is that counties on either side of the newspaper's market border would evolve similarly, but for the news coverage about fracking. Because counties are being compared to other local counties, similar preferences and propensities among these neighboring counties captured in ν_{oSt} will be differenced out. This will identify the causal effect of news coverage as long as propensities to migrate to fracking during booms is local, but not county specific.

These results are reported in Table 3.4. Relative to no articles, ten news articles signifi-

[34] Over 90 percent of these newspapers distribute to 40 counties or less.

cantly increased migration by 5.6 percent in counties that received the newspaper, relative to their neighbors. There was also a similar significant effect on cross-county commuting (4.9 percent). In this sample, the average circulation rate among in-market border counties was slightly higher than the benchmark five percent at 5.5 percent, making it easy to compare the magnitude of this effect to the estimates from the previous specification. In the average in-market county with a newspaper circulation rate slightly higher than five percent, one additional local news article increases migration by approximately 0.6 percent.[35]

3.7 Additional Explorations

3.7.1 Impact of TV News Exposure

In a 2013 Pew Research report, 69 percent of adults cite television as one of their main sources for news. This rate has only slightly fallen from 74 percent in 2001. Meanwhile, the share of adults citing newspapers as a main source of news has fallen from 45 percent to 28 percent. Also during this time period, the internet has become an increasingly important source of news going from 13 in 2001 to 50 percent in 2013. Data constraints prevent me from comparing internet news exposure to traditional news sources, but I am able to compare migration and commute responses to television and newspaper news exposure.

Using abbreviated news transcripts from the three major TV news networks (ABC, CBS, and NBC) available through Vanderbilt Television News Archive (VTNA), I construct a measure of TV news exposure similar to the measure of national newspaper news as follows

$$TV\ news\ exposure_{oSt} = \sum_{c \in C} \big(\text{broadcasts on } c \text{ about fracking in } S\big)_t * pre09\ view\ rate_{oc}. \quad (3.12)$$

The set $C = \{ABC, CBS, NBC\}$ and captures TV news coverage from the major national news networks. As with *newspaper exposure*$_{oSt}$, *TV news exposure*$_{oSt}$ captures variation in

[35] The point estimates are similar if I exclude national newspapers (as these borders are less local). Counties might appear multiple times in the stacked regression in equation (3.11). However, if I restrict the sample to only include one newspaper market border per county to ensure that counties only appear once, the pattern is similar. To do this I take the set of newspaper market borders each county belongs to, and restrict the sample to only include the newspaper market border that had the highest number of articles about fracking among these market borders.

national news, which is weighted by the channel's pre-fracking Nielsen's viewership rates obtained through the 2008 Television and Cable Factbook. During this period, TV stations were transitioning from analog to digital broadcasts on a market-by-market basis, and new digital equipment was needed to receive the transmission. This might introduce bias in the viewership rates as only some markets had transitioned by the time data was collected. For this reason, I also run specifications using ratings from the latest 2016 Factbook, after these updates were fully made.[36] Nielsen ratings are only available at the DMA-level, which is a mutually exclusive set of similar, nearby counties that represent a media market. To conduct this analysis, I aggregate all data up from the county to the DMA-level, including migration flows, newspaper circulation and exposure, and labor market measures. Typical viewership of ABC, CBS, and NBC was approximately 50 percent during this time period, so I scale TV news exposure such that a one unit increase represents the effect of one additional news broadcast from a TV network with 50 percent viewership.

This measure of TV exposure does not capture news exposure through cable news channels, such as CNN or Fox News. The VTNA only collects one hour of news broadcast data from these channels, and cable circulation is measured differently than traditional TV. If $TV newsexposure_{oSt}$ is negatively correlated with cable news exposure (i.e., if network and cable news are substitutes) and both sources of news lead to more migration, than these estimates will be biased downward. If instead network and cable news are complements, network news could be interpreted as a proxy for total TV news.

DMA-level estimates are presented in Table 3.5. I first report the effects of newspaper exposure on migration as the level of analysis has changed. One additional news article in a DMA with a five percent circulation rate increased migration to the fracking state mentioned by 5.0 percent. This point estimate is twice as large as the county-level estimate, but is not statistically different. Column (2) reports the estimated effects for TV news exposure. Using 2008 viewership rates, TV news exposure had no impact on migration behavior. In Column (3) I regress migration

[36]Using 2016 ratings potentially introduces endogeneity if viewership is responding to migration and commute behavior. However, circulation rates are highly persistent, suggesting this bias might be small.

on both newspaper exposure and TV news exposure, to determine which news source is more closely associated with migration.[37] The coefficients for both news sources remain similar and the effect for newspaper exposure is significant. In Column (4) I conduct the same analysis, but use a measure of TV news exposure using 2016 viewership rates, as all markets had fully transitioned to digital TV by this time. In this specification the effect of newspaper circulation is 4.9 percent, while the effect of TV news exposure jumps to 7.9 percent (the p-value on the first order effect is 0.11). If only TV news exposure is included, the magnitude of the effect is similar and significant, suggesting TV news exposure also affects migration, but the relationship is much weaker than for newspaper news.

When looking at commute behavior, the DMA-level point estimate on newspaper exposure is smaller than the county-level estimate in Table 3, but not statistically different.[38] The effect of TV exposure is large and significant. Relative to zero TV news exposure, one news broadcast in a DMA with a 50 percent circulation rate increased the number of people commuting to the fracking area discussed in the news report by 10.6 percent. The point estimates are largely unchanged when both news sources are included or when using 2008 or 2016 TV circulation rates. Many more people cite TV as a news source (Pew Research, 2013b), and circulation rates are an order of magnitude higher. However, there are far fewer TV broadcasts about fracking than newspaper articles, and these news clips are only 1-5 minutes. Observing similar impacts for both newspaper and TV exposure would suggest that when providing information about potential labor market opportunities, both the intensity of content and penetration influence the magnitude of the effect.

3.7.2 Positive versus Negative News

Many newspaper articles that mentioned specific states also referenced positive characteristics such as jobs, booms, or growth. However, there were also many articles that discussed negative aspects such as pollution, health, dangers, and earthquakes (see Appendix Table A3.1).

[37]Newspaper exposure and TV news exposure are moderately, positively correlated ($\rho = 0.36$).

[38]When aggregating to the DMA-level, many neighboring counties fall into "fracking" DMAs that are excluded from the sample of origin DMAs. This likely attenuates the estimated impact on commuting.

Although not necessary, it is possible that individuals are more responsive to positive news than negative news. It could also be the case that if people have preconceived beliefs about the value of fracking, and are only uncertain about where it is occurring, people might move even in response to "bad news". I parse the text of each news article to determine how many times positive and negative aspects are mentioned. I then classify an article as positive if it at least two positive mentions and has more positive mentions than negative. Negative articles are similarly defined. I then estimate the separate effect of positive and negative newspaper exposure on migration and cross-county commuting in Table 3.6.[39] Relative to no newspaper exposure in a county with a five percent circulation rate, one positive article significantly increased migration by 4.0 percent. This effect is one and a half times as large as the effect of one negative article, and significantly different. The effect of negative news is positive and significant, suggesting that "any news is good news", but people are more responsive to positive news. As changing origin characteristics that affect migration might be correlated with either positive or negative news rather than total news, I include origin county by year fixed effects and the effects are not statistically different.

In contrast to migration, positive and negative newspaper exposure affect commuting similarly. Relative to no exposure, one additional news article, positive or negative, leads to 7-9 percent higher cross-county commute flows. Including origin county by year fixed effects reduces the size of these effects but they remain significant and statistically indistinguishable. Unlike migrants, long distance commuters do not bear some of the drawbacks associated with fracking. such as potential home water contamination, earthquakes, or noise on residential streets. Negative news might provide a location signal, in addition to informing potential migrants about amenity costs that might be associated with moving; workers looking to commute might only value the location signal provided by the news.

In addition to exploring differences by source and content, I have also explored differences by newspaper and distance. The readership of the *New York Times* and *Wall Street Journal* is on average more-educated, higher income, and older than the typical migrant to fracking areas

[39]Exposure to neutral articles with less than two positive and two negative keywords are not included in this regression. Specifications including all three levels are similar but less precise.

(Wilson, 2017). In Appendix Table A3.8, I show that the estimated effects on migration and commuting are largest and most significant for news in the *USA TODAY*, with smaller effects from the *New York Times*, and very imprecise, insignificant effects from the *Wall Street Journal*. To explore difference across distance, I estimate equation (3.10) for origin county by destination state pairs in one hundred mile bins and plot the total marginal effect of news exposure for each distance in Appendix Figure A3.3. The effect climbs to about 6 percent between 400 and 1,000 miles and then gradually falls, consistent with information provision having an effect because people are aware of nearby opportunities, but lack information about distant potential opportunities.[40]

3.7.3 The Role of Origin County Labor Market Conditions

In recent years there has been considerable concern about decreasing labor market fluidity and mobility, especially when it appears that people in weak labor markets could encounter more abundant opportunities elsewhere (Molloy et al., 2016).[41] Previous evidence from the MTO experiment and educational differences in migration are consistent with information constraints playing a role (Malamud & Wozniak, 2012; Chetty, et al., 2016). I next explore heterogeneity in the migration response to newspaper exposure by the labor market strength of the origin to understand if providing information is particularly impactful for people in weak economic areas. To do this, I estimate a variant of my main specification as follows

$$\begin{aligned}
Y_{oSt} &= \beta_1 newspaper\ exposure_{oSt} + \beta_2 newspaper\ exposure^2_{oSt} \\
&\quad + \beta_3 newspaper\ exposure_{oSt} * emp/pop_{ot-1} + \beta_4 newspaper\ exposure_{oSt} * emp/pop^2_{ot-1} \\
&\quad + \beta_5 newspaper\ exposure^2_{oSt} * emp/pop_{ot-1} + \beta_6 newspaper\ exposure^2_{oSt} * emp/pop^2_{ot-1} \\
&\quad + \beta_7 emp/pop_{ot-1} + \beta_8 emp/pop^2_{ot-1} + X'_{ot}\Gamma + \phi_{oS} + \psi_{St} + \varepsilon_{oSt}. \qquad (3.13)
\end{aligned}$$

Where *emp/pop* is the lagged county employment to population ratio, for the adult population. I then calculate the total effect of one unit of newspaper exposure, which is allowed to vary quadratically with the employment to population ratio and use the delta method to obtain standard

[40]The effects of news on commuting for different age groups is similar (Appendix Figure A3.4).
[41]This topic has also come up in the popular press (Brooks, 2016; Cohen, 2016).

errors (the corresponding coefficients are provided in Appendix Table A3.9). The effects on both migration and commuting are plotted in percentage points in Figure 3.5 for county employment to population ratios between 60 and 85 percent (approximately the 15th to 90th percentile). Both the migration and commute responses are larger for weaker economic areas. A one unit increase in newspaper exposure led to a 2.8 percent increase in migration from counties with a low employment to population ratio, but had a small, one percent impact on migration from counties with a high employment to population ratio. Low employment counties saw commute flows increase by nearly 8 percent for an additional news article, while counties with high employment saw increases closer to 2 percent. Exposure to news about fracking in distant, potential labor markets had a larger impact on migration flows from economically weak areas, suggesting informational constraints might be a contributing factor to differences in migration behavior.[42]

Using this specification, I estimate the implied impact of various news exposure treatments at different points in the origin employment to population ratio distribution as reported in Table 3.7. These simulations highlight the heterogeneous impacts of different potential policy interventions. Even increasing the exposure level by one unit (one additional article in a county with a five percent readership rate) has substantial impacts across the distribution. The impacts are the largest for counties with employment to population ratios below the mean of 70.9 percent (effects between 2.7-2.8 percent), and monotonically decrease as the local labor market conditions improve. As exposure increases to five or ten the impacts become very large for counties with weak labor markets and the differential impact becomes substantial. Increasing exposure to ten increases migration flows from counties with weaker than average labor markets by 19-21 percent, 20.5 percent for counties near the mean, and 11.6 percent for counties with very strong labor markets. Once exposure reaches 15 (close to the top one percent of exposure in 2012) effect sizes start to plateau at most employment to population levels, suggesting the additional benefits of exposure at this level are small. For reference the average actual exposure level in 2012, along with the implied impact of this level of exposure are also reported. Average actual exposure is between 1.5 and 2.5

[42]If I use the unemployment rate rather than the employment to population ratio the pattern is more flat, but still downward sloping.

for all levels of labor market strength, although it does peak slightly around the mean. For origins near the mean employment to population level, migration flows would have been 6.4 percent lower if there had been no news about these local fracking booms.

From equation (3.13), I solve for the value of exposure that maximizes the impact for each point in the employment to population distribution. For counties with employment to population ratios below the average at 65 or 70 percent, the maximizing level of exposure is very high, at 18.4 or 19.6. This level decreases, falling to only 15.6 for counties with employment at 80 percent. The maximum implied impacts also vary greatly, going from 27 percent in areas below the average to only 13.3 percent at the top. The patterns are similar when looking at commute behavior although the impacts are larger and more heterogeneous, while the maximizing level are more uniform (see Appendix Table A3.10).

Heterogeneous impacts by origin labor market strength could result from differential exposure to new or heterogeneous returns to the information. As differences in actual exposure are small it is likely that only a small part of the heterogeneous impacts can be explained by differential exposure. Exposure at all levels of labor market strength is substantially lower than the maximizing level, suggesting all regions face limited information. However, the information is most valuable for people living in counties with weak labor markets, where the expected gains to moving are largest.

This has several implications for potential policy interventions. Information provision policies could increase geographic mobility, potentially resulting in more beneficial labor market transitions (Molloy et al., 2016) and higher economic mobility (Chetty & Hendren, 2016). Providing a modest amount of information about potential labor market opportunities in other parts of the country to all counties will significantly increase migration to those regions. However, a government facing limited resources would see the largest returns by focusing on providing information to weak labor markets. Not only would the migrant benefit, by encountering more favorable labor markets, but this might also generate positive externalities for workers in the weak origin labor market, as the market becomes less slack.

3.8 Online Searches, a Potential Mechanism

The data indicate that when counties are more exposed to news coverage about fracking in a certain state, migration and commute flows to fracking areas in that state increase. This relationship posits that news coverage provides information about potential labor market opportunities, and affects migration through changing expectations and uncertainty. This channel, however, cannot be directly tested in the data. Rather than verify that people's expectations change, I am able to quantify how interest in fracking and the states mentioned changes after news is disseminated. Using Google Trends data, I next explore search interest before and after TV news broadcasts about fracking. For a specified search term (i.e., "fracking"), Google Trends will provide a time-series of search intensity at the national, state, or DMA level. This time-series is an ordinal measure of intensity that equals 100 on the day with the highest number of searches per capita, and with every other day scaled as a percent of the maximum. For example, on a day that is assigned a value of 20, search intensity for the search term was only 20 percent the level from the maximum day. This measure facilitates comparisons within a region over time, but is not conducive to studying differences across both geography and time. As such, I will examine changes in search behavior a short period before and after a TV news broadcast, but cannot reliably determine if search intensity increased by more in areas with higher TV viewership rates.

For each of the 17 TV news broadcasts that mention "fracking" or "shale", I pull daily time-series for every DMA in the United States for 15 days before the broadcast and 14 days after for several search terms.[43] First I look at search intensity for the term "fracking", and then I look at search intensity for the name of any states that are mentioned in the broadcast. States are only mentioned in 14 of the broadcasts. To identify the impact of the news broadcast on average search intensity I estimate the following regression

$$search\,index_{opt} = \sum_{\tau=-14}^{14} \delta_\tau * \mathbf{1}\{t \text{ is } \tau \text{ days from broadcast}\}_{op} + X_t'\Gamma + \phi_{op} + DOW_t + \varepsilon_{opt} \quad (3.14)$$

where $search\,index_{opt}$ is the search index on date t in DMA o relative to the search period

[43] A special thanks to Tanner Eastmond for help working through the Python code.

p. The search period is the 15 days prior and the 14 days after each broadcast, such that *op* uniquely identifies each DMA/term pair, over which the relative search index is measured. The set of coefficients δ_τ trace out the daily deviations in the search index from the omitted day ($\tau = -15$). I include a DMA by search period fixed effect in order to compare days from the same search that have comparable indices. Day of the week fixed effects are also included to account for differences in search behavior during different times in the week.

Several of the news reports were broadcast in close proximity to other high publicity events connected to either fracking or the states mentioned in the reports. For example, On January 24, 2012, four days prior to a news report about fracking in Pennsylvania, President Barack Obama discussed shale gas extraction and fracking in the State of the Union. Similarly, late on December 31, 2011, there was an earthquake in Youngstown, Ohio, that many linked to fracking. Four days later there was a news report on fracking in Ohio. There are also other high publicity state specific events (such as college football bowl games or school shootings) that occur during some of the search period windows. When looking at searches for fracking I include indicator controls for the three days after the two events related to fracking, and when looking at searches for specific state names I include a set of indicator controls for the local, high publicity events that are listed in Appendix Table A3.11. If I do not control for these events the series becomes more volatile, but there is still a significant spike directly after the broadcast. Figures that do not control for other high publicity events are included in Appendix Figure A3.5.[44]

These effects are plotted in Figure 3.6 for "fracking" and Figure 3.7 for the state names. As seen in Figure 3.6, search intensity for "fracking" spikes the day of the broadcast and remains elevated for the next two days before falling back to the previous levels. Across all DMAs, search intensity for "fracking" jumps by nearly 2.5 points on average. Because the search index is a relative measure, this cannot be converted to back out how many additional searches were made. If I combine days into 3 day bins, for statistical power, I estimate a 2.5 point spike in days 0

[44]In theory I would like to look at search behavior after newspaper publications as well. However, national news articles about fracking were published quite frequently, leading to large overlap in sample windows.

to 2, followed by an statistically significant one point increase for the remainder of the days in the sample, suggesting that search interest remained elevated for some time (see Appendix Figure A3.6).

When looking at search interest in the names of states that were mentioned in the broadcast there is also a spike one day after the broadcast and interest remains elevated for the next five days. Although not directly comparable, the jump in search intensity is larger at nearly 6.5 points. For reference I also regress the search intensity for the same set of fracking states that are not mentioned in the news broadcasts. These estimates remain close to zero, with no spike or increase after the news broadcast, suggesting this is not capturing overall interest in fracking states. The Google Trends data indicate that following a TV news broadcast about fracking, people search more for fracking and for the states mentioned in the broadcast. Although this is not direct evidence that news coverage motivates people to move, it does suggest news coverage induces people to seek more information about the potential fracking destination, which might be due to an interest in moving or a desire to obtain more information.[45]

3.9 Conclusion

Migration is an investment that can improve the types of labor markets individuals encounter, but many of those that appear to face the largest benefit do not move. This could be the result of optimal behavior, but could also be due to various constraints or market frictions, such as limited information. In this paper I evaluate the role of information in the decision to move to more favorable labor market opportunities. The current literature speaks very little to the effect of labor market information on migration behavior. I outline a conceptual framework for understanding how information will affect the migration decision and potentially change migration outcomes.

To estimate the effect of news on migration I exploit information disseminated through

[45] I have also looked at search interest in moving specific terms such as "Uhaul" or "Uhaul rental". At both the DMA and state-level there appears to be a visual shift at the time of broadcast, however it is not statistically different. Search intensity for terms like "fracking jobs" or "oil jobs" are low and frequently suppressed by Google. These patterns are similar if the window is extended.

national news coverage of localized fracking booms. The technological and geological constraints associated with fracking have led to sudden, large labor market shocks in well-defined areas. The novelty of fracking also makes it straightforward to identify news coverage about fracking across different affected areas. I combine national news content with historic local circulation rates to construct a measure of news exposure that strips away endogenous changes in consumer readership and endogenous changes in producer content decisions.

The data suggest that for a county with a five percent circulation rate and no previous exposure, one news article about fracking in a specific state increased migration flows to fracking counties in that state by 2.4 percent. This estimate accounts for destination specific characteristics that are changing over time and does not change when controlling for origin-level shocks or local news exposure. Cross-county commute flows increase by 6.6 percent.

I also provide evidence that TV news exposure has an effect on commuting and potentially migration also. The magnitude of these effects are similar to the response to newspaper news. Migration flows are more responsive to exposure to positive news than negative news, though both lead to more migration. In contrast, commute flows respond similarly to positive and negative news, consistent with commuters not facing many of the negative costs associated with fracking at their homes (e.g., water contamination, increased risk of earthquakes). As further evidence that news coverage increases interest in these fracking destinations, I find that, directly after a TV news broadcast about fracking, Google search interest in both the term "fracking" and the names of states mentioned significantly increases. News exposure induces people to seek more information on the internet, which might influence the migration decision.

Importantly, the migration response is largest from origin counties that have been experiencing weak labor market conditions, suggesting the benefit to news provision is largest in those areas. This has potential implications for people trying to understand why less-educated and low-income households in poor performing labor markets are unlikely to move, and if there are policies that can encourage more migration to better economic opportunity. Simulations suggest that providing more information about potential labor market opportunities in other areas would increase geographic mobility in all areas, with the most pronounced response in weak labor markets where the returns to migration are the largest.

Tables

Table 3.1: County Characteristics by *USA TODAY* Pre-Fracking Circulation Rate

	County Characteristics in 2000		Change from 2000 to 2010		Predicted Difference from 25th to 75th Percentile
	Below Median	Above Median	Below Median	Above Median	
	(1)	(2)	(3)	(4)	(5)
Migrants to fracking areas (Percent of Population)	0.09	0.11	-0.01	-0.02	0.00
Employment to Population (16+)	55.68	59.14	-0.78	-1.43	-0.10
Unemployment Rate	3.29	3.44	0.96	1.33	-0.09**
Median Household Income	31,805	38,834	8,485	8,763	290***
Percent in Poverty	15.82	12.29	0.83	2.18	0.51***
Percent White	85.16	84.32	-0.83	-2.26	-0.58***
Percent Black	8.95	9.27	-0.15	0.42	0.17***
Percent Hispanic	5.87	5.72	1.77	2.38	0.29***
Percent Other Race	5.89	6.41	0.98	1.85	0.41***
Percent Population 20-34	16.67	19.60	-0.55	-0.67	0.02
Percent Population 35-64	38.76	38.56	1.82	1.47	-0.27***
Percent Population Over 64	15.96	13.64	1.20	1.15	0.00
Percent Households Renting	23.47	28.30	1.64	1.81	0.10**
Number of Counties	1,420	1,418	1,420	1,418	2,838

Notes: Migration data from the IRS Statistics of Income. Other county characteristics obtained through American FactFinder from the 2000 Census and 2010 Census and 5-Year American Community Survey. *USA TODAY* circulation data from the Alliance for Audited Media. The county level median pre-2009 circulation rate of the *USA TODAY* was 0.83 percent, and ranged from 0 to 27.8 percent. Median Household Income is reported in current dollars. Column (5) reports the predicted change in the characteristic between 2000 to 2010 when pre-2009 circulation increases from the 25th to the 75th percentile. p<0.01 ***, p<0.05 **, p<0.1 *.

Table 3.2: Impact of Destination State Specific National Newspaper Exposure on Migration to Fracking Counties in State

	Inverse Hyperbolic Sine of the Number of Migrants$_{oSt}$				Number of Migrants$_{oSt}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
National Newspaper Exposure$_{oSt}$	0.025***	0.025***	0.026***	0.025***	1.439***	1.491***	1.673***	1.644***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.419)	(0.437)	(0.534)	(0.532)
National Newspaper Exposure$^2_{oSt}$	-0.001***	-0.001***	-0.001***	-0.001***	-0.036*	-0.038*	-0.036*	-0.038*
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.020)	(0.020)	(0.022)	(0.021)
Local Newspaper Exposure$_{oSt}$				0.009**				-0.671
				(0.004)				(1.662)
Local Newspaper Exposure$^2_{oSt}$				-0.0001**				0.057
				(0.00004)				(0.043)
Origin Labor Market Controls		X				X		
Origin by Year Effects			X	X			X	X
Origin/Destination Local News				X				X
Mean Number of Migrants	7.6	7.6	7.6	7.6	7.6	7.6	7.6	7.6
Observations	590,224	590,224	590,224	590,224	590,224	590,224	590,224	590,224

Notes: Data from the IRS Statistics of Income, LexisNexis newspaper transcripts, and newspaper circulation from the Alliance for Audited Media. The level of observation is the origin county by destination state by year from 2000 to 2012. Origin counties with any fracking production or in the New York City designated market area are excluded. The variable *National Newspaper Exposure$_{oSt}$* is scaled such that a one unit increase represents the impact of one additional news story in a newspaper with a 5 percent circulation rate. For origin/destination pairs with any news exposure, mean national newspaper exposure is 0.99. In all specifications origin/destination pair fixed effects, and destination by year fixed effects are included to control for time-invariant differences across pairs and characteristics of the destination that vary over time. Origin controls include the origin county unemployment rate, employment to population ratio, and average annual earnings (2010\$). Origin/destination specific local news is all destination state specific fracking news content listed in LexisNexis from non-national domestic newspapers. The variable *Local Newspaper Exposure$_{oSt}$* is scaled such that a one unit increase represents the impact of one additional news story in a newspaper with a 40 percent circulation rate, approximately the 95th percentile of pre-fracking circulation among non-national newspapers with articles about fracking. The sample correlation between national news exposure and local news exposure is approximately 0.12. Origin county by year fixed effects control for time-varying characteristics of the origin county and account for potential changes in preferences toward fracking that might be correlated with newspaper readership and affect migration to fracking areas. Standard errors are adjusted for clustering at the designated market area to account for correlation across geography and time. p<0.01 ***, p<0.05 **, p<0.1 *.

Table 3.3: Impact of Destination State Specific Newspaper Exposure on Cross-County Commuting to Fracking Counties in State

	Inverse Hyperbolic Sine of the Number of Cross-County Commute Jobs$_{oSt}$			
		By Age		
	All Jobs	Under 30	30-54	Over 54
	(1)	(2)	(3)	(4)
National Newspaper Exposure$_{oSt}$	0.068***	0.032***	0.053***	0.037***
	(0.009)	(0.005)	(0.007)	(0.005)
National Newspaper Exposure$^2_{oSt}$	-0.002***	-0.001***	-0.001***	-0.001***
	(0.0004)	(0.0002)	(0.0003)	(0.0002)
Dependent Mean	31.4	8.6	18.0	4.9
Observations	499,440	499,440	499,440	499,440

Notes: Data from the LEHD Origin-Destination Employment Statistics (LODES), LexisNexis newspaper transcripts, and newspaper circulation from the Alliance for Audited Media. The level of observation is the origin county by destination state by year from 2002 to 2012. LODES data is only available starting in 2002. Origin counties with any fracking production or in the New York City designated market area are excluded. The variable *National Newspaper Exposure*$_{oSt}$ is scaled such that a one unit increase represents the impact of one additional news story in a newspaper with a 5 percent circulation rate. For comparison, circulation of the *USA TODAY* was 4.5 percent at the 95th percentile. Commuting jobs are also examined by age groups, pre-defined in the LODES data. In all specifications origin/destination pair fixed effects, and destination by year fixed effects are included to control for time-invariant differences across pairs and characteristics of the destination that vary over time. Controls for the origin county unemployment rate, employment to population ratio, and average annual earnings (2010$) are also included. Standard errors are adjusted for clustering at the designated market area to account for correlation across geography and time. The effect of national newspaper exposure on commuting is significantly larger for workers aged 30-54 than the other two age groups. $p<0.01$ ***, $p<0.05$ **, $p<0.1$ *.

Table 3.4: Newspaper Market Cross Border Analysis: Impact of Newspaper Articles on Migration and Commuting

	Inverse Hyperbolic Sine of the Number of Migrants$_{oSt}$			Inverse Hyperbolic Sine of the Number of Cross-County Commute Jobs$_{oSt}$		
	(1)	(2)	(3)	(4)	(5)	(6)
News articles$_{nSt}$ $*$ *In-Market*$_{on}$	0.061***	0.056***	0.069***	0.054**	0.042**	0.082***
(in 10s of Articles)	(0.013)	(0.012)	(0.017)	(0.023)	(0.021)	(0.032)
News articles$^2_{nSt}$ $*$ *In-Market*$_{on}$	-0.005***	-0.004***	-0.005***	-0.005**	-0.004**	-0.007**
(in 10s of Articles)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
In-Market$_{on}$	-0.001***	-0.001***	-0.00004**	-0.001**	-0.001*	-0.0001*
	(0.0002)	(0.0002)	(0.00002)	(0.0004)	(0.0005)	(0.00003)
Exclude National Newspapers		X			X	
Only One Border per Origin County			X			X
Dependent Mean	24.4	24.0	14.1	144.0	141.8	72.2
Observations	1,476,352	1,465,648	509,664	1,250,112	1,240,784	431,360

Notes: Data obtained from the IRS Statistics of Income, LEHD Origin-Destination Employment Statistics (LODES), LexisNexis newspaper transcripts, and newspaper circulation from the Alliance for Audited Media. The level of observation is the origin county by destination state by year from 2000 to 2012 for the migration data and 2002 to 2012 for the commute data. Sample includes all counties on both sides of the border of a newspaper market for any of the 220 newspapers with an article about fracking and circulation data. News articles are newspaper and destination state specific and measured in units of ten. In-market is an indicator that equals one if the county is inside the newspaper's market area (i.e., has positive circulation). In all specifications origin/destination pair fixed effects are included to control for time-invariant differences across pairs. Newspaper market border by destination by year fixed effects are also included to control for characteristics of the local border/destination that vary over time, and make this a comparison of origin counties within the same newspaper market border. Controls for the origin county unemployment rate, employment to population ratio, and average annual earnings (2010$) are also included. Average circulation among in-market counties across all newspapers was 5.5 percent. Columns (2) and (5) exclude national newspapers, as their market borders are not local. In Columns (3) and (6) the sample is restricted to only include one newspaper market border per county, and it is the border that had the highest number of articles about fracking among all of the borders the county belongs to. Standard errors adjusted for clustering at the origin designated market area are in parentheses. $p<0.01$ ***, $p<0.05$ **, $p<0.1$ *.

Table 3.5: Source of News: Impact of Newspaper and TV News Exposure on Migration to Fracking Regions

	Inverse Hyperbolic Sine of the Number of Migrants$_{oSt}$				Inverse Hyperbolic Sine of the Number of Cross-County Commute Jobs$_{oSt}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Newspaper Exposure$_{oSt}$	0.052***		0.052***	0.051**	0.045**		0.055**	0.045**
	(0.018)		(0.018)	(0.018)	(0.019)		(0.019)	(0.019)
Newspaper Exposure$^2_{oSt}$	-0.002***		-0.002***	-0.002***	-0.001**		-0.001**	-0.001**
	(0.001)		(0.001)	(0.001)	(0.001)		(0.001)	(0.001)
TV News Exposure$_{oSt}$		0.040	0.029	0.100		0.125**	0.115*	0.134*
		(0.072)	(0.072)	(0.064)		(0.062)	(0.061)	(0.069)
TV News Exposure$^2_{oSt}$		-0.010	-0.011	-0.021		-0.019	-0.019	-0.018
		(0.017)	(0.017)	(0.013)		(0.014)	(0.014)	(0.014)
2008 TV Viewership Rates		X	X			X	X	
2016 TV Viewership Rates				X				X
Dependent Mean (in Levels)	60.2	60.2	60.2	60.2	152.6	152.6	152.6	152.6
Observations	32,864	32,864	32,864	32,864	27,808	27,808	27,808	27,808

Notes: Data obtained from the IRS Statistics of Income, LEHD Origin-Destination Employment Statistics (LODES), LexisNexis newspaper transcripts, and newspaper circulation from the Alliance for Audited Media. TV news circulation is only available at the Designated Market Area (DMA) level, from the 2008 Television Factbook, and all data is aggregated to that level. The level of observation is the origin DMA by destination state by year from 2000 to 2012. The variable *Newspaper Exposure$_{oSt}$* is scaled such that a one unit increase represents the impact of one additional news story in a newspaper with a 5 percent circulation rate. For comparison, circulation of the *USA TODAY* was 4.5 percent at the 95th percentile. The variable *TV News Exposure$_{oSt}$* is scales such that a one unit increase represents the impact of one additional TV news broadcast on a network with a 50 percent circulation rate, approximately the average circulation rate of ABC, CBS, or NBC. In all specifications origin/destination pair fixed effects, and destination by year fixed effects are included to control for time-invariant differences across pairs and characteristics of the destination that vary over time. Controls for the origin DMA unemployment rate, employment to population ratio, and average annual earnings (2010$) are also included. In 2008, there was significant transition to digital TV and full viewership ratings were not available, so Columns (4) and (8) use TV circulation from 2016 to construct TV news exposure. Standard errors are adjusted for clustering at the designated market area to account for correlation across geography and time. p<0.01 ***, p<0.05 **, p<0.1 *.

Table 3.6: Positive vs. Negative News: Impact of Newspaper Exposure on Migration and Commuting to Fracking Regions

	Inverse Hyperbolic Sine of the Number of Migrants_{oSt}		Inverse Hyperbolic Sine of the Number of Cross-County Commuting Jobs_{oSt}	
	(1)	(2)	(3)	(4)
Positive Newspaper $Exposure_{oSt}$	0.044***	0.047***	0.072***	0.056***
	(0.010)	(0.011)	(0.018)	(0.020)
Positive Newspaper $Exposure^2_{oSt}$	-0.004**	-0.004**	-0.007**	-0.003
	(0.002)	(0.002)	(0.003)	(0.003)
Negative Newspaper $Exposure_{oSt}$	0.026***	0.025***	0.098***	0.051***
	(0.005)	(0.005)	(0.014)	(0.011)
Negative Newspaper $Exposure^2_{oSt}$	-0.001***	-0.001***	-0.006***	-0.002***
	(0.0005)	(0.0004)	(0.001)	(0.001)
Origin by Year Fixed Effects		X		X
Dependent Mean	7.6	7.6	31.4	31.4
Observations	590,224	590,224	499,440	499,440

Notes: Data obtained from the IRS Statistics of Income, LEHD Origin-Destination Employment Statistics (LODES), LexisNexis newspaper transcripts, and newspaper circulation from the Alliance for Audited Media. The level of observation is the origin county by destination state by year from 2000 to 2012 for the migration data and 2002 to 2012 for the commute data. $Exposure_{oSt}$ measures are scaled such that a one unit increase represents the impact of one additional news story in a newspaper with a 5 percent circulation rate. A positive news article is one that contains at least two positive phrases (referencing jobs, boom, or growth) and more positive than negative phrases (referencing pollution, health, danger, or earthquakes), while a negative article is the opposite. Some fracking destinations have many positive and negative articles, leading to a high correlation between *Positive Newspaper* $Exposure_{oSt}$ and *Negative Newspaper* $Exposure_{oSt}$ ($\rho = 0.70$). Controls include the origin county unemployment rate, employment to population ratio, and average annual earnings (2010$). In all specifications origin/destination pair fixed effects and destination by year fixed effects, are included to control for time-invariant differences across pairs and characteristics of the destination and origin that vary over time. Standard errors are adjusted for clustering at the designated market area to account for correlation across geography and time. p<0.01 ***, p<0.05 **, p<0.1 *.

Table 3.7: Simulated Impacts of News Exposure Migration Flows by Origin Employment to Population Ratio

	Employment to Population Ratio in $t-1$ ($\mu = 70.9$)				
	60	65	70	75	80
Exposure Level					
0	0.0	0.0	0.0	0.0	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
1	0.027***	0.028***	0.027***	0.023***	0.017***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)
5	0.116***	0.125***	0.120***	0.103***	0.072***
	(0.015)	(0.016)	(0.016)	(0.015)	(0.014)
10	0.189***	0.210***	0.205***	0.174***	0.116***
	(0.025)	(0.027)	(0.027)	(0.026)	(0.026)
15	0.218***	0.256***	0.255***	0.214***	0.133***
	(0.031)	(0.034)	(0.035)	(0.035)	(0.040)
Mean Exposure in 2012	1.57	2.08	2.48	2.44	1.89
Implied Impact	0.041***	0.057***	0.064***	0.054***	0.030***
	(0.005)	(0.007)	(0.008)	(0.008)	(0.006)
Maximizing Exposure	15.9	18.4	19.6	18.9	15.6
Implied Impact	0.219***	0.265***	0.270***	0.223***	0.133***
	(0.032)	(0.038)	(0.041)	(0.043)	(0.042)

Notes: Simulated impacts are obtained for each combination of origin employment to population ratio and exposure level from equation (3.13), where the outcome is the inverse hyperbolic sine of the number of migrants. The corresponding coefficients are reported in Appendix Table A3.9. The maximizing exposure is obtained by setting the first derivative of equation (3.13) with respect to newspaper exposure equal to zero and solving for the maximizing exposure for the specified employment to population ratio. This value is rounded down to the nearest whole number. The implied impact is the corresponding effect of the maximizing exposure level. For reference, the mean exposure level in 2012 for origin counties with employment to population ratios within 2.5 percent of the specified threshold. An employment to population ratio of 60 roughly corresponds to the 15th percentile while a ratio of 80 corresponds to roughly the 85th percentile. p<0.01 ***, p<0.05 **, p<0.1 *.

Figures

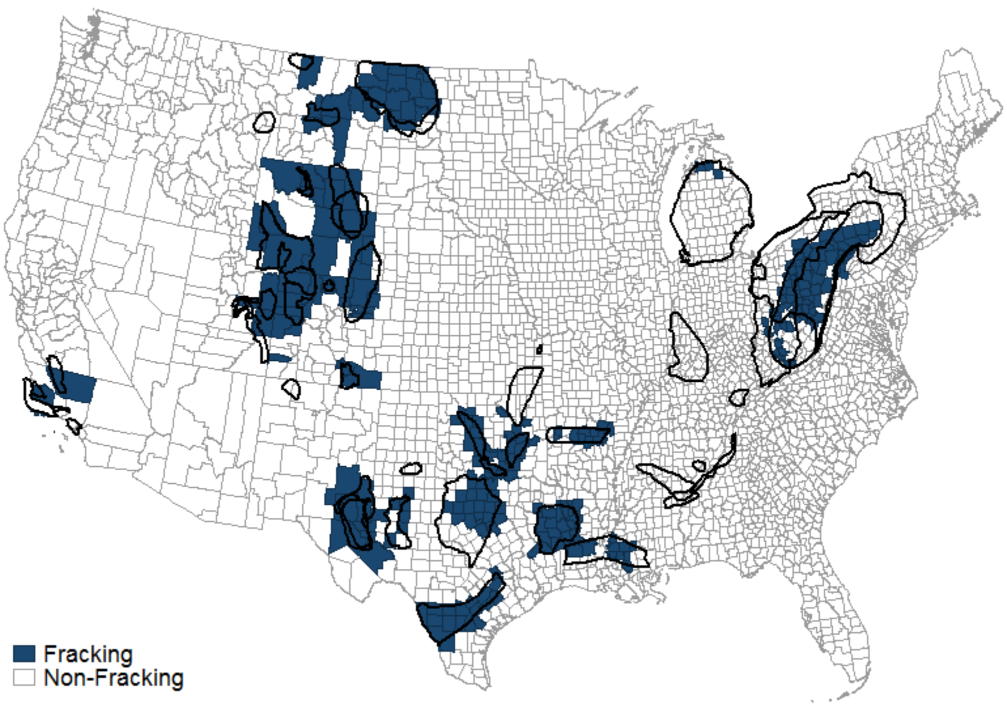

Figure 3.1: Fracking Counties and Shale Plays

Notes: Any county with production from fracking wells between 2000 and 2012 is labeled as a fracking county. Shale play boundaries are outlined in black.

Source: Author's calculations constructed from DrillingInfo well level data. Shale play boundaries are from the EIA.

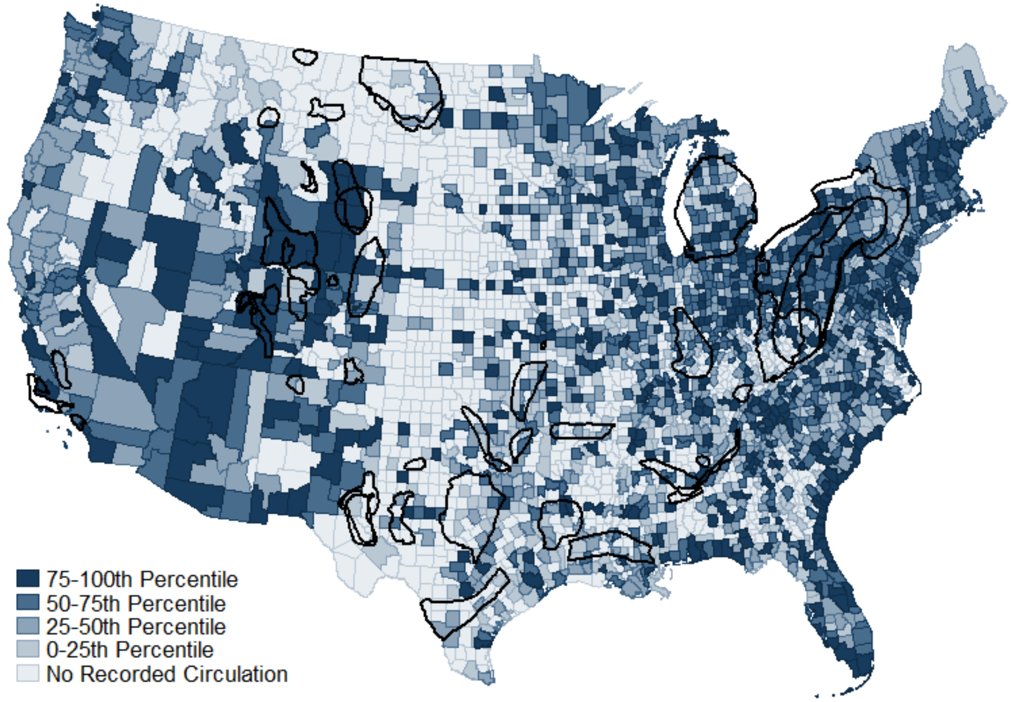

Figure 3.2: County-level Circulation of USA TODAY between 2005 and 2008

Notes: Location of shale plays outlined in black.

Source: Author's calculations using annual county-level circulation rates averaged between 2005 and 2008 obtained from the Alliance of Audited Media.

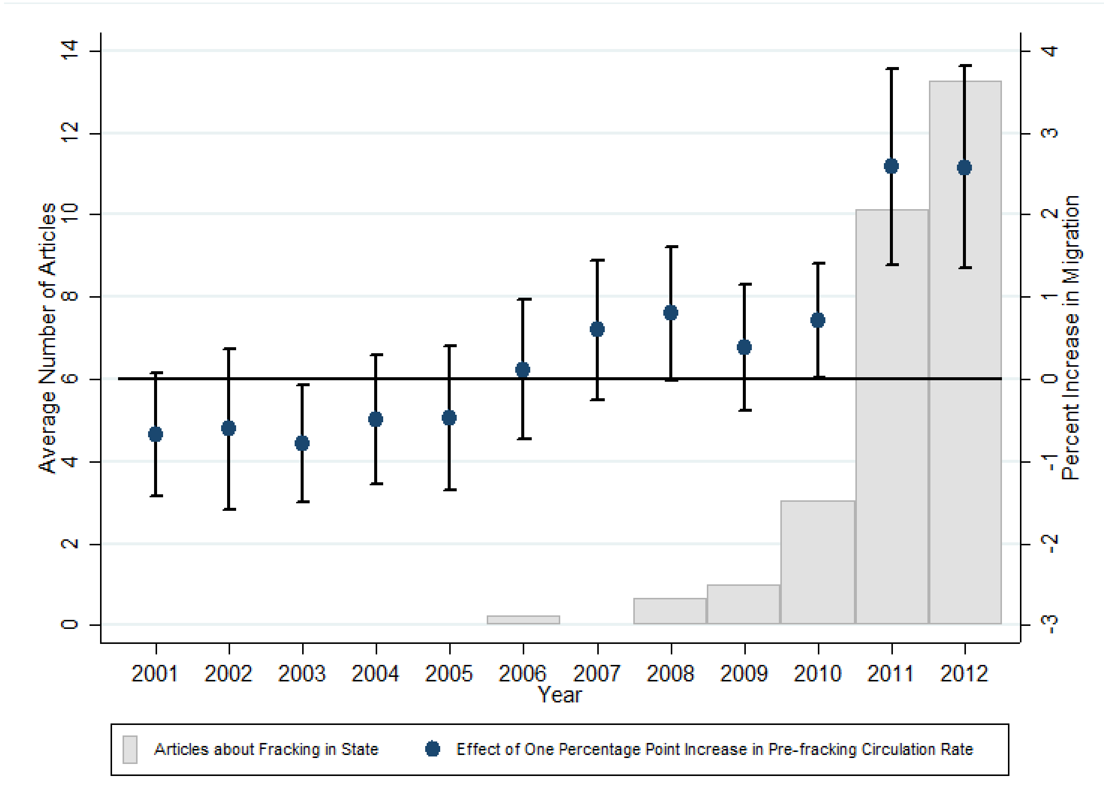

Figure 3.3: Trends in Migration by Pre-fracking Circulation

Notes: For each origin, the pre-fracking circulation rate is the weighted average of the pre-fracking circulation of the *USA TODAY*, *New York Times*, and *Wall Street Journal*, where weights are the share of the total articles about fracking in the destination state in each newspaper. This measure captures the extent to which an origin will eventually be exposed to fracking news. This measure is then interacted with year indicators. The inverse hyperbolic sine of the number of migrants is then regressed on this set of interactions along with origin-destination pair effects and destination-by-year fixed effects, as in the main specification, to trace out the effect of a one percentage point increase in the pre-fracking circulation rate on migration, as a percent. The coefficients on these year interactions are interpreted as the marginal effect of a one percentage point increase in the pre-fracking circulation rate on migration flows in that given year and are plotted for each year on the right axis, to look at trends by differences in eventual exposure. Standard errors from the regressions are corrected for clustering at the origin DMA level. For reference, the average number of articles about fracking in each state is also plotted for each year in bars on the left axis.

Source: Author's calculations using circulation rates from the Alliance of Audited Media, newspaper content from LexisNexis, and migration flows from the IRS SOI.

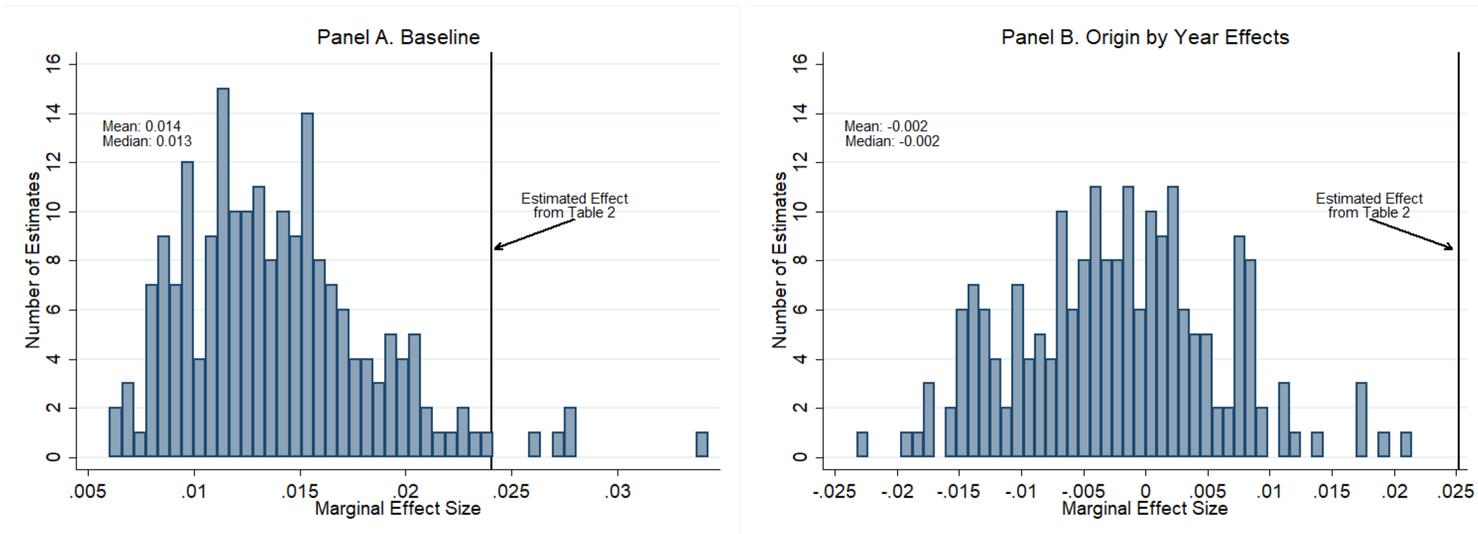

Figure 3.4: Migration Response to Randomly Assigned News about Fracking in a Different State

Notes: Each state is randomly assigned the fracking news exposure of a different state, and then the inverse hyperbolic sine of migration is regressed on a quadratic of this randomly assigned news exposure, similar to the baseline regression in equation (3.10). The histogram of estimated effects from the baseline model for 200 regressions are plotted in Panel A. For some states the trends in news exposure are similar, and across all 200 regressions the average correlation between actual news exposure and placebo news coverage was 0.44. Panel B. repeats the same 200 regressions but includes origin by year effects. This exploits variation in news coverage across destinations within an origin, relying on destination state specific deviations in news exposure.

Source: Author's calculation from 200 regressions of randomly assigned news exposure on the inverse hyperbolic sine of migration using circulation rates from the Alliance of Audited Media, newspaper content from LexisNexis, and county to county migration flows from the IRS SOI.

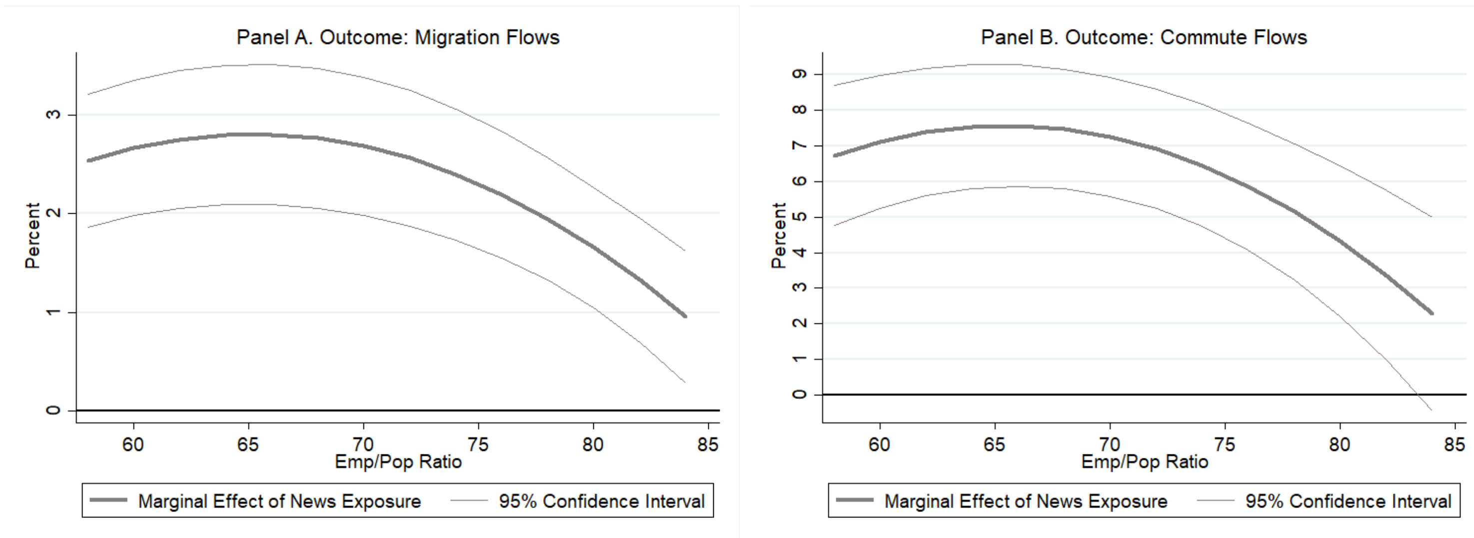

Figure 3.5: Heterogeneous Impacts of Newspaper Exposure by Origin Employment to Population Ratio in $t-1$

Notes: Marginal impact of newspaper exposure calculated by interacting a quadratic in newspaper exposure and a quadratic of lagged employment to population ratio at the origin. Approximately the 10th to 90th percentile of the employment to population ratio are plotted. Standard errors are calculated using the delta method.

Source: Author's calculations using circulation rates from the Alliance of Audited Media, newspaper content from LexisNexis, migration flows from the IRS SOI, and county employment to population ratio constructed from BLS QCEW data.

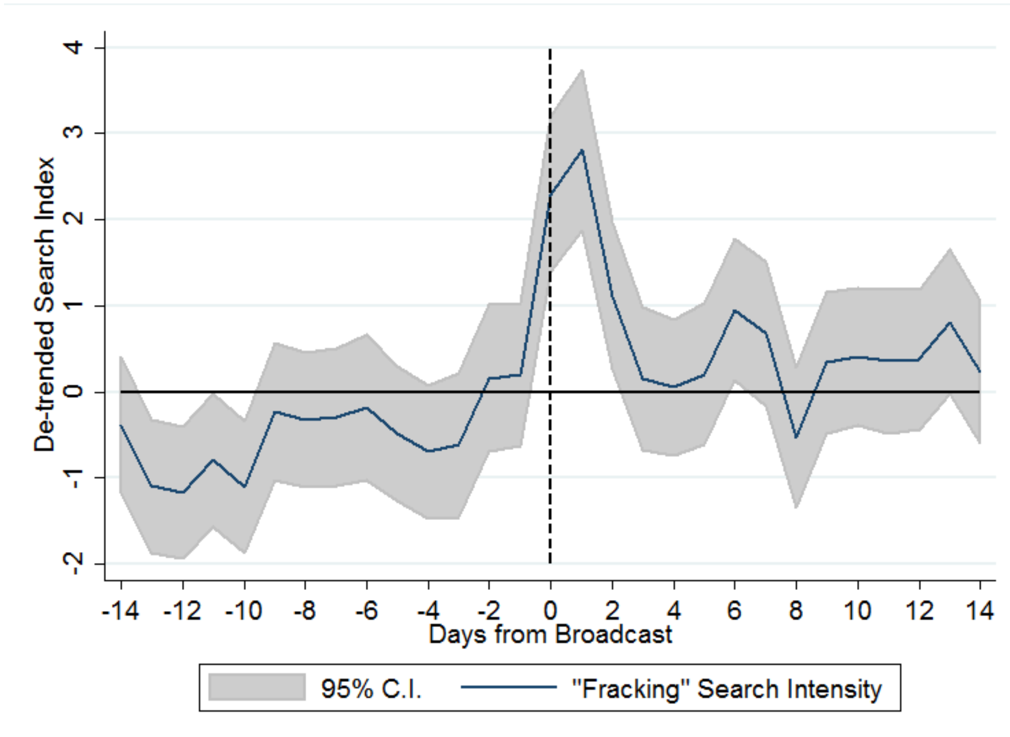

Figure 3.6: Google Search Interest in “Fracking” After TV News Broadcasts

Notes: Plot depicts the average daily search index for the term “fracking” by DMA before and after 17 TV broadcast mentioning fracking or shale gas between 2006 and 2012 as recorded by the Vanderbilt Television News Archive. Search intensity is de-trended by removing day of week and search (DMA by four week publication window) specific effects. To be consistent with other analysis in the paper, one broadcast from CNN and one broadcast from Fox News are excluded. Four days prior to a news broadcast on January 28, 2012, President Barack Obama mentioned shale gas exploration due to fracking in the State of the Union Address. Four days prior to a news broadcast on January 4, 2012, there was an earthquake in Ohio that reporters linked to fracking. For both of these event I include indicator variables for the next four days. Excluding these controls does not significantly change the daily average search index time series (see Figure A3.5). Standard errors are clustered at the search level.

Source: Source: Author’s calculations using daily search indices from Google Trends.

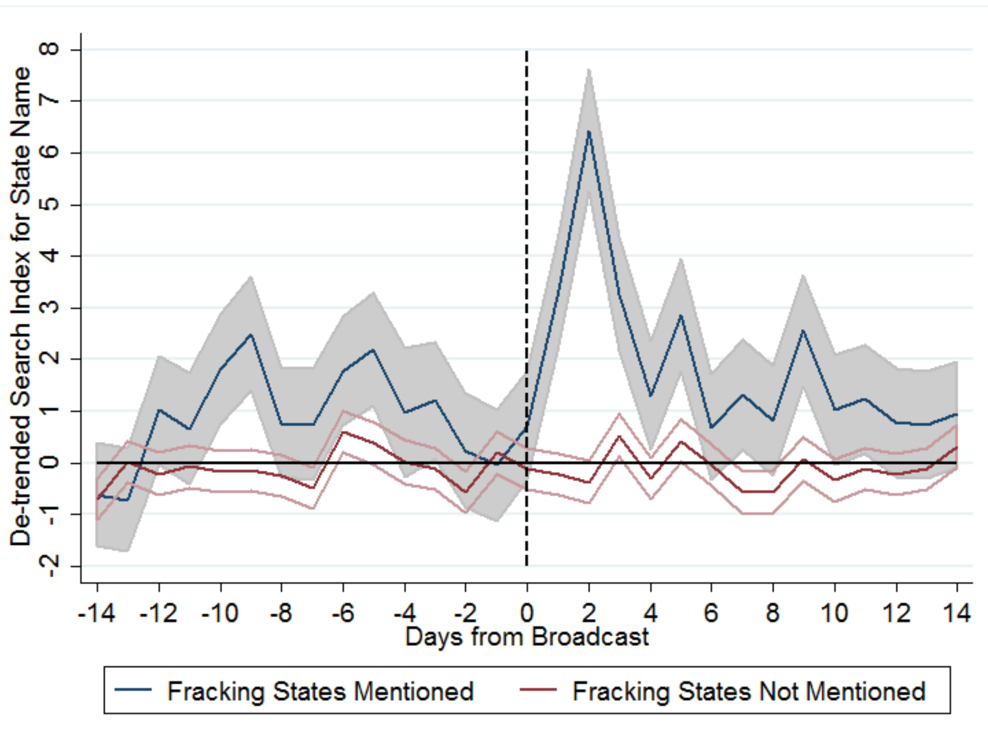

Figure 3.7: Google Search Interest in the Names of Fracking States Mentioned in TV News Broadcasts

Notes: Plot depicts the average daily search index for the name of the state by state before and after 14 TV broadcast mentioning fracking or shale gas and a specific state between 2006 and 2012 as recorded by the Vanderbilt Television News Archive. Search intensity is de-trended by removing day of week and search (DMA by four week publication window) specific effects. To be consistent with other analysis in the paper, one broadcast from CNN and one broadcast from Fox News are excluded. Additional control indicators are also included for specific high publicity state-specific events that fall in the search period window, such as the earthquakes, wildfires, special elections, and major sporting events. Excluding these controls does not significantly change the daily average search index time series (see Figure A3.5). For reference, the search intensity for fracking states *not* mentioned in the news broadcast is also plotted with 95 percent confidence intervals. Standard errors are clustered at the search level.

Source: Source: Author's calculations using daily search indices from Google Trends.

3.10 Appendix A. Additional Tables and Figures

Table A3.1: Content of Newspaper Articles

	Share of Articles that Mention							
	Jobs References[1]	*"boom"*	*"growth"*	*Pollution References*[2]	*"health"*	*"danger"*	*"earthquake"*	*Total Articles*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All Articles	0.17	0.24	0.14	0.39	0.20	0.12	0.06	562
Mention State:								
Arkansas	0.29	0.50	0.0	0.64	0.43	0.07	0.36	14
California	0.28	0.41	0.34	0.41	0.31	0.13	0.03	32
Colorado	0.10	0.39	0.23	0.65	0.28	0.13	0.06	31
Louisiana	0.32	0.47	0.26	0.58	0.21	0.26	0.05	19
Michigan	0.57	0.43	0.43	0.29	0.0	0.14	0.14	7
Mississippi	0.0	0.50	0.0	0.75	0.75	0.50	0.0	4
Montana	0.67	0.67	0.67	0.17	0.17	0.0	0.0	6
New Mexico	0.0	1.0	0.0	0.0	0.0	0.0	0.0	2
North Dakota	0.29	0.57	0.25	0.46	0.29	0.11	0.04	28
Ohio	0.33	0.36	0.12	0.55	0.39	0.21	0.24	33
Oklahoma	0.28	0.60	0.36	0.72	0.32	0.28	0.12	25
Pennsylvania	0.30	0.41	0.14	0.66	0.30	0.23	0.07	91
Texas	0.22	0.42	0.27	0.41	0.24	0.11	0.09	105
Utah	0.20	0.40	0.60	0.40	0.60	0.20	0.20	5
West Virginia	0.25	0.46	0.13	0.63	0.54	0.13	0.04	24
Wyoming	0.04	0.36	0.14	0.68	0.25	0.21	0.04	28

Notes: Newspaper content for articles between 2008 and 2012 obtained through LexisNexis, for the *New York Times*, *USA TODAY*, and *Wall Street Journal*. Not all articles reference a state, and some articles reference multiple states. Search terms are truncated to include various tenses and included both capitalized and lower case. [1] Jobs References include the following search terms: "new job", "creat∼ + job", "low + unemploy∼", "hire/hiring". [2] Pollution References include the following search terms: "contaminat∼" and "pollut∼".

Table A3.2: County Characteristics by the *New York Times* Pre-Fracking Circulation Rate

	Pre-2009 Circulation Rate of the *New York Times*				
	County Characteristics in 2000		Change from 2000 to 2010		Predicted Difference from 25th to 75th Percentile
	Below Median	Above Median	Below Median	Above Median	
	(1)	(2)	(3)	(4)	(5)
Migrants to fracking areas (Pct. of Population)	0.11	0.08	-0.02	-0.01	0.00
Employment to Population (16+)	55.62	59.22	-0.6	-1.62	-0.15**
Unemployment Rate	3.50	3.23	0.82	1.47	0.17***
Median Household Income	31,652	39,018	8,080	9,173	1,119***
Percent in Poverty	16.05	12.04	1.11	1.90	0.07
Percent White	84.49	85.0	-1.26	-1.84	-0.41***
Percent Black	9.40	8.82	0.11	0.16	0.10***
Percent Hispanic	5.35	6.24	1.83	2.32	0.28***
Percent Other Race	6.11	6.19	1.15	1.68	0.31***
Percent Population 20-34	17.55	18.72	-0.40	-0.82	-0.13***
Percent Population 35-64	38.28	39.04	1.65	1.64	0.03
Percent Population Over 64	15.65	13.95	1.08	1.28	0.07***
Percent Households Renting	25.37	26.40	1.75	1.69	-0.21***
Number of Counties	1,426	1,412	1,426	1,412	2,838

Notes: Migration data from the IRS Statistics of Income. Other county characteristics obtained through American FactFinder from the 2000 Census and 2010 Census and 5-Year American Community Survey. Circulation data for the *New York Times* from the Alliance for Audited Media. The county level median pre-2009 circulation rate of the *New York Times* was 0.32 percent, ranging from 0 to 3.29 percent. Circulation of the *New York Times* and the *Wall Street Journal* are highly correlated ($\rho = 0.8$), and characteristics look similar by circulation of the *Wall Street Journal*. Median Household Income is reported in current dollars. Column (5) reports the predicted change in the characteristic between 2000 to 2010 when pre-2009 circulation increases from the 25th to the 75th percentile. p<0.01 ***, p<0.05 **, p<0.1 *.

Table A3.3: Impact of Destination State Specific Newspaper Exposure on Cross-County Commuting to Fracking Regions

	Inverse Hyperbolic Sine of the Number of Cross-County Commute Jobs$_{oSt}$					
	By Monthly Earnings			By Broad Industry		
	$\leq$\$1,250	\$1,250–\$3,333	$\geq$\$3,333	Goods Producing	Trade and Transportation	Other Industry
	(1)	(2)	(3)	(4)	(5)	(6)
National Newspaper Exposure$_{oSt}$	0.034***	0.045***	0.045***	0.025***	0.029***	0.058***
	(0.005)	(0.006)	(0.006)	(0.005)	(0.004)	(0.007)
National Newspaper Exposure$^2_{oSt}$	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0003)
Dependent Mean	0.7	11.2	11.7	6.1	8.1	17.2
Observations	499,440	499,440	499,440	499,440	499,440	499,440

Notes: Data from the LEHD Origin-Destination Employment Statistics (LODES), LexisNexis newspaper transcripts, and newspaper circulation from the Alliance for Audited Media. The level of observation is the origin county by destination state by year from 2002 to 2012. LODES data is only available starting in 2002. Origin counties with any fracking production or in the New York City designated market area are excluded. The variable *National Newspaper Exposure*$_{oSt}$ is scaled such that a one unit increase represents the impact of one additional news story in a newspaper with a 5 percent circulation rate. For comparison, circulation of the *USA TODAY* was 4.5 percent at the 95th percentile. Earnings and Industry classifications are pre-defined in the LODES data. In all specifications origin/destination pair fixed effects, and destination by year fixed effects are included to control for time-invariant differences across pairs and characteristics of the destination that vary over time. Controls for the origin county unemployment rate, employment to population ratio, and average annual earnings (2010\$) are also included. Standard errors are adjusted for clustering at the designated market area to account for correlation across geography and time. p<0.01 ***, p<0.05 **, p<0.1 *.

Table A3.4: Robustness to Functional Form

	Inverse Hyperbolic Sine of the Number of Migrants$_{oSt}$				Inverse Hyperbolic Sine of the Number of Cross-County Commute Jobs$_{oSt}$			
	Linear	Quadratic	Cubic	IHS	Linear	Quadratic	Cubic	IHS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
National Newspaper Exposure$_{oSt}$	0.014***	0.025***	0.022***		0.037***	0.068***	0.099***	
	(0.003)	(0.004)	(0.005)		(0.007)	(0.009)	(0.011)	
National Newspaper Exposure$^2_{oSt}$		-0.001***	-0.0002			-0.002***	-0.006***	
		(0.0001)	(0.001)			(0.0004)	(0.001)	
National Newspaper Exposure$^3_{oSt}$			0.00001				0.0001***	
			(0.00001)				(0.00002)	
Inverse Hyperbolic Sine of				0.046***				0.155***
National Newspaper Exposure$_{oSt}$				(0.007)				(0.019)
Dependent Mean (in Levels)	7.6	7.6	7.6	7.6	31.4	31.4	31.4	31.4
Observations	590,224	590,224	590,224	590,224	499,440	499,440	499,440	499,440

Notes: Data obtained from the IRS Statistics of Income, LEHD Origin-Destination Employment Statistics (LODES), LexisNexis, and the Alliance for Audited Media. The level of observation is the origin county by destination state by year. *National Newspaper Exposure*$_{oSt}$ is scaled such that a one unit increase represents the impact of one additional news story in a newspaper with a 5 percent circulation rate. The inverse hyperbolic sine approximates a natural log transformation, but is defined for values of zero. Controls for the origin county unemployment rate, employment to population ratio, and average annual earnings (2010$) are also included. In all specifications origin/destination pair fixed effects, and destination by year fixed effects are included to control for time-invariant differences across pairs and characteristics of the destination that vary over time. Standard errors are adjusted for clustering at the designated market area to account for correlation across geography and time. $p<0.01$ ***, $p<0.05$ **, $p<0.1$ *.

Table A3.5: Sensitivity to Sample

	Inverse Hyperbolic Sine of the Number of Migrants_{oSt}				Inverse Hyperbolic Sine of the Number of Cross-County Commute Jobs_{oSt}			
	Include NYC DMA	Include Fracking Origins	Exclude Top One Percent of Exposure	Exclude Zero Exposure	Include NYC DMA	Include Fracking Origins	Exclude Top One Percent of Exposure	Exclude Zero Exposure
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
National Newspaper $Exposure_{oSt}$	0.013***	0.025***	0.031***	0.022***	0.045***	0.067***	0.123***	0.041***
	(0.004)	(0.003)	(0.006)	(0.003)	(0.006)	(0.009)	(0.013)	(0.007)
National Newspaper $Exposure^2_{oSt}$	-0.0001**	-0.001***	-0.0003	-0.001***	-0.0004***	-0.002***	-0.007***	-0.001**
	(0.0001)	(0.0001)	(0.001)	(0.0001)	(0.0001)	(0.0003)	(0.001)	(0.0003)
Dependent Mean (in Levels)	7.7	12.2	6.1	11.4	31.2	62.2	26.9	49.1
Observations	596,256	639,840	505,664	160,350	504,544	541,504	427,840	160,350

Notes: Data obtained from the IRS Statistics of Income, LEHD Origin-Destination Employment Statistics (LODES), LexisNexis, and the Alliance for Audited Media. The level of observation is the origin county by destination state by year. *National Newspaper* $Exposure_{oSt}$ is scaled such that a one unit increase represents the impact of one additional news story in a newspaper with a 5 percent circulation rate. Controls for the origin county unemployment rate, employment to population ratio, and average annual earnings (2010$) are also included. In all specifications origin/destination pair fixed effects, and destination by year fixed effects are included to control for time-invariant differences across pairs and characteristics of the destination that vary over time. Standard errors are adjusted for clustering at the designated market area to account for correlation across geography and time. p<0.01 ***, p<0.05 **, p<0.1 *.

Table A3.6: Accounting for Censoring: Impact of Destination State Specific Newspaper Exposure on Migration Flows

	Inverse Hyperbolic Sine of the Number of Migrating Tax Units$_{oSt}$		Number of Migrating Tax Units$_{oSt}$		Over 10 Migrating Tax Units$_{oSt}$	Inverse Hyperbolic Sine of the Number of Migrants$_{oSt}$
	As Reported	Lower Bound: Replace 0 with 9	As Reported	Lower Bound: Replace 0 with 9		Positive Flows in All Years
	(1)	(2)	(3)	(4)	(5)	(6)
National Newspaper Exposure$_{oSt}$	0.022***	0.001***	0.838***	0.732***	0.005***	0.018**
	(0.003)	(0.0003)	(0.222)	(0.209)	(0.001)	(0.007)
National Newspaper Exposure$^2_{oSt}$	-0.001***	-0.00002**	-0.019**	-0.016*	-0.0002***	-0.0004**
	(0.0001)	(0.00001)	(0.010)	(0.009)	(0.00003)	(0.0002)
Dependent Mean (in Levels)	4	145	4	145	0.03	348.7
Observations	590,224	590,224	590,224	590,224	590,224	12,092

Notes: Data obtained from the IRS Statistics of Income, LexisNexis Newspaper transcripts, and newspaper circulation from the Alliance for Audited Media. The level of observation is the origin county by destination state by year from 2000 to 2012. The variable *National Newspaper Exposure$_{oSt}$* is scaled such that a one unit increase represents the impact of one additional news story in a newspaper with a 5 percent circulation rate. For comparison, circulation of the *USA TODAY* was 4.5 percent at the 95th percentile. In Columns (1) and (2) the outcome is the inverse hyperbolic sine of migrating tax units (rather than migrants). Censored values are assigned a value of 0 in Column (1), and assigned a value of 9 in Column (2), to provide a lower bound. In Columns (3) and (4) the outcome is the number of migrating tax units in levels, to account for the fact that percentages are not comparable when censored values are reassigned a value of 9. The outcome in Column (5) is an indicator that equals one if there were over 10 migrating tax units. During the sample period, flows with less than 10 returns were censored, and this outcome captures transitions across the censoring threshold. The outcome in Column (6) is the inverse hyperbolic sine of migrants for a subsample of origin/destination pairs that reported positive flows in all years. Controls for the origin county unemployment rate, employment to population ratio, and average annual earnings (2010$) are also included. In all specifications origin/destination pair fixed effects, and destination by year fixed effects are included to control for time-invariant differences across pairs and characteristics of the destination that vary over time. Standard errors are adjusted for clustering at the designated market area to account for correlation across geography and time. $p<0.01$ ***, $p<0.05$ **, $p<0.1$ *.

Table A3.7: Advertising Effects of News Exposure: Market Expanding or Share Stealing

	Inverse Hyperbolic Sine of the Number of Migrants_{oSt}		Inverse Hyperbolic Sine of the Number of Cross-County Commute Jobs_{oSt}	
	(1)	(2)	(3)	(4)
National Newspaper Exposure$_{oSt}$	0.027***	0.027***	0.041***	0.049***
	(0.004)	(0.003)	(0.007)	(0.007)
National Newspaper Exposure$^2_{oSt}$	-0.001***	-0.001***	-0.001***	-0.001***
	(0.0001)	(0.0001)	(0.0003)	(0.0003)
All States Newspaper Exposure$_{ot}$	0.0003		0.005***	
	(0.0003)		(0.001)	
All States Newspaper Exposure$^2_{ot}$	-0.00001**		-0.00003***	
	(0.000003)		(0.00001)	
Max. State Newspaper Exposure$_{ot}$		0.001		0.015***
		(0.001)		(0.004)
Max. State Newspaper Exposure$^2_{ot}$		-0.0001**		-0.0004***
		(0.0001)		(0.0002)
Dependent Mean (in Levels)	7.6	7.6	31.4	31.4
Observations	590,224	590,224	499,440	499,440

Notes: Data obtained from the IRS Statistics of Income, LEHD Origin-Destination Employment Statistics (LODES), LexisNexis, and the Alliance for Audited Media. The level of observation is the origin county by destination state by year. *Newspaper Exposure*$_{oSt}$ is scaled such that a one unit increase represents the impact of one additional news story in a newspaper with a 5 percent circulation rate. *All States Newspaper Exposure*$_{ot}$ is the total news exposure for all 16 destination states within an origin year, to determine if news about fracking in general affects migration. *Max. States Newspaper Exposure*$_{ot}$ is the highest level of news exposure across all 16 destination state within an origin year, to determine if higher news exposure leads to shifting away from other fracking destinations. Controls for the origin county unemployment rate, employment to population ratio, and average annual earnings (2010$) are also included. In all specifications origin/destination pair fixed effects, and destination by year fixed effects are included to control for time-invariant differences across pairs and characteristics of the destination that vary over time. Standard errors are adjusted for clustering at the designated market area to account for correlation across geography and time. p<0.01 ***, p<0.05 **, p<0.1 *.

Table A3.8: Heterogeneity by Newspaper: Impact of Newspaper Exposure on Migration and Commuting to Fracking Regions

	Inverse Hyperbolic Sine of the Number of Migrants$_{oSt}$		Inverse Hyperbolic Sine of the Number of Cross-County Commuting Jobs$_{oSt}$	
	(1)	(2)	(3)	(4)
USA TODAY Exposure$_{oSt}$	0.027***	0.021***	0.081***	0.030***
	(0.004)	(0.004)	(0.013)	(0.007)
USA TODAY Exposure$^2_{oSt}$	-0.001***	-0.001***	-0.003**	-0.001
	(0.0003)	(0.0002)	(0.001)	(0.001)
New York Times Exposure$_{oSt}$	0.007***	0.009***	0.019***	0.016***
	(0.002)	(0.002)	(0.004)	(0.004)
New York Times Exposure$^2_{oSt}$	-0.0001**	-0.0002***	-0.001***	-0.0004**
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Wall Street Journal Exposure$_{oSt}$	0.098*	0.056	0.097	0.018
	(0.055)	(0.046)	(0.130)	(0.086)
Wall Street Journal Exposure$^2_{oSt}$	-0.027	-0.006	-0.062	-0.033
	(0.036)	(0.029)	(0.098)	(0.060)
Origin by Year Effects		X		X
Dependent Mean	7.6	7.6	31.4	31.4
Observations	590,224	590,224	499,440	499,440

Notes: Data obtained from the IRS Statistics of Income, LEHD Origin-Destination Employment Statistics (LODES), LexisNexis Newspaper transcripts, and newspaper circulation from the Alliance for Audited Media. The level of observation is the origin county by destination state by year from 2000 to 2012 for the migration data and 2002 to 2012 for the commute data. Each newspaper's exposure level is scaled to represent the impact of one additional news story in a county with circulation at the 95th percentile (3.9 percent for the *USA TODAY*, 1.9 percent for the *New York Times*, and 2.4 percent for the *Wall Street Journal*). Controls include the origin county unemployment rate, employment to population ratio, and average annual earnings (2010$). In all specifications origin/destination pair fixed effects and destination by year fixed effects are included to control for time-invariant differences across pairs and characteristics of the destination that vary over time. Standard errors are adjusted for clustering at the designated market area to account for correlation across geography and time. p<0.01 ***, p<0.05 **, p<0.1 *.

Table A3.9: Heterogeneous Impacts by Origin Employment to Population Ratio

	Inverse Hyperbolic Sine		Levels	
	Number of Migrants	Number of Commuters	Number of Migrants	Number of Commuters
	(1)	(2)	(3)	(4)
Newspaper Exposure$_{oSt}$	0.0270***	0.0731***	1.7029***	5.3651**
	(0.0037)	(0.0088)	(0.5103)	(2.2426)
Newspaper Exposure$^2_{oSt}$	-0.0007***	-0.0020***	-0.0392*	-0.2425**
	(0.0001)	(0.0004)	(0.0234)	(0.1048)
Newspaper Exposure$_{oSt}$ *	-0.0006***	-0.0017**	-0.0100	-0.1350
Emp/Pop$_{ot-1}$	(0.0002)	(0.0006)	(0.0221)	(0.0958)
Newspaper Exposure$_{oSt}$ *	-0.0001***	-0.0002***	-0.0040***	-0.0189***
Emp/Pop$^2_{ot-1}$	(0.000)	(0.000)	(0.0013)	(0.0066)
Newspaper Exposure$^2_{oSt}$ *	0.000	0.000	-0.0009	0.0058
Emp/Pop$_{ot-1}$	(0.000)	(0.0001)	(0.0011)	(0.0043)
Newspaper Exposure$^2_{oSt}$ *	-0.000	0.000	-0.000	0.0008**
Emp/Pop$^2_{ot-1}$	(0.000)	(0.000)	(0.0001)	(0.0004)
Emp/Pop$_{ot-1}$	0.0004**	0.0002	0.0052	-0.1376
	(0.0002)	(0.0006)	(0.0192)	(0.1236)
Emp/Pop$^2_{ot-1}$	0.000	-0.000	0.0006*	0.0032
	(0.000)	(0.000)	(0.0004)	(0.0021)
Dependent Mean (in levels)	7.635	31.23	7.635	31.23
Observations	544,688	499,296	544,688	499,296

Notes: Data obtained from the IRS Statistics of Income, LEHD Origin-Destination Employment Statistics (LODES), LexisNexis, and the Alliance for Audited Media. The level of observation is the origin county by destination state by year. The origin county employment to population ratio (Emp/Pop) is obtained from the BLS, and lagged by one year. Emp/Pop is demeaned, such that the direct effect of newspaper exposure is the effect for a county at the mean employment to population ration (70.9 percent). Controls for the current origin county unemployment rate, employment to population ratio, and average annual earnings (2010$) are also included. In all specifications origin/destination pair fixed effects, and destination by year fixed effects are included to control for time-invariant differences across pairs and characteristics of the destination that vary over time. Standard errors are adjusted for clustering at the designated market area to account for correlation across geography and time. p<0.01 ***, p<0.05 **, p<0.1 *.

Table A3.10: Simulated Impacts of News Exposure Cross-County Commute Flows by Origin Employment to Population Ratio

	Employment to Population Ratio in $t-1$ ($\mu = 70.9$)				
	60	65	70	75	80
Exposure Level					
0	0.0	0.0	0.0	0.0	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
1	0.071***	0.076***	0.073***	0.062***	0.043***
	(0.010)	(0.009)	(0.009)	(0.009)	(0.011)
5	0.315***	0.336***	0.322***	0.274***	0.191***
	(0.041)	(0.040)	(0.039)	(0.040)	(0.046)
10	0.527***	0.566***	0.544***	0.461***	0.319***
	(0.70)	(0.071)	(0.071)	(0.070)	(0.074)
15	0.637***	0.689***	0.664***	0.562***	0.384***
	(0.089)	(0.099)	(0.102)	(0.098)	(0.093)
Mean Exposure in 2012	1.57	2.08	2.48	2.44	1.89
Implied Impact	0.110***	0.153***	0.172***	0.144***	0.079***
	(0.015)	(0.018)	(0.020)	(0.021)	(0.020)
Maximizing Exposure	17.9	18.3	18.5	18.4	17.7
Implied Impact	0.654***	0.713***	0.689***	0.582***	0.393***
	(0.099)	(0.118)	(0.128)	(0.120)	(0.105)

Notes: Simulated impacts are obtained for each combination of origin employment to population ratio and exposure level from equation (3.13), where the outcome is the inverse hyperbolic sine of the number of cross-county commuters. The corresponding coefficients are reported in Appendix Table A3.9. The maximizing exposure is obtained by setting the first derivative of equation (3.13) with respect to newspaper exposure equal to zero and solving for the maximizing exposure for the specified employment to population ratio. This value is rounded down to the nearest whole number. The implied impact is the corresponding effect of the maximizing exposure level. For reference, the mean exposure level in 2012 for origin counties with employment to population ratios within 2.5 percent of the specified threshold. An employment to population ratio of 60 roughly corresponds to the 15th percentile while a ratio of 80 corresponds to roughly the 85th percentile. p<0.01 ***, p<0.05 **, p<0.1 *.

Table A3.11: State-level Events Controlled for in Google Trends State Specifications

Date (1)	Event (2)	States (3)
November 7-9, 2006	Four-way Texas Gubernatorial Election	Texas
September 9-10, 2010	San Bruno Pipeline Explosion	California
September 6-11, 2011	2011 Texas Wildfires	Texas
Dec. 31, 2011-Jan. 2, 2012	4.0 Earthquake in Eastern Ohio	Ohio
February 19-20, 2012	Texas A&M v. Oklahoma State Basketball Game	Texas, Oklahoma
February 27-28, 2012	Chardon High School shooting	Ohio
March 7, 2012	Ohio Primary Elections	Ohio
May 21-22, 2012	Tornado Outbreak of May 2013	Arkansas

Notes: High-level interest events that are closely tied to a specific state during the Google Trend search windows are controlled for to increase precision. Indicators that equal one for each of the listed dates for the destination state listed are included in the state name Google Trend analysis.

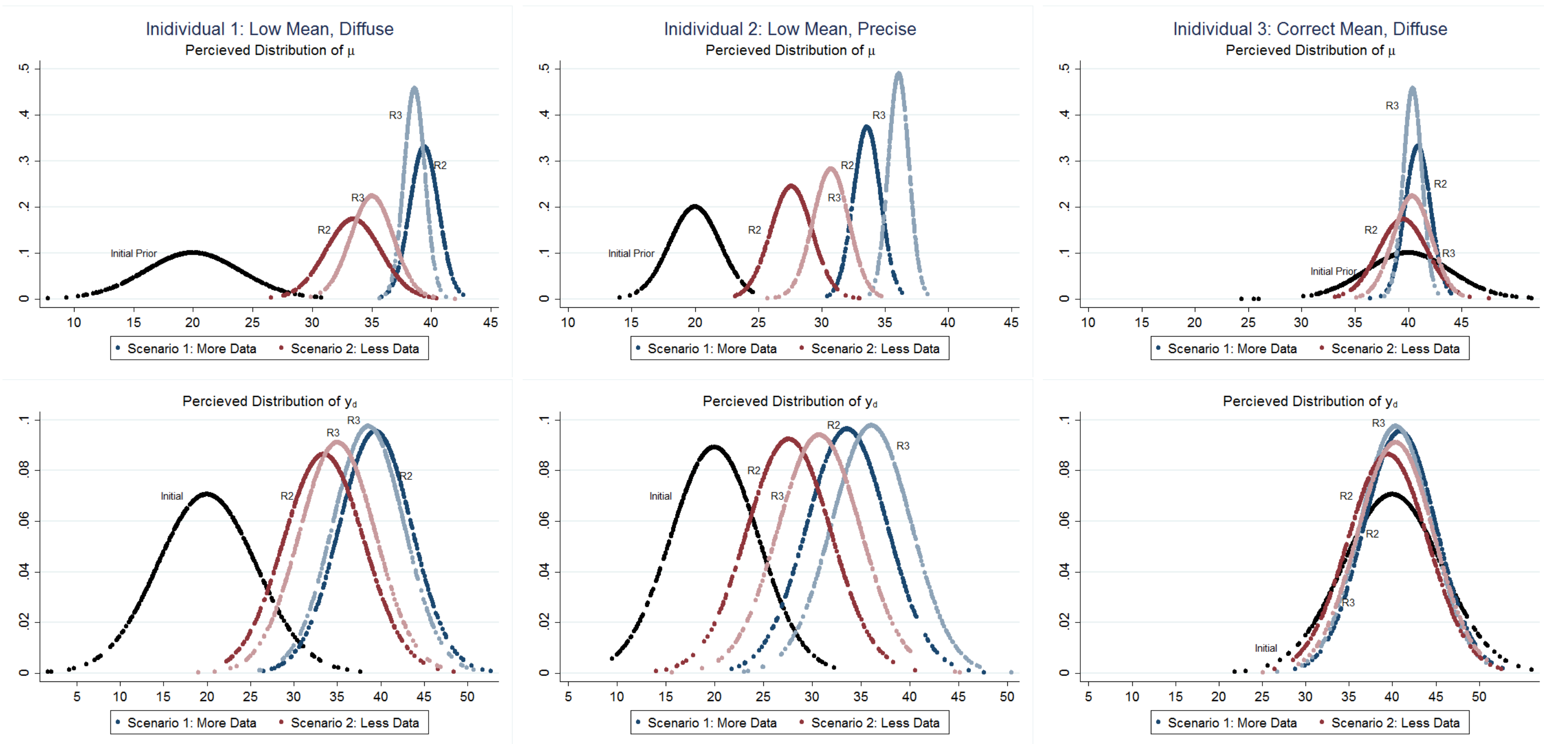

Figure A3.1: Model Simulations: Information and Bayesian Updating

Notes: Simulated data points from the distributions of μ_d and y_d are presented for three separate individuals in two separate scenarios. Individual 1 had a diffuse prior with a low mean, individual 2 had a more precise prior with a low mean, and individual 3 had a diffuse prior with a correct mean. In scenario 1 the individual viewed ten data points from the true distribution of y_d in each round (R2 and R3), and updates the posterior probability accordingly. In scenario 2 the individual views only 2 data points and updates the posterior. The initial prior and two additional iterations are shown.

Source: Author's calculations.

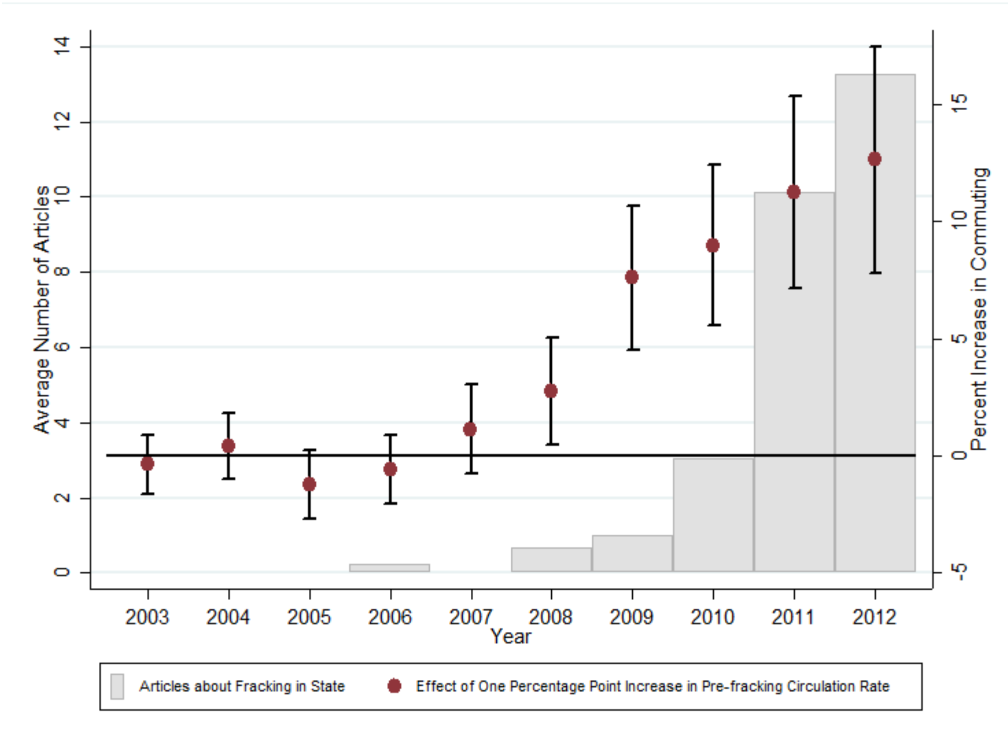

Figure A3.2: Trends in Commuting by Pre-fracking Circulation

Notes: For each origin, the pre-fracking circulation rate is the weighted average of the pre-fracking circulation of the *USA TODAY*, *New York Times*, and *Wall Street Journal*, where weights are the share of the total articles about fracking in each newspaper. This measure is then interacted with year indicators. The inverse hyperbolic sine of the number of cross-county commuting jobs is then regressed on this set of interactions along with origin-destination pair effects and destination-by-year fixed effects, as in the main specification, to trace out the effect of a one percentage point increase in the pre-fracking circulation rate on migration, as a percent. Commuting data is only available starting in 2002. The marginal effect of one unit of a one percentage point increase is converted to percentage points and plotted for each year on the right axis, to look at trends by pre-fracking circulation. Standard errors from the regressions are corrected for clustering at the origin DMA level. For reference, the average number of articles about fracking in each state is also plotted for each year in bars on the left axis.

Source: Author's calculations using circulation rates from the Alliance of Audited Media, newspaper content from LexisNexis, and commuting flows from the LODES.

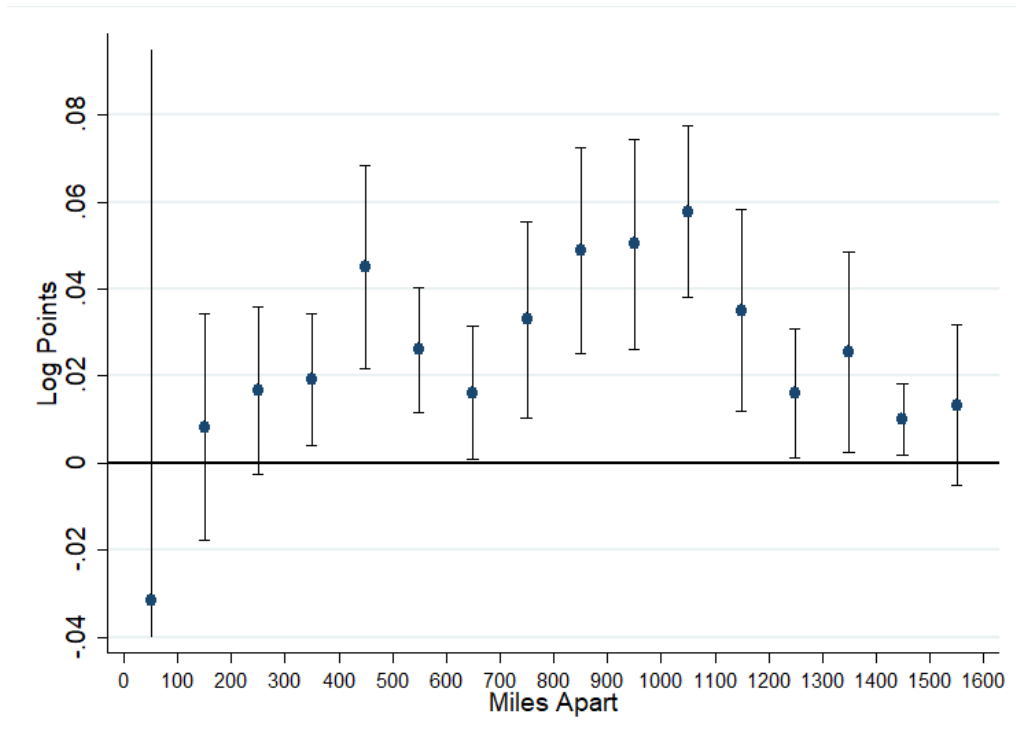

Figure A3.3: Marginal Impact of Newspaper Exposure on Migration by Origin to Destination Distance

Notes: Coefficients and confidence intervals plotted for the marginal effect of newspaper exposure on migration flows from equation (3.10), estimated over one hundred mile bins. Standard errors calculated using the Delta Method.

Source: Author's calculations using circulation rates from the Alliance of Audited Media, newspaper content from LexisNexis, and migration flows from the IRS SOI.

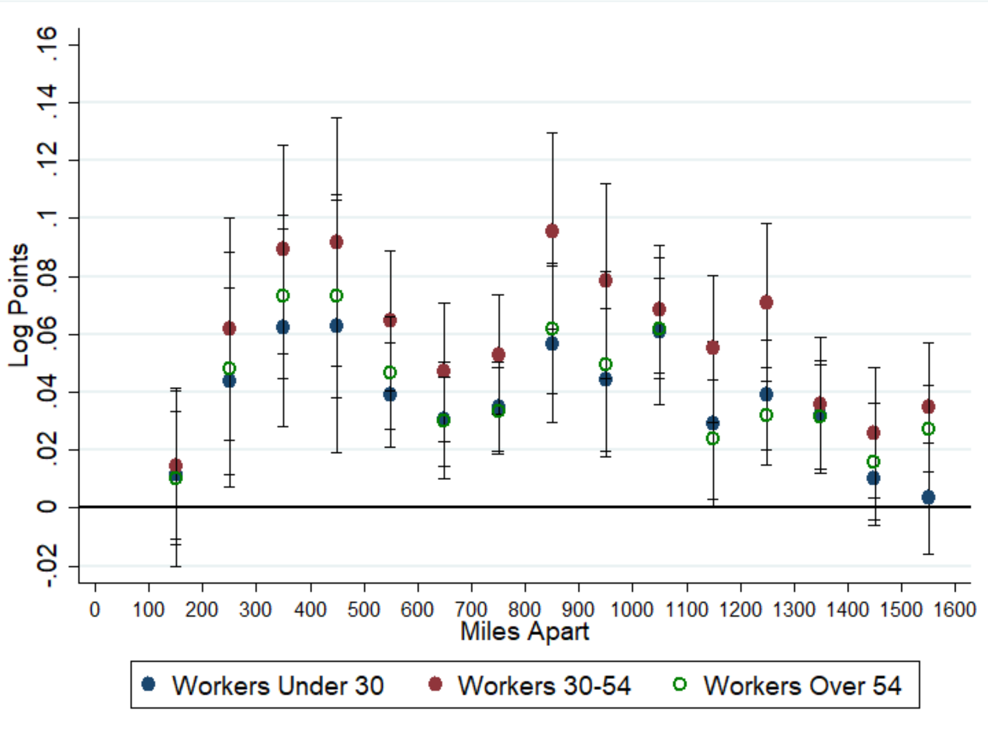

Figure A3.4: Marginal Impact of Newspaper Exposure on Commuting by Age and Origin to Destination Distance

Notes: Coefficients and confidence intervals plotted for the marginal effect of newspaper exposure on cross-county commute flows from equation (13), estimated over one hundred mile bins, separately by worker age. Standard errors calculated using the Delta Method. The estimated marginal impact for pairs less than one hundred miles apart for each age group are highly negative, at -0.06 (0.04), -0.10 (0.06), and -0.16 (0.06), respectively.

Source: Author's calculations using circulation rates from the Alliance of Audited Media, newspaper content from LexisNexis, and commute flows from the LEHD Origin-Destination Employment Statistics (LODES).

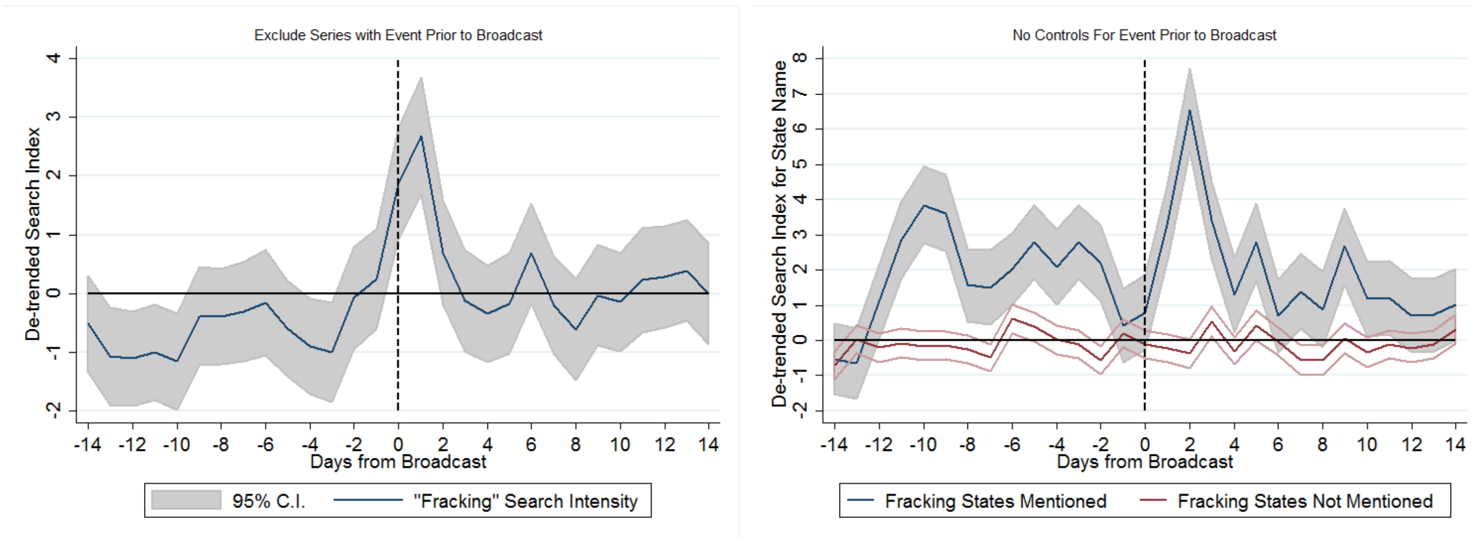

Figure A3.5: Google Search Interest: Exclude Event Controls

Notes: Plot depicts the same average daily search index for “fracking” and specific states mentioned as in Figures 3.6 and 3.7, but does not include controls for high publicity events that occurred during the search window and were either related to fracking, or a specific state. Standard errors are clustered at the search level.

Source: Author’s calculations using daily search indices from Google Trends.

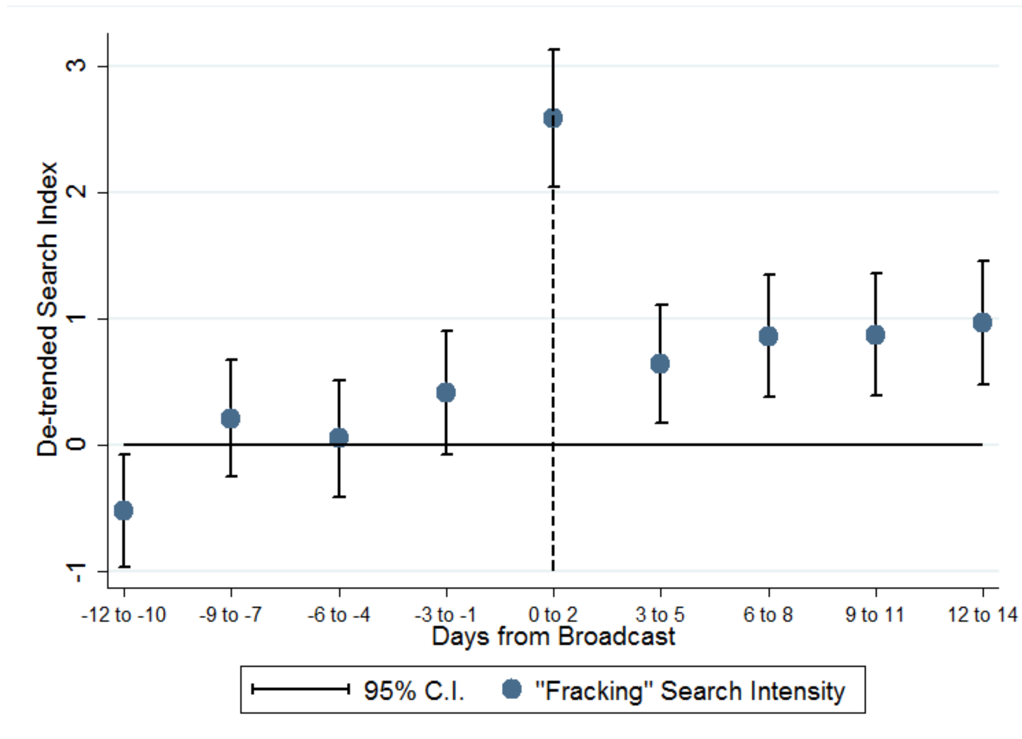

Figure A3.6: Google Search Interest in "Fracking" After TV News Broadcasts, 3 Day Bins

Notes: Plot depicts the average daily search index for the term "fracking" by DMA before and after 17 TV broadcast mentioning fracking or shale gas between 2006 and 2012 as recorded by the Vanderbilt Television News Archive, as in Figure 3.7, but groups days into 3-day bins. Search intensity is de-trended by removing day of week and search (DMA by four week publication window) specific effects. To be consistent with other analysis in the paper, one broadcast from CNN and one broadcast from Fox News are excluded. Four days prior to a news broadcast on January 28, 2012, President Barack Obama mentioned shale gas exploration due to fracking in the State of the Union Address. Four days prior to a news broadcast on January 4, 2012, there was an earthquake in Ohio that reporters linked to fracking. For both of these event I include indicator variables for the next four days. When these events are not controlled for, there is a marginally significant increase in search intensity in the days prior to the broadcast. Standard errors are clustered at the search level.

Source: Author's calculations using daily search indices from Google Trends.

3.11 Appendix B. Data Appendix

Below I describe each of the key datasets used in my analysis, as well as important characteristics of data construction.

DrillingInfo Oil and Gas Production Data

Well level information on drilling date, lease agreements, location, direction, and geological formation as well as other characteristics are provided through a restricted use data agreement from DrillingInfo. This data is proprietary, and obtained through an academic use agreement with DrillingInfo, available through their academic outreach initiative. DrillingInfo does not indicate if a well is a fracking well, as fracking is a means of stimulating production. To infer wells that are affected by the technological innovation associated with fracking, I use details on drilling direction and well location. Localized fracking booms occurred in part because of the combination of horizontal (directional) drilling and hydraulic fracturing. The DrillingInfo data reports whether a well is horizontally or vertically drilled. In addition, fracking was particularly impactful over shale plays, as these resources were not extractable previously. For this reason I assign non-vertical wells drilled in counties that intersect with shale plays as fracking wells. This production data is then combined with shale play boundary shapefiles provided by the Energy Information Administration to identify wells in shale plays and counties with fracking production.

Internal Revenue Service Statistics of Income County Flows

The Internal Revenue Service (IRS) Statistics of Income (SOI) division provides annual counts of county-to-county flows. This provides the raw number of tax returns and exemptions that were filed in one county in year $t-1$ and in another county in year t. Each year, the IRS provides county-to-county flows of exemptions in a file with two years (e.g., 2002to2003). This represents exemptions that were in one county when filing in 2002 and in another county when filing in 2003. As most people file in the beginning of the year before April, I assign this flow to the year 2002.

Using exemptions to approximate people in a household, I can identify origin-destination county level flows. For privacy purposes, the IRS suppresses county pairs that have fewer than ten returns move in each year. As such, county pairs that have small, positive flows will be recorded as zero. This potentially introduces measurement error. For this reason, I also consider lower bound specifications where all county to county flows of zero are replaced with nine. This operates under the assumption that all flows had at least nine households move, which is likely an extreme over-estimate. In 2013 the suppression threshold increased to 20 households. This led to considerably more suppression. For this reason I limit my analysis to 2012.

Unfortunately, the IRS county to county flows only provide aggregate numbers, and do not break up the migration levels by demographic characteristics (gender, marital status, education). As such, I am unable to use the IRS measures to look at differences across demographics. The only measure provided is the total adjusted gross income for all of the moved- returns. This is the adjusted gross income in the first year, but only the average for all movers in the county pair is provided.

The IRS data does not capture every move from one county to another. Low income individuals and households are not required to file a tax return, and thus might be under represented in the data. It is likely that individuals that move to fracking areas will earn well beyond the filing threshold after moving, but they might not have been required to file in the previous year. If there are individuals that did not file in the first year, but moved in response to fracking and filed in the second year, my estimates would be attenuated. Households that file for extensions past September will also not be included in the data, which might exclude very high income households with complicated returns.

The IRS data does not capture every move from one county to another. Low income individuals and households are not required to file a tax return, and thus might be under represented in the data. It is likely that individuals that move to fracking areas will earn well beyond the filing threshold after moving, but they might not have been required to file in the previous year. If there are individuals that did not file in the first year, but moved in response to fracking and filed

in the second year, my estimates would be attenuated. Households that file for extensions past September will also not be included in the data, which might exclude very high income households with complicated returns.

Alliance for Audited Media Newspaper Circulation Data

Newspaper circulation rates between 2005 and 2008 were obtained through a temporary academic membership at the Alliance for Audited Media (AAM). These circulation rates are provided in PDFs, which I scraped to collect county level estimates. In some cases the scrap was unable to identify the circulation rate, so hand corrections were made.

The AAM conducts regular (annual or biannual) audits of newspapers and collects circulation rates, along with other measures such as prices. This circulation rates includes the number of copies sold on the audit date and the number of copies as a percent of households for each county with over 25 copies. Counties with fewer than 25 copies sold are assigned a zero value. For most newspapers, circulation rates are reported at the county level. However, for the *New York Times* and *Wall Street Journal* these rates were provided at the DMA-level. For my county level analysis I assign each county the DMA-level, which reduces the variation and adds measurement error. However, as seen in Table 3.5, DMA-level estimates provide similar conclusions. A small subset of local newspapers that reported about fracking do not have AAM audits. For these newspapers, which often only distributed to one or two counties, statistics about local circulation was compiled from online searches. For local newspapers that were not audited annually, the intermittent values were imputed through linear interpolation. The three national newspapers report circulation every year.

TV and Cable Factbook TV Circulation Data

TV circulation data is taken from the Television Cable Factbook for 2008 and 2016. The Factbook contains information on local TV stations as well as DMA-level circulation as reported by Nielsens. TV circulation is reported at the DMA level for each TV station and includes viewership from both cable and non-cable households. This data is available at the station-level and not specific to news programming. The circulation rate is constructed by dividing total weekly viewership by the total number of households in the DMA. I use average weekly circulation rates throughout my analysis. For each station the "own" DMA and "other" DMA circulation is reported. Because it is not specified what "other" DMA is included, I only include circulation in "own" DMA. This is likely to attenuate the estimated effects. However, for many stations viewership outside the DMA is very low or non-existent. The 2016 circulation rates were obtained through a temporary online membership which provided only the current 2016 circulation rates.

For this reason, I also hired an undergraduate RA to collect circulation rates from the 2008 Factbook. Between 2007 and 2009, TV stations were transitioning from analog to digitally transmitted broadcasts on a market-by-market basis. When a market transitioned, viewers were required to obtain digital reception equipment, and it is unclear how this affected viewership and if 2008 viewership is representative of later years. For this reason I include estimates using both the 2008 and 2016 measures.

LexisNexis Newspaper Content Data

Newspaper content is collected through LexisNexis by searching on key terms, "frack*", "fracing", and "hydraulic fractur*". I then take the universe of articles, remove non-US sources (e.g., Daily Mail in the UK), and remove articles that only reference things like "Frick and Frack", unrelated acronyms, or last names. I then parse the entire text of these articles for each of the 16 state names (both capitalized and lower cased). References to states in the title of newspapers or place of publication are excluded, (ex: articles published in Colorado are not included as citing Colorado unless there is a reference in the body of the text). I then parse the entire text of the articles for positive and negative terms: "new job", "creat + job", "low + unemploy", "hire", "hiring", "boom", "growth", "earthquak", "environment", "health", "contaminat", "danger", and "pollut". Positive articles are articles that reference at least two positive terms and more positive terms than negative. Negative terms are the opposite. There are "neutral" articles that refer to fewer references that are not included when looking at news content, but have been included in previous specifications. When positive, neutral, and negative news are all included the patterns are similar but less precise.

Vanderbilt Television News Archive TV News Content

TV news content was pulled from the Vanderbilt Televisions News Archive (VTNA) and includes broadcasts that mention “fracing”, “frack”, or “shale”. The VTNA database contains TV news recordings and transcript abstracts for nightly news broadcasts from the three major news networks (ABC, CBS, and NBC) and the cable news channels CNN and Fox News. The database only includes one hour of programming each day for both cable news outlets. Because the available content of cable news is limited, and circulation rates are only available for the TV networks, I restrict the sample to TV broadcasts from the three major news networks. I parse the transcript abstracts for search terms such as “fracking” and “shale” as well as which state is being discussed. These clips are short often ranging from one to five minutes in length.

LEHD Origin Destination Employment Statistics Commute Data

The LEHD Origin Destination Employment Statistics (LODES) contains the number of workers for each residence/work place Census block pair. This data is available for all years since 2002, and also provides statistics by broad age (under 30, 30-54, over 54), monthly earnings (under $1,250, $1,250-3,333, over $3,333), and industry (goods, trade/transportation, other) groups. For each Census block I identify the corresponding county, and then aggregate up commute flows to the county to county level. For privacy, some noise is introduced at the Census block level, which likely remains at the county level, although to a lesser extent.

Chapter 4: The EITC and Employment Transitions: Labor Force Attachment, Annual Exit, and the Role of Information

4.1 Introduction

The Earned Income Tax Credit (EITC) is one of the largest components of the social safety net in the United States, transferring over $67 billion to 27 million households (IRS, 2017) and lifting over 6 million people out of poverty each year (Short, 2011; Hoynes & Patel, 2016). The EITC unambiguously incentivizes households to work at some point during the year, and the literature has consistently found evidence that the EITC increases annual employment rates of less educated single women, a group often eligible for the program (Nichols & Rothstein, 2015). However, both longitudinal data and ethnographic evidence suggest less-educated single women frequently transition in and out of employment and on and off of welfare within a year (Schochet & Rangarajan, 2004; Edin & Lein, 1997). At present it is unclear how the work incentives of the EITC impact these frequent, within year and across year employment transitions. In this paper, I evaluate how increases in EITC generosity affect less educated single women's within year labor force attachment and transitions out of the labor force.

For a less-educated single woman, the returns to a low-wage job are often offset by monetary and psychic employment costs as well as forgone means-tested benefits. When the net returns to work are low, even small week-to-week cost fluctuations, such as the need to stay home with a sick child or a minor car repair, can induce a woman to reduce labor supply. If faced with labor market rigidities that prevent adjustment at the intensive margin, this can result in volatile labor force attachment. For low levels of income the EITC operates as a wage subsidy, essentially raising the opportunity cost of dropping out of the labor force. In theory, this should increase the probability of staying in the labor force at any given time and mechanically increase the length of employment spells leading to less annual level exit. In this paper I empirically explore these hypotheses.

The annual, repeated cross-section data used by most of the EITC literature cannot inform us about within year decisions. I exploit the short panel nature of the CPS by linking individuals across months and years to estimate how year-to-year increases in federal EITC generosity in the 1990s affect these high frequency employment decisions. Unlike the previous work, this strategy allows me to evaluate within year measures and exploit within person differences. By linking individuals' monthly CPS surveys across years I observe a four month snap-shot of within year employment decisions and can estimate the impact of the EITC on within year employment transitions. For a one hundred dollar (2010$) increase in the maximum EITC a woman is eligible to receive, the probability of being employed during a four month period increased by one percentage point. The share of months working also increases by one percentage point. These estimates would suggest the large expansion in 1994 increased the share of months employed by 10 percentage points (12 percent). However, this response is only observed among women with recorded labor force participation during the initial year. The increase in EITC generosity slightly increased the share of months employed, but many women still eventually dropped out during the sample period. Accordingly, the incidence of multiple exits fell, suggesting that within year exit was reduced but not eliminated. These women still transitioned between employment and non-employment within the year, but less frequently than before. These patterns are robust and not present among similarly educated single women that report household incomes above the EITC region, or among more educated single women who generally fall outside the EITC region.

Given this increased labor force attachment within the year, I next use the linked annual ASEC March CPS supplement, to see if this affects annual level transitions out of the labor force. Consistent with the monthly analysis, less-educated single women responded to an increase in EITC generosity by working more weeks during the year. This response is concentrated among women who worked before the increase in EITC generosity, and results in more women staying in the labor force from year to year, effectively reducing annual level exit. No increase was detected among women who were not initially working, suggesting the EITC expansions during this period increased labor force participation largely by reducing exit among women with a previous labor force attachment, rather than pulling in new participants.

In the data, employment decisions respond to changes in the maximum credit eligible to *receive* in the current year rather than changes in the maximum credit eligible to *earn* in the current year. Previous work has suggested that during this time period, many EITC eligible households were unaware of the EITC (Mead, 2014). Since the EITC is a tax credit transferred with a one year lag, individuals can learn about the program and respond *ex post* (Nichols & Rothstein, 2015). The data are consistent with women lacking information about the program and updating their employment decisions after observing the added work incentives when they filed their taxes. Alternative explanations, such as the EITC relaxing liquidity constraints and allowing women to continue working by increasing cash on hand (e.g., they can continue paying for childcare) are less supported by the data.

Overall, the data suggest that the EITC increases labor force attachment among less educated single women with previous labor force participation, in part by delaying exit from the labor force. This increase in employment stability might generate potential benefits of the EITC not captured by the previous annual level analysis. However, as employment behavior only responds to changes in the maximum credit the individual is eligible to receive, it is important for individuals to have information about the program.

4.2 The EITC and Labor Force Attachment of Single Women

A large literature explores the structure and impacts of the EITC. For brevity I highlight the related work exploring labor supply impacts and refer the reader to Hotz and Scholz (2003) or Nichols and Rothstein (2015) for a more complete review. The EITC was introduced in 1975 as a refundable annual tax credit and is structured to reward work.[1] A household with no earned income does not receive a credit, but for each additional dollar of income the credit increases,

[1]The Advance EITC allowed recipients to receive their credit in advance with their paycheck throughout the year. However, take-up was extremely low (2-3%) and this provision was removed in 2011.

creating large negative tax rates for a short phase-in region, followed by a zero marginal tax rate plateau, and a gradual phase-out region (see Figure 4.1 for sample parameters in 1989, 1991, and 1994). The program has expanded several times since 1975, with the largest expansions occurring in 1994, 1995 and 1996, as seen in Figure 4.2.

Eissa and Liebman's (1996) pioneering work the EITC exploits variation from the 1986 expansion and compare single mothers with a high school degree or less to similarly educated single women without children and to more educated single mothers, to show that this early expansion increased labor force participation by 2.8 percentage points. Later expansions between 1993 and 1996 differentially affected households with one child and multiple children and also introduced a small credit for households without children. Much of the work exploring the EITC has focused on these expansions and compared single mothers with one child to single mothers with multiple children before and after the expansion, finding that annual employment increased by approximately 3-6 percentage points (Meyer & Rosenbaum, 2001; Hotz, Mullin, & Scholz, 2005). Most of the work has relied on repeated cross-sections from the March ASEC CPS supplement (Eissa & Liebman, 1996; Meyer & Rosenbaum, 2001; Eissa & Hoynes, 2004), although some have made use of longitudinal administrative data (Hotz, et al., 2005), and the National Longitudinal Survey of Youth (NLSY) to explore annual transitions when households become no longer EITC eligible (Moulton, Graddy-Reed, & Lanahan, 2016).[2] During this time period, other things that might differentially affect the single mothers with one child and multiple children were also changing, such as welfare reform (Looney & Manoli, 2013). This might lead to differential trends by family status and must be accounted for in the estimation (Meyer & Rosenbaum, 2001).

Mead (2014) raises the concern that many potential recipients lacked information about the program during this time period, casting doubt that it was the EITC driving employment

[2] To the best of my knowledge, only one concurrent working paper uses the longitudinal nature of the CPS to explore the EITC. Yucong (wp, 2016) links individuals in the March ASEC supplements to capture employment in the previous year and then estimates a repeated cross-section difference-in-differences by previous employment status to explore the effect of welfare reform and the 1993 EITC expansion on annual entry and exit. These results are difficult to interpret because the sample is conditioned on previous employment (the lagged outcome) which is potentially responding to the policy change.

rates during the 1990s. However, because the EITC is designed as a tax credit, annual tax filing provides an automatic mechanism for informing potential recipients about the program (Nichols & Rothstein, 2015). Even if potential recipients were not aware of the credit *a priori*, they might become aware of the program when filing taxes and adjust labor supply behavior *ex post* after perceiving that the returns to work are larger than initially believed.

Despite the ample evidence of a strong, extensive margin effect, there is less conclusive evidence on how individuals adjust at the intensive margin. Nichols and Rothstein (2015) highlight several reasons why this might be. Because the EITC induces participation, the composition of workers changes and repeated cross-sectional data cannot separately identify behavioral responses by those already working and compositional changed due to new entrants. In order to capture these intensive margin adjustments, an individual's work history must be observed both before and after the policy. Labor market frictions also make it difficult to adjust at the intensive margin. Presumably, many low wage workers face fixed work schedules and cannot flexibly adjust the number of hours worked. This rigidity might prevent them from adjusting at the intensive margin, even if they would prefer to do so. Finally, the EITC is a function of annual income, which might be difficult for individuals with volatile labor force attachment to predict. Uncertainty about annual income can lead to uncertainty about marginal incentives, which might lead to less adjustment at the intensive margin.

Recent work has exploited detailed administrative data to understand intensive margin responses to the EITC. Saez (2010) looks at kink points in the EITC schedule and finds no evidence of bunching or intensive margin responses among wage workers, but substantial bunching among self-employed workers at the first kink in the EITC schedule. Chetty and Saez (2013) randomly assign H&R Block clients to receive individualized information from the tax preparer about where they fall on the EITC schedule and how additional work would affect their credit in the subsequent year. This information did not change earnings in the following year on average, but there is significant heterogeneity across tax preparers. Using tax data, Chetty, Friedman, and Saez (2013) find that when people move to areas that have more "knowledge" of the EITC schedule they report incomes closer to the refund-maximizing level. This is even true among wage workers. Overall the previous work would suggest that either intensive margin elasticities are small or highly dependent on the information people have access to.

Despite the EITC's annual incentives, the empirical patterns suggest that less educated single women make extensive margin employment decisions at a much more frequent interval. From the CPS in the 1990s, 16-28 percent of single women with a high school degree or less who ever reported employment during a four month period, also entered or exited employment (see Table 4.1). In comparison, among single women with a college degree, this rate is less than 10 percent. From the 1996 SIPP, 38 percent of women in low-wage jobs had left the job by four months, and 53 percent had left by eight months (Schochet & Rangarajan, 2004). Over half of these exits were to non-employment. Edin and Lein (1997) present ethnographic evidence from the early 1990s that single mothers frequently transition between welfare and employment, and cycled through multiple jobs in a year. They conclude that when deciding to work or claim welfare, these single mothers often weighed the costs and benefit of each option and, "made reasonable assessments of how much they would need to earn to offset the added costs of work (p. 63)."[3]

It is reasonable to presume that some subset of these less educated women face labor market rigidities (such as inflexible work schedules) that make it infeasible to adjust at the intensive margin, and the only way to adjust labor supply when faced with an unexpected psychic or monetary cost associated with work (i.e., a sick child or a car repair), is to drop out of the labor force. For less-educated women facing minimum wage employment opportunities, the net benefits of working are already small or non-existent, and even small employment cost shocks can undo the benefits of working.[4]

Given the empirical evidence it seems likely that many less educated single women make frequent (monthly or weekly) employment decisions. To understand how the work incentives of the

[3]Sometimes this comparison was crude or approximate, while other mothers were able to recall or provide exact calculations on scraps of paper or the back of envelopes.

[4]An alternative way to consider this decision is to think of women facing a probability of being fired if they do not show up at work. For low wages (low returns to work) the expected loss associated with missing work will be smaller, making them more likely to skip work and be fired when facing a large employment cost shock.

EITC affect within year labor supply decisions, consider a conceptual model where a less educated single woman is making the decision to work ($y_t = 1$) or not ($y_t = 0$) in the current week (t) by comparing the lifetime expected utility she could achieve in both cases. Assume labor market rigidities prevent her from adjusting labor supply on the intensive margin and she can only adjust by entering or dropping out of the labor force. If she works she receives wages (w_t) and faces both psychic and monetary costs associated with working. The monetary costs (c_t) such as paying for transportation or childcare, can vary from period to period according to a known distribution F, but the current costs are observed by the woman before making her labor supply decision. The psychic utility cost (ϕ_t) associated with working is distributed according to G. This cost can range from the disutility associated with interacting with difficult managers or coworkers to the emotional cost of leaving children without adequate supervision as suggested by Edin and Lein (1997) (p. 133-136). For simplicity assume c_t and ϕ_t are independent.

If the woman does not work she receives an outside option benefit (b_t), that is taxed away if she works, similar to Aid to Families with Dependent Children (AFDC), Temporary Assistance for Needy Families (TANF), or Supplemental Nutritional Assistance Program (SNAP) benefits. The woman begins period t with wealth a_{t-1}, and can save or borrow from the future, where a_t^w represents the optimal wealth she carries forward to the next period if she chooses to work, and a_t^n is the optimal amount if she chooses not to work. She will decide to work in the current period if

$$u_t(w_t - c_t + a_{t-1} - a_t^w) - \phi_t + EV(a_t^w, y_t = 1) \geq u_t(b_t + a_{t-1} - a_t^n) + EV(a_t^n, y_t = 0). \tag{4.1}$$

In the case of both employment and non-employment, $u_t(\cdot)$ is the current period utility where $u' > 0$ and $u'' < 0$. $EV(\cdot, \cdot)$ is the expected future value, which is a function of wealth and employment status in period t. Allowing $EV(\cdot, \cdot)$ to depend on the employment status accounts for costs or frictions that potential entrants might face as well as potential long run benefits associated with employment.[5] I will remain agnostic about the functional form of $EV(\cdot, \cdot)$, but assume that it is increasing in wealth ($\frac{\partial EV(a_t, y_t)}{\partial a_t} > 0$). This assumption is fairly standard as the individual's future

[5]For example, stable employment might lead to longer job tenure and change the lifetime trajectory of wages.

self faces the same problem, but is endowed with more resources.

From equation (4.1), there will be a threshold psychic cost where the woman is indifferent between employment and non-employment in the current period, defined as

$$\phi^* = u_t(w_t - c_t + a_{t-1} - a_t^w) - u_t(b_t + a_{t-1} - a_t^n) + EV(a_t^w, y_t = 1) - EV(a_t^n, y_t = 0). \tag{4.2}$$

Given Ω_t (defined as wages, monetary costs, wealth, and the outside option) as well as the distribution of ϕ_t the probability the woman works will be $Pr(y_t = 1|\Omega_t) = Pr(\phi_t \leq \phi^*|\Omega_t)$.

Comparative statics suggest that an increase in the wage will have a direct effect through its impact on utility, but will also have an indirect effect through its impact on the optimal amount of wealth carried over, a_t^w. Because of the smoothing motives associated with u_t, this indirect effect will be positive, but less than one for one ($0 < \frac{\partial a_t^w}{\partial w_t} < 1$).[6] As future utility is increasing in wealth, the total impact of wages on the cost threshold is

$$\frac{\partial \phi^*}{\partial w_t} = u_t' * \left(1 - \frac{\partial a_t^w}{\partial w_t}\right) + \frac{\partial EV(a_t, y_t)}{\partial a_t} \frac{\partial a_t^w}{\partial w_t} > 0. \tag{4.3}$$

A small increase in the wage will increase the cost threshold, meaning the woman is now willing to incur a higher psychic cost and still work. This in turn increases the probability of working. Similar calculations show that an increase in c_t or b_t will reduce the probability of working. For a single women facing wages at or near the minimum wage, working might be a "financial wash (p.67, Edin & Lein, 1997)," and even a small psychic cost shock, such as the inability to leave work when a child is sick, might induce her to exit from employment.

Now suppose a tax policy is introduced, which gives positive transfers associated with work, similar to the EITC. Through this policy, low-income households received an income transfer, equal to some percentage (τ_t) of earnings (w_t). This income transfer is refunded with a lag like the EITC, where benefits are paid out in a future period.[7] The threshold cost now becomes

$$\phi^* = u_t(w_t - c_t + a_{t-1} - a_t^w) - u_t(b_t + a_{t-1} - a_t^n) + EV(a_t^w + \tau_t w_t, y_t = 1) - EV(a_t^n, y_t = 0). \tag{4.4}$$

[6]See Appendix B for a more detailed explanation.

[7]This transfer system is a simplified version of the EITC. In reality, the refund rate of the EITC is not a constant, but a function of earnings.

It is worth noting that this policy does not increase net wages in the current period, but only enters the expected utility in the future by providing more wealth in the future if the woman worked.[8] As with wages, an increase in τ_t will have a direct effect through additional future income, as well as an indirect effect through its impact on wealth carried over from period t. If we assume for simplicity that the EITC refund is transferred with a one period lag, then $-w_t < \frac{\partial a_t^w}{\partial \tau_t} < 0$.[9] Intuitively, the refund increases wealth in the future and reduces the need to save today, but, smoothing motives make this a less than one for one trade off. As such

$$\frac{\partial \phi^*}{\partial \tau_t} = \frac{\partial EV(a_t^w + \tau_t w_t, y_t = 1)}{\partial a_t}\left(\frac{\partial a_t^w}{\partial \tau_t} + w_t\right) > 0 \tag{4.5}$$

because future utility is increasing in wealth. All else equal, the woman's threshold cost is now higher, and she is willing to incur a higher psychic cost to work. If τ_t increases, then $Pr(y_t = 1|\tau_t, \Omega_t)$ will increase. This mechanically reduces the probability of exit ($Pr(y_t = 0|y_{t-1} = 1)$), which lengthens employment spells in expectation. The work incentives of the EITC increase the opportunity cost of dropping out of the labor force in any given period, leading to more weeks of employment, longer employment spells, and more stable employment with less frequent exit. Enriching this model to reflect the nuances of real-world decisions (such as fixed entry costs or liquidity constraints) generally affect the magnitude, not sign of these predictions.[10]

[8]There has been a recent literature exploring the incidence of the EITC and resulting impacts on market wages (Rothstein, 2008; Leigh, 2010). However, during this time period many less educated single women were paid wages at or near the minimum wage, constraining general equilibrium adjustments.

[9]In reality, the EITC transfers benefits as a lump sum in the following year when the individual files her taxes and the size of this income transfer is a function of aggregate earnings over all subperiods in a calendar year. As such, the average effect and marginal effect of an additional dollar of earnings might be different (see discussion below). The result presented here will be similar if we assume that the EITC is refunded with a longer lag, but does not capture the annual level determination of the transfer.

[10]Introducing fixed costs associated with entering the labor force increases labor force attachment of those already employed and deters non-working single women from entering employment. The tax transfer (τ_t) will still increase the opportunity cost of dropping out of the labor force. If less educated single women are liquidity constrained and cannot borrow against the future, there will be some probability (γ) that the woman cannot cover the monetary costs of work even if the psychic cost is low and lifetime value is high. However, with probability $1 - \gamma$ she can still account for the future benefits of the tax transfer, so an increase in τ_t will still increase the dropout threshold, but by a smaller magnitude because there is a positive probability the liquidity constraint will bind.

Average versus Marginal Tax Rates. The implications of this model will depend on whether people consider average or marginal tax rates. The increase in EITC generosity between 1993 and 1994 unambiguously lowered the EITC component of the average tax rates (making them more negative) for EITC eligible households with two children (see Figure 4.3). However, the effect on EITC component of the marginal tax rates varied across the EITC schedule. The marginal tax rate became more negative in the phase-in region, the marginal tax rate increased in the phase-out region (became more positive), and the plateau region shortened, with some households experiencing higher, lower, or the same marginal tax rate.[11] If people consider average tax rates, an increase in the maximum EITC credit unambiguously increases the refunded transfer, and will lead to more weeks of employment and longer employment spells with less frequent exit. However if people consider marginal tax rates, households in the phase-in region will experience a more negative marginal tax rate leading to a higher opportunity cost of employment exit, while households in the phase-out region will face more positive marginal tax rates and the returns to working in the current period actually fall. In this case the theoretical predictions about average behavior are less clear and ultimately an empirical question. Previous work suggests that people facing a multipart tax schedule, such as the EITC, respond to average tax rates rather than marginal tax rates, a phenomenon referred to as "ironing" in the public finance literature (Liebman & Zeckhauser, 2004).[12]

Information about EITC Policy and Learning. If women did not know about the EITC, as suggested by Mead (2014), and never learn about it, we would expect no employment response.

[11]For reference, from the 1994 CPS ASEC, approximately 30.5 percent of single women with children had family income in the phase-in region where both average and marginal tax rates were negative. An additional 10.3 percent were in the plateau region where the average tax rate was negative and the marginal tax rate was zero.

[12]One explanation for this is that it might be difficult for households with volatile labor force attachment to accurately predict their location on the schedule to identify the marginal tax rate, whereas average tax rates might be more salient. This logic has been used to explain the strong extensive margin response and weaker evidence of intensive margin responses (Liebman, 1998). During the 1990s and even through the 2000s, only the amount of the EITC credit was reported on the IRS Form 1040. As such, a filer using a tax preparer would only know the amount of the credit in relation to her earned income. Even self-prepared filers using Schedule 596 EITC instructions would only see a table similar to a tax table, which does not report marginal tax rates, although they could be calculated.

However, the EITC is issued as a tax credit, meaning that all households that qualify or are close to qualifying will potentially learn about the credit and how it changes the returns to work when they file their taxes.[13] If a woman does not know about the EITC, or a change to the EITC, she will operate under the decision represented by equation (4.1) until the lagged EITC is refunded. For simplicity, once again assume the refund is transferred with a one period lag. If the woman did not work in the previous period, her optimization problem remains unchanged. She does not receive the EITC income transfer ($\tau_t w_t$) and continues to operate under equation (4.1). However, if she did work in the previous period, she will start the new period with an unexpected transfer $\tau_t w_t$. If by receiving the transfer she learns that it pays to work, she will now make decisions based on equation (4.4). This is equivalent to increasing τ_t which will lead to a higher probability of employment and more stable employment in expectation. The scenario is similar if individuals know about the EITC, but are unaware of an increase in EITC generosity. The model would suggest that if people lack information and learn about the EITC with a delay, the employment responses to changes in the EITC should be concentrated among those who were previously employed and potentially learned about the program, and might occur with a year lag, after households observe the increased EITC generosity and update their perceptions of the returns to work.

As noted earlier, there is an established literature evaluating the role of information in the context of labor supply responses to the EITC (Chetty & Saez, 2013; Chetty, Friedman, & Saez, 2013). This information considered in this paper varies from the previous work in several important ways. First, the previous work focuses on intensive margin responses to information during the mid-2000s, when the overall economic climate and knowledge of the EITC is markedly different than in the 1990s. Unlike the previous work, which has focused on information about an individual's location on the EITC schedule, this work explores the role of information and learning associated with changes in EITC policy. Optimal responses are also likely to change in response to the policy changes. The information exploited here reflects variation in the timing of learning while the previous work has addressed salience, which might affect decisions differently. Although

[13] A majority of EITC recipients used a tax preparer (Nichols & Rothstein, 2015), who was likely aware of the EITC and shared some level of information with the filer.

the policy changes explored in this paper are several decades old, and overall knowledge of the program has changed, examining this type information can help identify the interaction between information and new policy creation.

4.3 Data

Testing these predictions requires frequent, within-year observations of labor supply. Tax data only provides annual measures, while administrative unemployment insurance data is only quarterly. To approximate these high frequency decisions, I exploit the monthly panel nature of the CPS obtained through IPUMS (Flood et al., 2015). Much of the previous work has relied on annual repeated cross-sections of the CPS and ignored the short panel nature of the survey. Since 1953, the CPS has conducted repeated interviews with households (Rivera Drew, Flood, & Warren, 2014). Each household is interviewed for four consecutive months (a survey wave), rotated out of the sample for eight months, and then re-enters the survey for four consecutive months.[14] In theory, each household will be interviewed for the same four consecutive months two years in a row (e.g., January-April in both 1993 and 1994). During each monthly survey round, participants are asked about hours worked and employment status in the previous week, making it possible to create a four month employment history for each participant at the same point in two consecutive years. In the ASEC March supplement, participants are also asked about the total number of weeks worked during the previous calendar year, allowing researchers to look at annual outcomes.

Previous work has documented the concerns and constraints associated with linking the CPS (Lefgren & Madrian, 2000) as well as identified more accurate ways of linking individuals (Rivera Drew et al., 2014). I follow the practices explained in these papers to link participants across months and across ASEC supplements from year to year.[15] As seen in Table 4.1, the linked

[14] For an in-depth description of the design, panel nature, and linking methods of the CPS, see Lefgren & Madrian (2000) or Rivera Drew et. al, (2014).

[15] Following Lefgren and Madrian (2000) I preserve matches across months as valid if the individual's sex, race, and age (within two years) is consistent across all months. To further improve match quality, I then use detailed industry information, education, and number of children to see if invalid matches are the same along these characteristics. If these characteristics are exactly the same, I keep these matches as any difference in sex, race, or age is likely a coding error. The results are virtually the same if these additional matches are dropped.

sample of single women is similar to the full sample, although slightly more positively selected. The linked sample is slightly more educated, more likely to be Non-Hispanic white, and more attached to the labor force. In Appendix Table A4.1 I reweight the linked sample to look like the full sample, and show that the pattern of results is similar and the coefficients are only slightly smaller.

I use the household roster to determine how many EITC eligible children each woman has in the household during any month. Current college enrollment status is not available in the monthly survey, so only children ages 0-18 are counted as EITC eligible children. For each individual I define the number of EITC eligible children as the maximum number of eligible children reported at any time during the sample period.[16] I then restrict my sample to single mothers who were 19 or older and less than 45 during the survey months.[17] Using the survey year and number of eligible children, I combine the CPS data, with historic federal EITC parameters from the Tax Policy Center (2015) which vary with the number of children and across years to determine the maximum credit each woman could receive. This is converted to 2010 dollars using the personal consumption index from the Bureau of Economic Analysis. This EITC measure captures the height of the plateau, and is a function of the number of qualifying children and current program parameters, not individual household income.

Monthly Data. In the monthly data, women who are first surveyed between October and December will potentially face different EITC generosity across months within a survey wave. Rather than make assumptions about how this affects decisions, I limit the sample to single women that entered the CPS between January and September so that the EITC schedule is consistent for all months in a survey wave.[18] The sample is restricted to less educated single women that entered the CPS from January 1989 to May of 1994. In 1994 there was an institutional change

[16] This removes changes in the maximum EITC that are due to changes in household composition, such as births.

[17] Previous work also excludes women who are enrolled as full time students, ill or disabled, or had positive income but zero hours worked. This level of information is not available in the monthly basic interview and I cannot make these restrictions. This will likely introduce measurement error into my estimation.

[18] The results are similar when women whose survey wave crossed years are included.

in the survey format, making it infeasible to link households across months from the latter half of 1994 and 1995. This leaves monthly data for 14,476 single women with a high school degree or less. As seen in Figure 4.2, during this period there was one moderate expansion in 1991 for households with children, small increases for households with children in 1992 and 1993, a very large expansion in 1994 for all households (which increased with the number of children), and a large expansion in 1995 for households with two or more children.[19]

The month to month information captures several short term labor market outcomes not available at the annual level. I create an indicator for whether or not the woman is employed at any point during the four months surveyed in the year, to see if the probability of being employed changed. I then calculate the share of months during the four month period the woman was employed.[20] I create dummy variables that indicate if the woman ever exited employment (was employed in the previous month but is currently not employed) during the four month period.[21]

Annual Data. Using the annual data, I can also construct measures to capture employment during the year. The annual number of weeks worked is only available in the CPS ASEC supplement. As this supplement was conducted only in March over my sample period, only single mothers interviewed in March for two consecutive years will be included. For this analysis the sample includes 6,919 single women with a high school degree or less that enter the CPS between 1989 and 1994 and can be linked from one March supplement to the next. The number of weeks

[19]Unfortunately, changes in the CPS methodology prevent linking households between 1995 and 1996, when some of the largest increases in the EITC for single mothers with multiple children occurred. Results are similar if I exclude the cohort surveyed in 1994 and 1995 that experienced the two largest expansions.

[20]Participants are asked about work in the previous week. As such, it is possible that women have simply found a new job by the next month. During this sample period, participants of the March ASEC who were currently unemployed, were asked about the duration of their unemployment. Among my sample, nearly 75 percent of unemployment spells were four or more weeks. Although this does not directly relate to the number of individuals that would become unemployed and re-employed during a one month period, it does suggest that this is a small fraction of individuals. Also, if I examine employment in the same occupation from one month to the next, the results are virtually the same.

[21]These measures capture the same patterns observed in the previous literature. When looking at the probability of being ever employed during the four month period in the traditional difference in differences comparison of high school or less single mothers with one qualifying child and multiple before and after 1993, the estimated coefficient is 0.051 (s.e. = 0.014), consistent with the previous literature.

reported correspond to the previous calendar year and all data is appropriately lagged. I also construct an annual employment indicator (weeks worked greater than zero). The impacts on annual entry and exit can be estimated by conditioning the sample on employment status linked from the initial year. Because annual outcomes are reported with a one year lag, the data corresponds to employment between 1988 and 1994.

4.4 Graphical Evidence

Using this data, I first descriptively explore the relationship between EITC generosity and labor force attachment among single women with a high school degree or less in Figure 4.4. Using the four month CPS employment history, I construct the share of months employed during each survey year for each individual in the sample. I then plot the average share of months employed in solid black for women with children relative to women without children. These plots are presented separately for women with any reported work in the first survey year in Panel A and for women without reported work in the first survey year in Panel B. For reference, the maximum EITC a single woman with two children is eligible to receive in the current year is plotted in the background. This measure captures the lagged EITC generosity, and is associated with the EITC policy the individual would encounter when they file their taxes in that year. There was a \$304 (2010\$) increase in 1992, gradual increases in 1993 and 1994, and a \$1,312 increase in 1995 (from the 1994 expansion). During this period, households without children were ineligible for the EITC until it was first available in 1995 at \$407 (2010\$).

When looking at women with any employment during the first survey wave (in Panel A), the gap between women with and without eligible children closes in years when the lagged EITC becomes more generous. This is particularly visible in 1995 when the largest EITC expansion occurs. This is consistent with single women with children responding to the EITC by increasing labor force attachment relative to single women without children. Among women who did not work in the first year, in Panel B, there is no discernible relationship between EITC increases and the share of months employed.

When aggregating up to the annual level a similar pattern emerges. Using the annual weeks worked reported for the previous calendar year from the March Supplement, I construct the individual's employment status for each year. I then plot in black the share of individuals employed in the second survey year for single women with children relative to single women without children in each year. These plots are also presented separately for women with any work in the first survey year in Panel A and for women without work in the first survey year in Panel B with the lagged maximum EITC in the background for reference. Because the weeks worked is reported for the previous calendar year, the sample covers 1989 to 1994, and the largest expansion in 1995 is no longer in the sample. As seen in Panel A, among single women who worked during the first survey year, the share of women with children that were employed during the second year increased relative to women without children when the EITC increased. In other words, more single women with children stayed in the labor force when the EITC expanded, and single women with children became less likely to exit at the annual level, relative to women without children. Among women who did not participate during the first survey wave there is no discernible relationship. Consistent with the monthly data, women with previous labor force attachment appear to become more attached when the lagged EITC they face becomes more generous.

4.5 Identification Strategy

To identify the effect of the EITC on within year employment decisions, I exploit the panel nature of the CPS and estimate an individual fixed effects model that allows women with different numbers of children to have different secular trends. For each woman there are four month employment histories for two consecutive years that I collapse to two observations per person. Given this short, two-period individual-level panel, I am able to examine within person changes in employment outcomes from one year to the next when there has been an expansion in EITC generosity. Because each individual is observed for two periods, I can estimate an individual fixed effects specification as follows

$$Y_{it} = \beta_1 Max\ Credit_{i,t-1} + \beta_2 Max\ Credit_{it} + X_{it}'\Gamma + \theta_n * t + \delta_i + \phi_t + \varepsilon_{it}. \tag{4.6}$$

In equation (4.6), the outcomes of interest are employment measures, constructed from the monthly data such as an indicator for being ever employed, the share of months employed, or an indicator for any exit from employment. This employment outcome is a function of the maximum credit eligible to *receive* in the current year ($Max\,Credit_{i,t-1}$) as well as the maximum credit eligible to *earn* in the current year ($Max\,Credit_{it}$), both measured in hundreds of dollars (2010$). Given the previous discussion on the potential importance of information and the graphical relationships, it seems relevant to look at employment responses to both the current policy parameters as well as one year lagged policies, which might be the parameters people are aware of or learn about when they file their taxes. Both measures can be included simultaneously without concerns about collinearity because EITC generosity does not change uniformly every year. Changes in the current and lagged maximum credit have a low correlation ($\rho = 0.22$ in the monthly data, $\rho = 0.27$ in the annual data).[22] These variables are a function of the number of eligible children and year specific parameters of the EITC schedule.

I include a vector of controls including the federal minimum wage, the state minimum wage, and an indicator for whether a state TANF waiver is in place.[23] As highlighted by previous researchers, single women with zero, one, and two children might be differentially affected by other changes over this time period, so I also include number of eligible children specific linear trends ($\theta_n * t$). Given each individual has two observations, these linear trends by the number of children are equivalent to controlling for level differences in transition rates across eligibility groups. For example, if single women with children are on average more likely to exit from one year to the next, this level difference will be captured by these trends.[24] Individual fixed effects are included to make this a within person comparison and year fixed effects are included to compare households

[22]In Appendix Table A4.2, I estimate the results from Table 4.2 including the credit currently earning and the credit currently receiving separately. The point estimates are not substantively different from the model that includes the measures jointly (Table 4.2), suggesting the result is not driven by high levels of collinearity.

[23]Data on state TANF waivers and minimum wages were graciously provided by Kearney and Levine (2015).

[24]The current specification measures the number of qualifying children during the first survey wave. The results hold if I instead include linear trends by the number of children currently living in the house.

within the same year. Standard errors are corrected for possible clustering at the state level in this and all other specifications.[25] In this analysis there are two observations per person. When looking at the annual data, the March ASEC supplement reports employment from the previous calendar year, so the sample includes women from the 1989 to 1995 surveys, who reported employment from 1988 to 1994.

As seen in Table 4.1, some characteristics of less-educated single women with no children, one child, and two or more children are different. Single mothers are slightly older on average, less educated, and less likely to be Non-Hispanic white. Single mothers with multiple eligible children are even less likely to have a high school degree, more likely to be Non-Hispanic black or Hispanic, and more likely to have a child under five. Among single mothers with multiple eligible children, the average number of children is 2.7, with over 85 percent of these mothers having two or three children.[26] Labor force attachment also falls for single women as the number of eligible children increases. These demographic differences across family size are not inherently problematic to identification, but become so if they result in differential trends across eligibility groups. However, the empirical strategy described above allows for potentially different linear trends by the number of eligible children. As such, the identifying assumption is that differential changes in EITC generosity across eligibility groups are uncorrelated with other potential unobserved factors that affect employment behavior. Although this individual fixed effects specification accounts for unobserved individual characteristics, one limitation is that only short-run responses can be measured and this strategy does not speak directly to long run changes in the stock of employed women over the course of the decade.

[25]Standard errors are smaller if I do not correct for clustering, and are similar if I cluster at the family size by income bin level.

[26]When splitting by employment during the first year, the distribution of the number of children is similar, and excluding mothers with high ordered number of children does not affect the results substantially.

4.6 Results

4.6.1 Within Year Employment Response to EITC Increases

As suggested by the conceptual framework and graphical evidence, an increase in the EITC might increase the opportunity cost of leaving the labor force, leading to stronger labor force attachment within a year. To explore within year labor force attachment, I turn to the linked monthly data. As responses to increases in the EITC might vary by initial labor force attachment (as seen in Figure 4.4), I stratify the estimation by labor force participation during the initial survey wave and report the results in Table 4.2.[27] For example, women who were never employed might face very high employment costs and even with a wage subsidy such as the EITC, the positive returns to work might fall short of these high costs. Alternatively, if information about the policy is only revealed when taxes are filed, women who were not employed previously will not receive the credit or new information.

All employment responses are concentrated among women who were previously employed and are associated with changes in the maximum EITC eligible to *receive*, rather than the maximum EITC eligible to *earn*. In other words, only women who reported employment in the initial survey year and would be more likely to received the credit responded. This pattern is consistent with both a liquidity constraint framework as well as an information framework where single mothers might be unfamiliar with tax policies that affect them but adjust behavior after information about the current policy is revealed.

Among these women, a one hundred dollar increase in the maximum EITC eligible to receive increased the probability of being employed by 1.0 percentage point (column 1), meaning these women were more likely to be employed during the four months window after an EITC expansion. As seen in column (2), a one hundred dollar increase in the maximum EITC also increased the share of months worked by one percentage point, or approximately 1.2 percent (.01/.86). Meaning

[27] I cannot observe annual employment status for all participants as not all women were surveyed for the ASEC supplement. For the monthly data I measure employment during the first four survey months to proxy for annual employment. Women reported as non-employed might be misclassified if they worked later in the year.

there is a small, but significant increase in labor force attachment. As this increase is small, there is not a detectable decrease in the probability of exiting, suggesting many of these women still eventually transitioned out of the labor force. However, increasing employment duration does reduce the frequency of turnover and the probability of exiting multiple times within a survey wave, once again indicative of stronger labor force attachment.[28]

Between 1994 and 1995 the maximum EITC a household with children was eligible to receive (the 1993 to 1994 expansion) increased by just over one thousand dollars on average. These estimates would imply that this increase in the EITC increased the probability of working in the next year by 10 percentage points, and increased the share of months worked by nearly 12 percent. The magnitude of this effect is consistent with about 40 percent of single women with children who worked during the first survey wave working one extra month in the second year after the expansion. The data suggests that the EITC induced single women to work more months, leading to stronger labor force attachment and potentially longer employment spells.[29]

4.6.2 Robustness

Most households with earned income above the EITC schedule should be unaffected by increases in EITC generosity, although some households might become eligible. In Table 4.3, I partition the sample of previously employed women with a high school degree or less into groups based upon whether the reported family income from the previous 12 months falls within the corresponding EITC schedule. This income measure is endogenous and likely exhibits considerable measurement error, as households might contain several families or tax filing units, and should therefore be interpreted with caution. The employment effects are only observed among women

[28]The CPS does allow individuals to report broad categories for why they are unemployed (e.g., laid off, fired, voluntarily left). However, among individuals transitioning out of employment, over 58 percent are reported as not in the labor force in the following month and do not specify why.

[29]Previous work suggests self-employed workers are the most able to adjust income to EITC incentives (Saez, 2010). Looking at self-employment suggests a hundred dollar increase in the maximum EITC credit increased the probability of being self-employed during the four months by 0.5 percentage points. One might be concerned that this potential job stability is preventing upward mobility. However, the average Occupational Prestige Score (Davis et al, 1991) for occupational changes does not change when the EITC becomes more generous, suggesting the EITC is not preventing upward job moves.

that reported EITC eligible income levels suggesting this is not simply capturing a trend faced by all single women with a high school degree or less. For these women, an additional hundred dollars of EITC credit received increased the probability of being employed during the four month period by 1.4 percentage points, increased the share of months employed by 17 percent (.014/.82), and led to a reduction in multiple exits during the four month period. As expected, women above the EITC phase-out region did not significantly adjust the probability of employment, share of months worked or exit behavior.[30] As further evidence that identification is not capturing more general aggregate trends, single mothers with a bachelor's degree, who are generally located higher in the income distribution, do not respond to changes in EITC generosity (see Table 4.4).[31]

4.6.3 Annual Employment Response to EITC Increases

The monthly data suggest that increased EITC generosity leads to stronger labor force attachment and potentially longer employment duration. I next turn to the linked ASEC CPS to see how this affects annual outcomes as reported in Table 4.5. As with the monthly data, I stratify the estimation by reported employment during the first ASEC survey wave. Once again, employment decisions respond to changes in the maximum credit eligible to *receive* in the current year rather than the maximum credit eligible to *earn* in the current year. Women who did not work during the initial survey year did not significantly respond to increases in EITC generosity. However, among women who previously worked, a hundred dollar increase in the maximum EITC credit eligible to receive is associated with 0.83 additional weeks of work a year on average. This

[30] I have also look separately at households in the phase-in, plateau, and phase-out regions of the EITC schedule. The samples become quite small, but the effects are largest and significant among households in the phase-in and phase-out regions.

[31] The Survey of Income and Program Participation (SIPP) also provides monthly employment histories. The 1990-1993 cohorts were surveyed every four months for 32 to 40 months and recalled employment status for the previous four months. Researchers have documented considerable recall bias and "seam" effects in the SIPP (Martini, 2002). After 1993, the SIPP was redesigned and the next cohort began in 1996. Using the 1990 through 1993 cohorts, I construct similar labor supply measures and estimate equation (4.6). The SIPP sample is considerably smaller, $n = 1,807$ for the sample of single women with a high school degree or less and working in the first survey year. On this smaller sample I get the same pattern of results, but the coefficients are attenuated and very imprecise. Smaller estimates and imprecision might be due to recall bias measurement error, the smaller sample, and the fact that the sample period does not correspond to the largest increases in EITC generosity.

increase in weeks of work leads to a significant 1.5 percentage point increase in the probability of working at least 10 weeks, a 2.1 percentage point increase in the probability of working 30 weeks, and 2.1 percentage points increase in the probability of working at least 40 weeks.[32]

To understand how this affects annual labor force transitions, I examine the impact of the EITC on any annual employment in Table 4.6. On average, the EITC expansions in the early 1990s did not significantly change the employment stock among less educated single women. However, when stratified by employment status in the initial survey year, increases in the maximum EITC eligible to receive increase the probability that those previously employed continued employment in column (2) (i.e., reduced annual exit), but did not induce single women without previous participation in column (5) to enter the labor force (although the standard errors are large). A hundred dollar increase in the maximum EITC credit received reduced year-to-year exit by 2.5 percentage points. During this sample period from 1988 to 1993, the largest year-to-year expansion was $365 (2010$), implying that at most annual exit was reduced by 9.1 percentage points. In columns (3) and (4) I separately look at women by labor force attachment during the first survey year. The reduction in annual exit associated with the EITC is concentrated among women who worked less than 52 weeks during the year and were initially less attached to the labor force. This is consistent with women facing lower net benefits from work becoming more attached after the EITC becomes more generous. This reduction in exit provides one channel through which the EITC expansions in the 1990s might have increased the annual employment stock of less-educated single women.

4.6.4 Information vs. Liquidity Mechanism

Both the monthly and annual data provide evidence that increases in the lagged EITC (the EITC received in the current year) led to stronger labor force attachment, more weeks worked, and less annual exit. This is consistent with the information mechanism proposed by Nichols and

[32]Women that worked less than full year and were only weakly attached to the labor force, might be more marginal and responsive to changes in the EITC. In Appendix Table A4.3 I estimate the impact separately for women who reported 1-51 weeks of work and exactly 52 weeks of work during the initial survey year. These effects are concentrated among women who worked less than full year and were less attached to the labor force. Approximately 35 percent of working women worked less than full year.

Rothstein (2015), but actually receiving the EITC transfer could also affect decisions by relaxing liquidity constraints that would otherwise induce women to drop out of employment. Previous work has noted that EITC recipients file taxes and receive refunds sooner than other tax payers (Barrow & McGranahan, 2000; LaLumia, 2013; Hoynes, Miller & Simon, 2015). As seen in Figure 4.5, in 1989 15.6 percent of refund dollars with EITC payments were distributed in February, 40.2 percent in March, and 23.4 percent in April, and 14.4 percent in May, meaning that by May, nearly 94 percent of all refund dollars with an EITC had been distributed. The literature also suggests that EITC recipients spend the money rather quickly, using their refund to pay down bills and fund the purchase of large durable goods, while saving is less common (Smeeding et al., 2000; Goodman-Bacon & McGranahan, 2008). In years when the maximum EITC credit increased, women surveyed during tax season (February through May) might have both new information and extra cash on hand, while women surveyed later in the year might have new information, but were less likely to still have the additional cash on hand. To determine if a response to lagged EITC policy is more likely due to information or liquidity, I test to see if the employment response changes for women surveyed during tax season as follows

$$Y_{it} = \beta_1 Max\ Credit_{i,t-1} + \beta_2 Max\ Credit_{i,t-1} * Share_i + X'_{it}\Gamma + \theta_n * t + \delta_i + \phi_t + \varepsilon_{it} \tag{4.7}$$

where $Share_i$ equals the share of months individual i was interviewed during the tax season (February through May). The purpose of this estimation is to see if β_2 is positive and statistically significant. An insignificant β_2 would suggest that women, whose survey period overlapped with tax season, responded similarly to women surveyed throughout the year, suggesting the response is not due to liquidity. However, as this infusion of income might also introduce income effects, these results are suggestive and should be interpreted with caution. As seen in Table 4.7 the probability of being employed and the share of months employed increased when the EITC expanded, but the additional effect for those surveyed during the tax season is negative (but insignificant).[33] Rather

[33] The observation that individuals interviewed during tax season might be less responsive would be consistent with income effects or the unemployment spell pattern documented by LaLumia (2013). Results are similar if instead the model includes an indicator that equals one if a woman is surveyed during the tax season.

than observing larger effects during tax season due to liquidity and relaxed constraints, the data suggest that survey participants throughout the year responded to the EITC, even those who are unlikely to have extra cash on hand. This is less consistent with a framework where the EITC refund relaxes liquidity constraints, allowing women to work, but is perhaps suggestive that EITC receipt revealed information about the returns to work.

4.6.5 Relationship in Current Conditions

Since the 1990s, the economy has undergone drastic changes and overall awareness of the EITC has also greatly expanded (Chetty & Saez, 2013; Chetty et al., 2013), making it unclear if we should expect the same behavioral responses in today's economy. Since 1996 there has only been one federal expansion of the EITC, in 2009 as part of the post-recession American Recovery and Reinvestment Act (ARRA), leaving little variation at the federal level. However, in the 2000s, many states adopted statewide EITC policies. In most cases, these EITC policies are calculated as a percentage of the federal credit. In 1999, only 8 states had an EITC in place, but by 2016 over half of the states had these policies in place. There is also significant heterogeneity across states, with some states distributing as little as 3.5 percent of the federal credit up to 85 percent among households in the phase-in region. Using these state level policies, I construct the total maximum credit eligible to earn and receive in the current year, which is composed of both the federal and state level credits. I then estimate equation (4.6) using the sample of single women with a high school degree or less from the 2000 through 2016 monthly CPS surveys. As before, I estimate the effects separately for women by employment status during the initial survey wave. These estimates are reported in Appendix Table A4.4.

During this more recent period, when knowledge of the EITC is more ubiquitous, we observe a slightly different response. As before, women who were already employed in the initial wave responded by increasing the probability of being employed and the share of months employed, but this response was to changes in the maximum EITC eligible to *earn* in the current year. However these responses are an order of magnitude smaller. Although this identifying variation is different (i.e., smaller state level increases) making it not directly comparable, this pattern is consistent with more overall awareness of the policy. As awareness of EITC policies rises, people with previous labor force attachment respond to contemporaneous policy.

4.7 Conclusion

Although less educated single women make frequent employment decisions, the current literature is silent as to how within year employment decisions respond to the EITC. Exploiting the panel nature of the CPS, I provide evidence that the expansions of the EITC in the early 1990s increased employment levels of less educated single women by increasing employment stability and reducing annual exit from the labor force. Increases in the maximum EITC eligible to receive made less educated single women increase the share of months worked, and exit the labor market less frequently. A hundred dollar increase in the maximum EITC credit increased the number of weeks worked in a year by 0.83 weeks and reduced year-to-year exit among single women that were previously employed by 2.1 percentage points. Importantly, employment behavior is responding to changes in the maximum EITC eligible to *receive* rather than the maximum EITC *earned* in the current year and only observed among women who reported employment in the initial year and were therefore likely to receive the credit.

To better understand if the response to receiving the EITC is driven by information or the lack there of I look to see if women surveyed during and after the tax season responded differently. Although women surveyed during the tax season are more likely to experience an increase in both liquidity and information, their employment behavior is not more responsive, suggesting this response is not driven by relaxed liquidity constraints. This paper provides evidence that transfer programs designed like the EITC can increase employment stability of less educated single women, which might have important welfare implications if, for example, labor market experience at this level increases wages or if employment stability leads to positive outcomes for children. However, as these employment responses appear to be influenced by knowledge or information about the policy, it is important that mechanisms be in place for potential recipients to learn or know about the program and how it affects the returns to work. This work also sheds light on the dynamics driving the documented effect of the EITC on employment levels. Often the EITC is cited as a tool to bring households into the labor force, but these results would suggest that the EITC operates by keeping individuals with previous labor force attachment in the labor force, rather than by bringing in new entrants.

Table 4.1: Characteristics and Sample Selection of Linked Singer Mothers with High School or Less in the 1989-1994 CPS

	Single Women 19-44 with High School or Less					
	0 Eligible Children		1 Eligible Child		2+ Eligible Children	
	Full Sample	Linked Sample	Full Sample	Linked Sample	Full Sample	Linked Sample
	(1)	(2)	(3)	(4)	(5)	(6)
Age	26.26	28.49	28.86	30.91	30.17	31.30
Less than High School	0.22	0.22	0.29	0.25	0.40	0.40
Non-Hispanic White	0.66	0.68	0.56	0.55	0.40	0.37
Non-Hispanic Black	0.17	0.17	0.28	0.33	0.42	0.46
Hispanic	0.13	0.11	0.13	0.10	0.15	0.15
Non-Hispanic Other	0.03	0.03	0.02	0.01	0.02	0.02
Number of Eligible Children	0	0	1	1	2.63	2.68
Any Children under 5	0	0	0.49	0.31	0.60	0.54
During Survey Months in Year						
Ever Employed	0.78	0.80	0.65	0.71	0.51	0.53
Ever Enter	0.10	0.08	0.09	0.09	0.09	0.07
Ever Exit	0.10	0.09	0.10	0.08	0.08	0.07
Ever Enter or Exit	0.16	0.13	0.15	0.13	0.14	0.11
Share of Months Continue Employment	0.71	0.74	0.58	0.65	0.44	0.47
Observations	16,423	7,025	6,686	3,293	7,639	4,158

Notes: Data from women who entered the CPS between January 1989 and May 1994. The Full Sample is restricted to the second survey round of single women who were 19-44 and had a high school degree or less, and who entered the CPS between January and September. Sample weighted using population weights provided by the CPS. The Linked Sample is the subset of women who can be linked across waves to their sixth survey round and corresponds to the analysis sample. For reference, the maximum EITC eligible to receive among households with children increased by two hundred dollars (2010$) on average across all years, and by approximately one thousand dollars between 1994 and 1995.

Table 4.2: Labor Force Attachment Response to EITC Increases

	Ever Employed (1)	Share Months Employed (2)	Ever Exit Employment (3)	Multiple Exits (4)
Panel A.	Reported Any Employment During First 4 Monthly Surveys			
Max Credit$_{t-1}$ *($100s)*	0.010**	0.010**	0.0003	-0.001*
(Credit Eligible to Receive)	(0.005)	(0.004)	(0.005)	(0.0004)
Max Credit$_{t}$ *($100s)*	-0.000	-0.002	-0.006	-0.0001
(Credit Eligible to Earn)	(0.003)	(0.002)	(0.004)	(0.0005)
Second Wave Dependent Mean	0.91	0.86	0.07	0.001
Panel B.	Reported No Employment During First 4 Monthly Surveys			
Max Credit$_{t-1}$ *($100s)*	-0.003	-0.002	-0.002	-0.0002
(Credit Eligible to Receive)	(0.006)	(0.004)	(0.003)	(0.0001)
Max Credit$_{t}$ *($100s)*	0.003	0.001	0.003	-0.0002
(Credit Eligible to Earn)	(0.005)	(0.004)	(0.002)	(0.0003)
Second Wave Dependent Mean	0.19	0.13	0.06	0.001

Notes: Data from the linked 1989-1995 monthly CPS. Sample includes two observations for all single women with a high school degree or less between the ages of 19 and 44, and for whom survey waves 1-4 were contained within a single year. There are 20,438 observations (10,219 women) included in Panel A (Any employed during first four months), and 8,514 observations (4,257 women) included in Panel B (no employed during first four months). Changes in the maximum credit do not occur every year and the within sample correlation coefficient between changes in the credit currently earning and the credit received this year is 0.22. The maximum credit is converted to real 2010 dollars using the personal consumption index and measured in hundreds of 2010 dollars. Controls include the federal and state minimum wage, an indicator for a TANF waiver, and linear trends by the number of qualifying children during the first survey wave. Individual and year fixed effects are included. Regressions weighted using population weights provided by the CPS. Standard errors are clustered at the state level. *** p<.01, ** p<.05, * p<.1.

Table 4.3: Sensitivity Check: Response by Reported Household Income in Previous Year

	Ever Employed (1)	Share Months Employed (2)	Ever Exit Employment (3)	Multiple Exits (4)
Panel A.	Previously Employed Single Women with High School or Less in EITC Range			
Max Credit$_{t-1}$ *($100s)*	0.014***	0.014**	-0.003	-0.002***
(Credit Eligible to Receive)	(0.006)	(0.006)	(0.006)	(0.001)
Max Credit$_{t}$ *($100s)*	0.000	-0.004	-0.005	0.000
(Credit Eligible to Earn)	(0.004)	(0.003)	(0.006)	(0.001)
Second Wave Dependent Mean	0.88	0.82	0.08	0.002
Observations	10,890	10,890	10,890	10,890
	Panel B. Previously Employed Single Women with High School or Less above EITC Range			
Max Credit$_{t-1}$ *($100s)*	0.006	0.006	0.005	-0.0004
(Credit Eligible to Receive)	(0.006)	(0.006)	(0.009)	(0.001)
Max Credit$_{t}$ *($100s)*	-0.0001	-0.001	0.001	-0.001
(Credit Eligible to Earn)	(0.005)	(0.005)	(0.004)	(0.001)
Second Wave Dependent Mean	0.95	0.91	0.06	0.001
Observations	9,548	9,548	9,548	9,548

Notes: Data from the linked 1989-1995 monthly CPS. Sample includes two observations for all single women with a high school degree or less between the ages of 19 and 44, and for whom survey waves 1-4 were contained within a single year. The maximum credit is converted to real 2010 dollars using the personal consumption index and measured in hundreds of 2010 dollars. Households are sorted into regions on the EITC schedule using household income and EITC parameters from the initial survey year. Household income is reported in bins, so they are assigned (imperfectly) based on where the midpoint of the bin falls on the EITC schedule. Controls include the federal and state minimum wage, an indicator for a TANF waiver, and linear trends by the number of qualifying children during the first survey wave. Individual and year fixed effects are included. Regressions weighted using population weights provided by the CPS. Standard errors are clustered at the state level. *** p<.01, ** p<.05, * p<.1.

Table 4.4: Sensitivity Check: Response of Previously Employed Single Women with a College Degree

	Ever Employed (1)	Share Months Employed (2)	Ever Exit Employment (3)	Multiple Exits (4)
	Previously Employed Single Women with College Degree			
Max Credit$_{t-1}$ *($100s)*	0.003	0.002	0.003	0.000
(Credit Eligible to Receive)	(0.003)	(0.005)	(0.008)	(0.001)
Max Credit$_{t}$ *($100s)*	-0.002	-0.001	-0.002	-0.00002
(Credit Eligible to Earn)	(0.003)	(0.004)	(0.006)	(0.0004)
Second Wave Dependent Mean	0.97	0.95	0.04	0.001
Observations	12,708	12,708	12,708	12,708

Notes: Data from the linked 1989-1995 monthly CPS. Sample includes two observations for single women with a college degree between the ages of 19 and 44, and for whom survey waves 1-4 were contained within a single year, and reported any employment during the initial 4 monthly surveys. The maximum credit is converted to real 2010 dollars using the personal consumption index and measured in hundreds of 2010 dollars. Controls include the federal and state minimum wage, an indicator for a TANF waiver, and linear trends by the number of qualifying children during the first survey wave. Individual and year fixed effects are included. Regressions weighted using population weights provided by the CPS. Standard errors are clustered at the state level. *** p<.01, ** p<.05, * p<.1.

Table 4.5: Within Individual Annual Labor Supply Response to EITC Increases

	Weeks Worked (1)	Weeks Worked $\geq$ 10 (2)	Weeks Worked $\geq$ 20 (3)	Weeks Worked $\geq$ 30 (4)	Weeks Worked $\geq$ 40 (5)	Weeks Worked $\geq$ 50 (6)
Panel A.	Reported Positive Weeks Worked During Initial Survey Year					
Max Credit$_{t-1}$ ($100s)	0.832***	0.015**	0.009	0.021**	0.021***	0.015
(Credit Eligible to Receive)	(0.264)	(0.007)	(0.008)	(0.008)	(0.006)	(0.010)
Max Credit$_t$ ($100s)	-0.138	-0.004	-0.001	-0.010*	-0.003	0.003
(Credit Eligible to Earn)	(0.199)	(0.004)	(0.004)	(0.005)	(0.005)	(0.007)
Second Wave Dependent Mean	42.6	0.89	0.86	0.81	0.77	0.69
Panel B.	Reported No Weeks Worked During Initial Survey Year					
Max Credit$_{t-1}$ ($100s)	0.027	-0.0001	0.001	0.001	-0.0001	-0.006
(Credit Eligible to Receive)	(0.647)	(0.017)	(0.016)	(0.009)	(0.009)	(0.011)
Max Credit$_t$ ($100s)	0.025	0.002	-0.002	0.002	-0.002	0.003
(Credit Eligible to Earn)	(0.150)	(0.004)	(0.003)	(0.003)	(0.004)	(0.006)
Second Wave Dependent Mean	6.5	0.18	0.14	0.10	0.09	0.07

Notes: Data from the linked 1989-1994 ASEC March CPS supplement. Sample includes one observation for all single women with a high school degree or less between the ages of 19 and 44, who were interviewed for two ASEC supplements. There are 10,052 observations (5,026 women) included in Panel A (any work during initial survey year), and 3,786 observations (1,893 women) included in Panel B (no work during initial survey year). This sample is smaller than the monthly sample (Table 4.2) because not every woman is surveyed in March. Changes in the maximum credit do not occur every year and the within sample correlation coefficient between the change in the credit currently earning and the credit received this year is 0.27. The maximum credit is converted to real 2010 dollars using the personal consumption index and measured in hundreds of 2010 dollars. Controls include the change in the federal and state minimum wages,an indicator for a TANF waiver, linear trends by the number of qualifying children during the first survey wave. Individual and year fixed effects are included. Regressions weighted using household population weights provided by the CPS ASEC. Standard errors are clustered at the state level. *** p<.01, ** p<.05, * p<.1.

Table 4.6: Effect of EITC Increases on Annual Labor Market Transitions

	Outcome: Any Employment				
		Any Work During Initial Survey Year			No Work During
	All	Any	< 52 weeks	52 weeks	Initial Survey Year
	(1)	(2)	(3)	(4)	(5)
Max Credit$_{t-1}$ *($100s)*	0.011	0.025***	0.047**	0.007	0.001
(Credit Eligible to Receive)	(0.008)	(0.009)	(0.020)	(0.006)	(0.022)
Max Credit$_t$ *($100s)*	0.002	0.002	0.008	-0.002	0.005
(Credit Eligible to Earn)	(0.003)	(0.003)	(0.009)	(0.003)	(0.005)
Second Wave Dependent Mean	0.73	0.91	0.81	0.97	0.23
Observations	13,838	10,052	3,474	6,578	3,786

Notes: Data from the linked 1989-1994 ASEC March CPS supplement. Sample includes one observation for all single women with a high school degree or less between the ages of 19 and 44, who were interviewed for two ASEC supplements. This sample is smaller than the monthly sample (Table 4.2) because not every woman is surveyed in March. Changes in the maximum credit do not occur every year and the within sample correlation coefficient between the credit currently earnings and the credit received this year is 0.27. The year-to-year change in annual employment captures annual level exit in column (2) and annual level exit in column (3). The maximum credit is converted to real 2010 dollars using the personal consumption index and measured in hundreds of 2010 dollars. Controls include the race indicators, age, age2, an indicator for any children under 5, the change in the federal and state minimum wages, and an indicator for the introduction of a TANF waiver. Fixed effects for the year, state, and number of qualifying children are included. Regressions weighted using household population weights provided by the CPS ASEC. Standard errors are clustered at the state level. *** p<.01, ** p<.05, * p<.1.

Table 4.7: Information vs. Liquidity: Differential Impacts for Previously Employed Women Surveyed During Tax Season

	Ever Employed (1)	Share Months Employed (2)	Ever Exit Employment (3)	Multiple Exits (4)
Max Credit$_{t-1}$ *($100s)*	0.011**	0.012**	-0.0004	-0.0014*
(Credit Eligible to Receive)	(0.005)	(0.005)	(0.006)	(0.001)
Max Credit$_t$ *	-0.001	-0.003	-0.001	0.001
Share of Surveys Feb-May	(0.002)	(0.003)	(0.003)	(0.001)
Second Wave Dependent Mean	0.91	0.86	0.07	0.001
Observations	20,438	20,438	20,438	20,438

Notes: Data from the linked 1989-1995 monthly CPS. Sample includes two observations for all single women with a high school degree or less between the ages of 19 and 44, and for whom survey waves 1-4 were contained within a single year, and reported any employment during the initial 4 monthly surveys. The maximum credit is converted to real 2010 dollars using the personal consumption index and measured in hundreds of 2010 dollars. Controls include the federal and state minimum wage, an indicator for a TANF waiver, and linear trends by the number of qualifying children during the first survey wave. Individual and year fixed effects are included. Regressions weighted using population weights provided by the CPS. Standard errors are clustered at the state level. *** p<.01, ** p<.05, * p<.1.

Figures

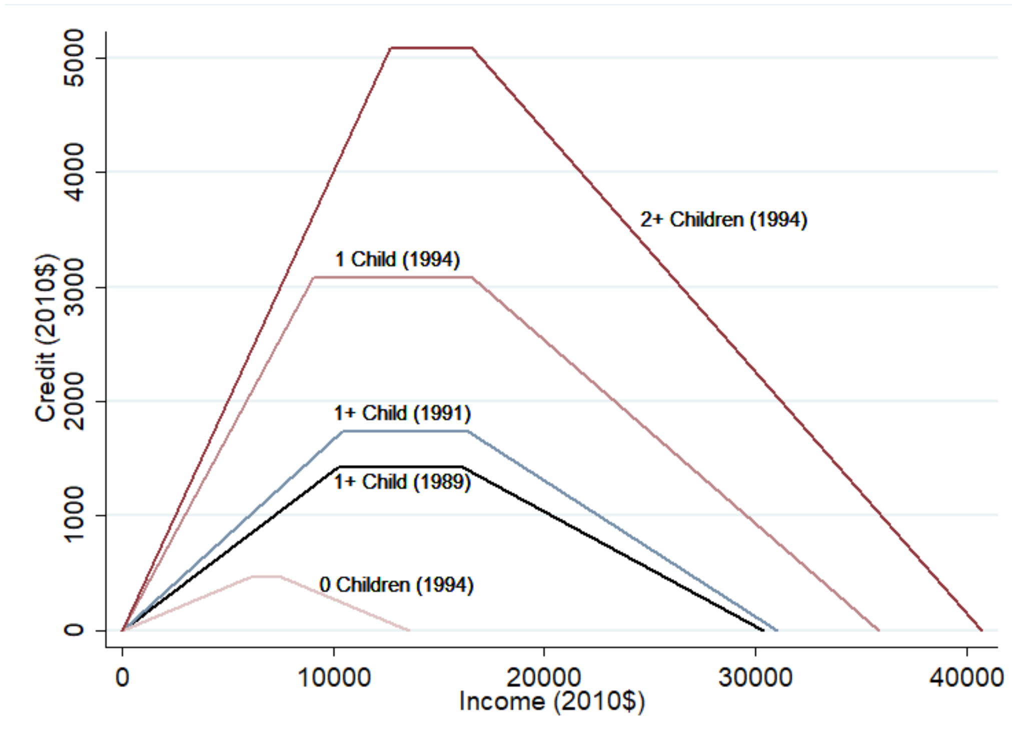

Figure 4.1: EITC Schedule Parameters over Time for Various Family Sizes
Source: Author's calculations using EITC parameters for 1989, 1991, and 1994 provided by Tax Policy Center.

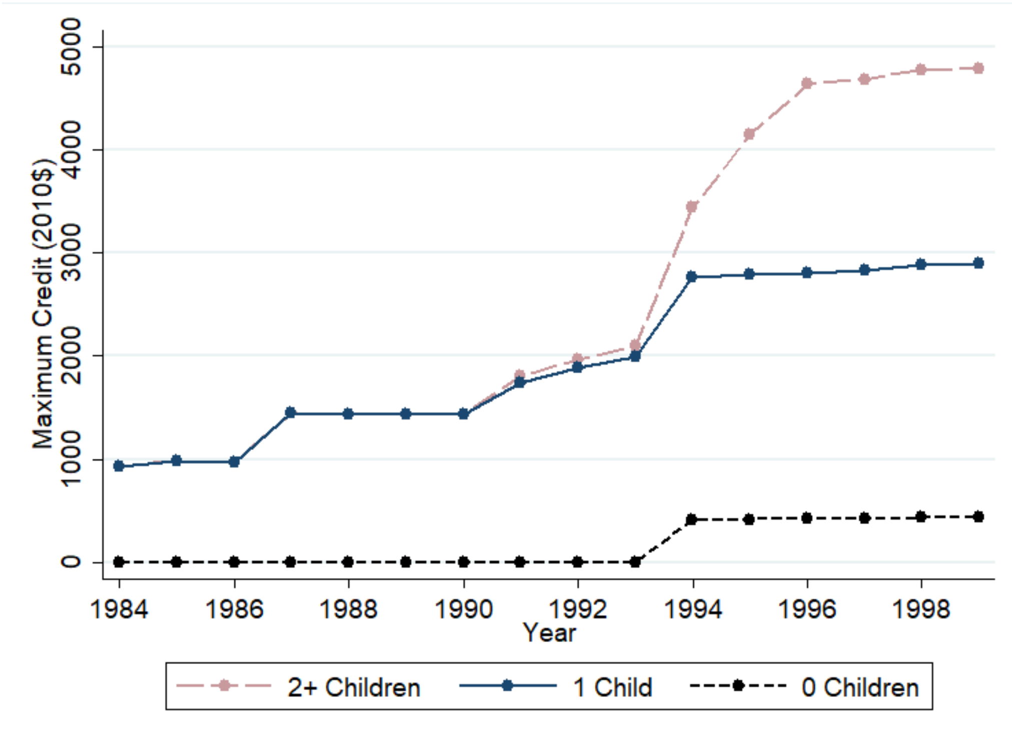

Figure 4.2: Maximum EITC Credit by Year and Family Size

Source: Author's Calculations using EITC program parameters as provided by the Tax Policy Center.

Figure 4.3: Change in Marginal and Average Subsidy Rates When the EITC Maximum Credit Increases

Notes: The 1994 expansion increased average subsidy rates for households throughout the EITC schedule. The impact on marginal subsidy rates varied across the EITC schedule. For households in the phase-in region the marginal tax rate become more negative, for households in the phase-out region the marginal tax rate became more positive, and for households in the plateau region the marginal tax rates either became larger, smaller or remained constant.

Source: Author's calculations using parameters provided by the Tax Policy Center for a household with two qualifying children from 1993 and 1994.

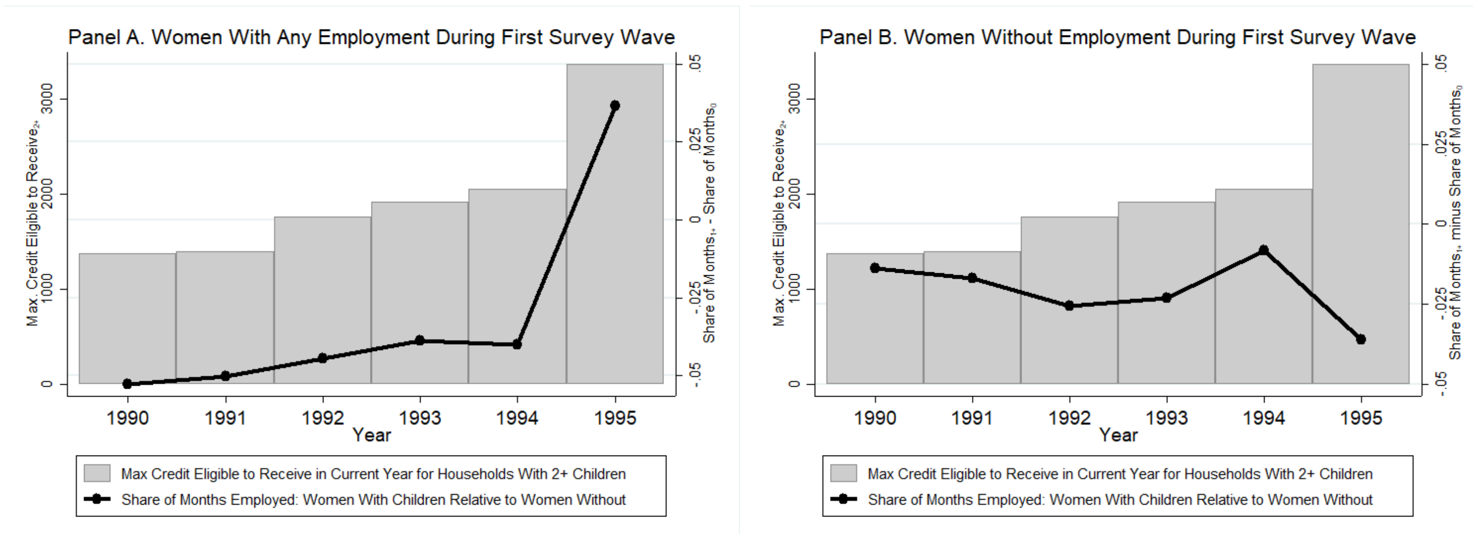

Figure 4.4: Average Share of Months Employed for Single Women with Children Relative to Single Women without Children

Notes: The solid black line plots the share of months employed for single women with a high school degree or less with children relative to women single women with a high school degree or less without children. Averages are weighted using CPS sampling weights. Averages are weighted using CPS sampling weights. The bar graph depicts the maximum credit eligible to *receive* in the current year (lagged EITC policy) for households with two or more eligible children. The maximum credit for households with one eligible child would look similar in most years, but 600 dollars lower in 1995.

Source: Author's calculation using the linked 1989-1995 monthly CPS and EITC program parameters from the Tax Policy Center.

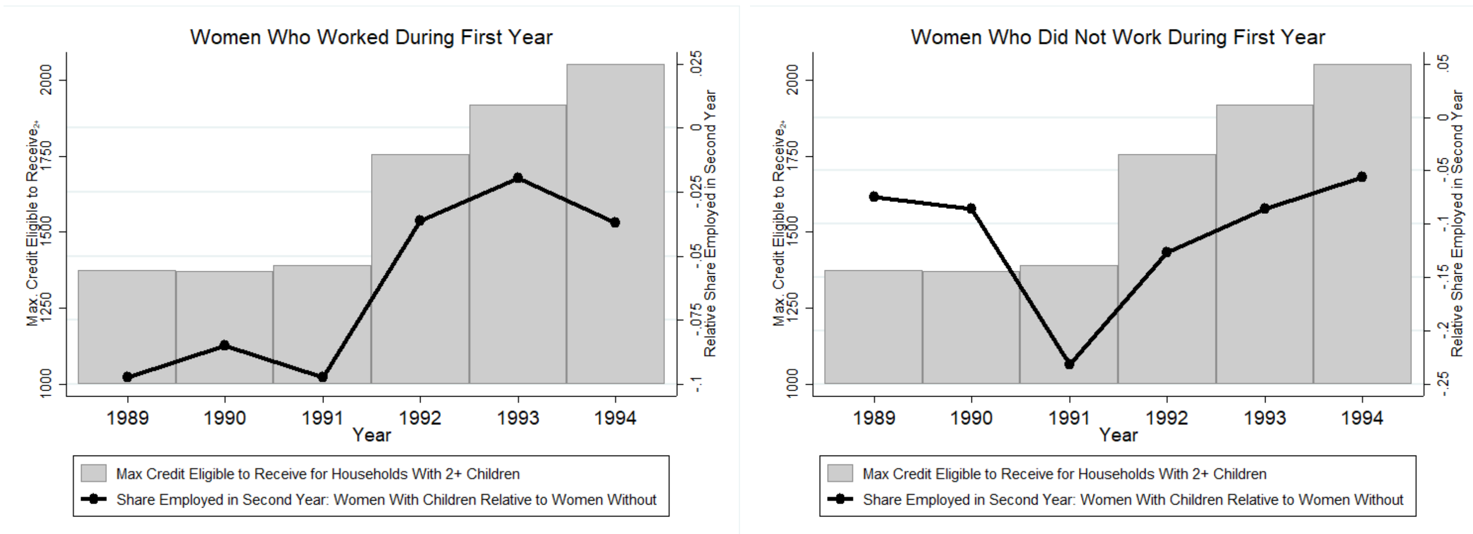

Figure 4.5: Share Employed During Second Year for Single Women with Children Relative to Single Women without Children

Notes: The solid black line plots the share employed in the second year for single women with a high school degree or less with children relative to women single women with a high school degree or less without children. Averages are weighted using CPS sampling weights. The bar graph depicts the maximum credit eligible to *receive* in the current year (lagged EITC policy) for households with two or more eligible children.

Source: Author's calculation using the linked 1989-1995 ASEC CPS and EITC program parameters from the Tax Policy Center.

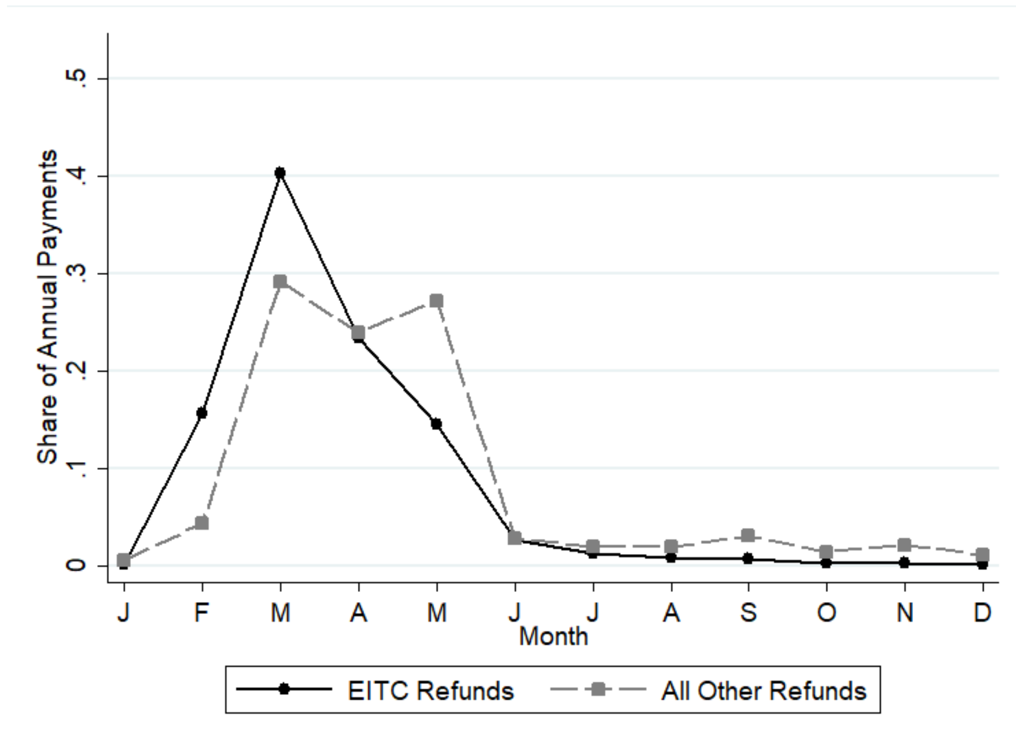

Figure 4.6: Distribution of Tax Refunds by Calendar Month in 1989

Notes: Plots represent the monthly share of the total tax refund payments made during the year for EITC refunds and all other refunds in 1989.

Source: Author's calculation using Monthly Treasury Statements for January 1989 through December 1989.

4.8 Appendix A. Additional Tables and Figures

Table A4.1: Re-weighting to Account for Sample Selection in Linked Analysis Sample

	Ever Employed (1)	Share Months Employed (2)	Ever Exit Employment (3)	Multiple Exits (4)
Panel A.	Reported Any Employment During First 4 Monthly Surveys			
Max Credit$_{t-1}$ *($100s)*	0.008**	0.008**	0.0003	-0.001**
(Credit Eligible to Receive)	(0.004)	(0.004)	(0.004)	(0.0005)
Max Credit$_{t}$ *($100s)*	0.0004	-0.0001	-0.007**	0.0001
(Credit Eligible to Earn)	(0.002)	(0.002)	(0.003)	(0.0004)
Dependent Mean (in levels)	0.91	0.86	0.07	0.001
Panel B.	Reported No Employment During First 4 Monthly Surveys			
Max Credit$_{t-1}$ *($100s)*	0.001	0.001	-0.001	-0.0002
(Credit Eligible to Receive)	(0.006)	(0.004)	(0.003)	(0.0002)
Max Credit$_{t}$ *($100s)*	-0.0001	-0.001	0.002	-0.0002
(Credit Eligible to Earn)	(0.004)	(0.003)	(0.003)	(0.0003)
Second Wave Dependent Mean	0.19	0.13	0.06	0.001

Notes: Data from the linked 1989-1995 monthly CPS. Sample includes two observations for all single women with a high school degree or less between the ages of 19 and 44, and for whom survey waves 1-4 were contained within a single year. The analysis sample is reweighted so that the demographic and employment characteristics in the initial year match those of the full sample in Table 2.1. Observations given zero weight are excluded from the analysis. The maximum credit is converted to real 2010 dollars using the personal consumption index and measured in hundreds of 2010 dollars. Controls include the federal and state minimum wage, an indicator for a TANF waiver, and linear trends by the number of qualifying children during the first survey wave. Individual and year fixed effects are included. Standard errors are clustered at the state level. *** p<.01, ** p<.05, * p<.1.

Table A4.2: Separate Effects of the Credit Currently Earning and the Credit Receiving in the Current Year

	Ever Employed (1)	Share Months Employed (2)	Ever Exit Employment (3)	Multiple Exits (4)
Panel A.	Reported Any Employment During First 4 Monthly Surveys, Credit Currently Receiving			
Max Credit$_{t-1}$ *($100s)*	0.010**	0.009**	-0.001	-0.001*
(Credit Eligible to Receive)	(0.004)	(0.004)	(0.005)	(0.0004)
Panel B.	Reported Any Employment During First 4 Monthly Surveys, Credit Currently Earning			
Max Credit$_{t}$ *($100s)*	0.002	-0.0003	-0.006	-0.0002
(Credit Eligible to Earn)	(0.003)	(0.002)	(0.004)	(0.0005)
Panel C.	Reported No Employment During First 4 Monthly Surveys, Credit Currently Receiving			
Max Credit$_{t-1}$ *($100s)*	-0.002	-0.001	-0.0002	-0.0002
(Credit Eligible to Receive)	(0.007)	(0.005)	(0.003)	(0.0002)
Panel D.	Reported No Employment During First 4 Monthly Surveys, Credit Currently Earning			
Max Credit$_{t}$ *($100s)*	0.003	0.001	0.003	-0.0002
(Credit Eligible to Earn)	(0.005)	(0.004)	(0.002)	(0.0003)

Notes: Data from the linked 1989-1995 monthly CPS. Sample includes two observations for all single women with a high school degree or less between the ages of 19 and 44, and for whom survey waves 1-4 were contained within a single year. There are 20,438 observations (10,219 women) included in Panel A and B (any employed during first four months), and 8,514 observations (4,257 women) included in Panel C and D (no employed during first four months). Changes in the maximum credit do not occur every year and the within sample correlation coefficient between changes in the credit currently earning and the credit received this year is 0.22. The maximum credit is converted to real 2010 dollars using the personal consumption index and measured in hundreds of 2010 dollars. Controls include the federal and state minimum wages, an indicator for a TANF waiver, and linear trends by the number of qualifying children during the first survey wave. Individual and year fixed effects are included. Regressions weighted using population weights provided by the CPS. Standard errors are clustered at the state level. *** p<.01, ** p<.05, * p<.1.

Table A4.3: Within Individual Annual Labor Supply Response to EITC Increases, By Initial Labor Force Attachment

	Weeks Worked (1)	Weeks Worked ≥ 10 (2)	Weeks Worked ≥ 20 (3)	Weeks Worked ≥ 30 (4)	Weeks Worked ≥ 40 (5)	Weeks Worked ≥ 50 (6)
Panel A.	Reported Between 1 and 51 Weeks Worked During Initial Survey Year					
Max Credit$_{t-1}$ ($100s)	2.54**	0.040*	0.025	0.056*	0.069**	0.069**
(Credit Eligible to Receive)	(1.06)	(0.024)	(0.021)	(0.031)	(0.031)	(0.029)
Max Credit$_t$ ($100s)	-0.22	-0.005	-0.003	-0.019**	-0.003	0.007
(Credit Eligible to Earn)	(0.38)	(0.011)	(0.011)	(0.010)	(0.009)	(0.009)
Second Wave Dependent Mean	32.9	0.76	0.70	0.61	0.54	0.43
Observations	3,474	3,474	3,474	3,474	3,474	3,474
Panel B.	Reported 52 Weeks Worked During Initial Survey Year					
Max Credit$_{t-1}$ ($100s)	0.23	0.002	0.003	0.010	0.006	0.003
(Credit Eligible to Receive)	(0.46)	(0.009)	(0.011)	(0.011)	(0.012)	(0.012)
Max Credit$_t$ ($100s)	-0.01	-0.002	0.002	-0.002	-0.001	0.002
(Credit Eligible to Earn)	(0.20)	(0.004)	(0.004)	(0.004)	(0.006)	(0.008)
Second Wave Dependent Mean	47.7	0.96	0.94	0.92	0.89	0.83
Observations	6,578	6,578	6,578	6,578	6,578	6,578

Notes: Data from the linked 1989-1994 ASEC March CPS supplement. Sample includes two observations for all single women with a high school degree or less between the ages of 19 and 44, who were interviewed for two ASEC supplements and reported positive weeks worked in their first ASEC survey. Changes in the maximum credit do not occur every year and the within sample correlation coefficient between the change in the credit currently earning and the credit received this year is 0.27. The maximum credit is converted to real 2010 dollars using the personal consumption index and measured in hundreds of 2010 dollars. Controls include the federal and state minimum wages, an indicator for a TANF waiver, and linear trends by the number of qualifying children during the first survey wave. Individual and year fixed effects are included. Regressions weighted using population weights provided by the CPS. Standard errors are clustered at the state level. *** p<.01, ** p<.05, * p<.1.

Table A4.4: Recent Responses to State and Federal EITC Credit Increases, 1999-2016

	Ever Employed (1)	Share Months Employed (2)	Ever Exit Employment (3)	Multiple Exits (4)
Panel A.	Reported Any Employment During First 4 Monthly Surveys			
Max Credit$_{t-1}$ *($100s)*	-0.0003	-0.001	-0.001	-0.00004
(Credit Eligible to Receive)	(0.002)	(0.001)	(0.002)	(0.0001)
Max Credit$_{t}$ *($100s)*	0.002***	0.002***	0.001	0.000
(Credit Eligible to Earn)	(0.001)	(0.001)	(0.001)	(0.0001)
Second Wave Dependent Mean	0.94	0.86	0.11	0.003
	Panel B. Reported No Employment During First 4 Monthly Surveys			
Max Credit$_{t-1}$ *($100s)*	-0.003	-0.003	-0.001	-0.00002
(Credit Eligible to Receive)	(0.005)	(0.003)	(0.002)	(0.0001)
Max Credit$_{t}$ *($100s)*	-0.00005	-0.001	-0.002***	-0.00004
(Credit Eligible to Earn)	(0.001)	(0.001)	(0.0004)	(0.0001)
Second Wave Dependent Mean	0.12	0.08	0.03	0.001

Notes: Data from the linked 2000-2016 monthly CPS. Sample includes two observations for all single women with a high school degree or less between the ages of 19 and 44, and for whom survey waves 1-4 were contained within a single year. There are 47,312 observations (23,656 women) included in Panel A (Any employed during first four months), and 21,074 observations (10,537 women) included in Panel B (no employed during first four months). Changes in the maximum credit do not occur every year and the within sample correlation coefficient between the credit currently earnings and the credit received this year is -.05. The maximum credit is converted to real 2010 dollars using the personal consumption index and measured in hundreds of 2010 dollars. Controls include the federal and state minimum wages, an indicator for a TANF waiver, and linear trends by the number of qualifying children during the first survey wave. Individual and year fixed effects are included. Regressions weighted using population weights provided by the CPS. Standard errors are clustered at the state level. *** p<.01, ** p<.05, * p<.1.

4.9 Appendix B. Model Derivations

Proposition 1.

$$0 < \frac{\partial a_t^w}{\partial w_t} < 1, \quad \frac{\partial \phi^*}{\partial w_t} = u_t'\big(1 - \frac{\partial a_t^w}{\partial w_t}\big) + \frac{\partial EV(a_t, y_t)}{\partial a_t}\frac{\partial a_t^w}{\partial w_t} > 0 \tag{4.8}$$

Suppose there is no uncertainty and define the value function at time t of a woman who chooses to work in period t as

$$V_t(a_{t-1}; y_{t-1}) = \max_{a_t} u_t(w_t - c_t + a_{t-1} - a_t + \beta V_{t+1}(a_t; y_t) \tag{4.9}$$

First Order Condition: $\frac{\partial V_t}{\partial a_t} = -u_t' + \beta\frac{\partial V_{t+1}}{\partial a_t} = 0 \rightarrow u_t' = \beta\frac{\partial V_{t+1}}{\partial a_t}$

Envelope Condition: $\frac{\partial V_t}{\partial a_{t-1}} = u_t'$

Combine the first order condition and envelop condition (iterated one period forward) to derive the Euler Equation

Euler Equation: $u_t' = \beta u_{t+1}'$

Totally differentiate the Euler Equation and rearrange to determine how the optimal choice of a_t responds to changes in w_t

$$u_t'' dw_t - u_t'' da_t = \beta u_{t+1}'' da_t \tag{4.10}$$

$$\frac{\partial a_t}{\partial w_t} = \frac{u_t''}{u_t'' + \beta u_{t+1}''} \tag{4.11}$$

By the concavity of u, $u'' < 0$ so $0 < \frac{\partial a_t}{\partial w_t} < 1$. Now consider the working cost threshold

$$\phi^* = u_t(w_t - c_t + a_{t-1} - a_t^w) - u_t(b_t + a_{t-1} - a_t^n) + EV(a_t^w, y_t = 1) - EV(a_t^n, y_t = 0). \tag{4.12}$$

Differentiating with respect to w_t yields

$$\frac{\partial \phi^*}{\partial w_t} = u_t'\big(1 - \frac{\partial a_t^w}{\partial w_t}\big) + \frac{\partial EV(a_t, y_t)}{\partial a_t}\frac{\partial a_t^w}{\partial w_t} > 0 \tag{4.13}$$

Differentiating with respect to c_t will yield

$$\frac{\partial a_t}{\partial c_t} = \frac{-u_t''}{u_t'' + \beta u_t''} < 0. \tag{4.14}$$

If the woman decides not to work, the value function becomes

$$V_t(a_{t-1}; y_{t-1}) = \max_{a_t} u_t(b_t + a_{t-1} - a_t) + \beta V_{t+1}(a_t; y_t) \tag{4.15}$$

The optimal wealth level will respond to b_t as it responded to w_t and $0 < \frac{\partial a_t}{\partial b_t} < 1$. However, the value of not working enters the cost threshold negatively so an increase in b_t will have a negative effect on ϕ^*.

Proposition 2.

$$-w_t < \frac{\partial a_t^w}{\partial \tau_t} < 0, \quad \frac{\partial \phi_\tau^*}{\partial \tau_t} = \frac{\partial EV(a_t^w + \tau_t w_t, y_t\!=\!1)}{\partial a_t}\big(\frac{\partial a_t^w}{\partial \tau_t} + w_t\big) > 0 \tag{4.16}$$

Suppose a wage subsidy is now available and transferred with a one period lag, the value function for working becomes

$$V_t(a_{t-1}; y_{t-1}) = \max_{a_t} u_t(w_t - c_t + a_{t-1} - a_t) + \beta V_{t+1}(a_t + \tau_t w_t; y_t) \tag{4.17}$$

and the Euler Equation will be

$$u_t'(w_t - c_t + a_{t-1} - a_t) = \beta u_{t+1}'(a_t + \tau_t w_t). \tag{4.18}$$

Totally differentiating with respect to τ_t and a_t yields

$$-u_t'' da_t = \beta u_{t+1}'' da_t + \beta u_{t+1}'' w_t d\tau_t \tag{4.19}$$

$$\frac{\partial a_t}{\partial \tau_t} = -\frac{\beta u_{t+1}''}{\beta u_{t+1}'' + u_t''} w_t \tag{4.20}$$

Because u is concave, $u'' < 0$ meaning that $\frac{\partial a_t}{\partial \tau_t} < 0$. However, $\frac{\beta u_{t+1}''}{\beta u_{t+1}'' + u_t''} < 1$ so it must be the case that

$$-w_t < \frac{\partial a_t}{\partial \tau_t} < 0. \tag{4.21}$$

If we substitute this into the equation for ϕ^* and take the partial derivative with respect to τ_t

$$\frac{\partial \phi_\tau^*}{\partial \tau_t} = \frac{\partial EV(a_t^w + \tau_t w_t, y_t\!=\!1)}{\partial a_t}\big(\frac{\partial a_t^w}{\partial \tau_t} + w_t\big) > 0. \tag{4.22}$$

Bibliography

[1] Allcot, Hunt and Daniel Keniston. 2014. "Dutch Disease or Agglomeration? The Local Economic Effect of Natural Resource Booms in Modern America," Working Paper No. 20508, National Bureau of Economic Research, Cambridge, MA.

[2] Autor, David, Lawrence Katz, and Melissa Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics*, 90 (2): 300-323.

[3] Autor, David, David Dorn, and Gordon H. Hansen. 2013 "China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103(6): 2121-2168.

[4] Autor, David, David Dorn, Gordon H. Hansen, and Jae Song. 2014. "Trade Adjustment: Worker-level Evidence," *Quarterly Journal of Economics*, 129(4): 1799-1860.

[5] Barrow, Lisa and Leslie McGranahan. 2000. "The Effect of the Earned Income Tax Credit on the Seasonality of Household Expenditures," *National Tax Journal*, 53 (4): 1211-1243.

[6] Bartik, Alexander. 2017. "Worker Adjustment to Changes in Labor Demand: Evidence from Longitudinal Census Data," Working Paper.

[7] Bartik, Alexander, Janet Currie, Michael Greenstone, and Christopher Knittel. 2017. "The Local Economic and Welfare Consequences of Hydraulic Fracturing," Working paper No. 23060, National Bureau of Economic Research, Cambridge.

[8] Becker, Gary S. 1962. "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy*, 5(2): 9-49.

[9] Black, Dan Terra McKinnish, and Seth Sanders. 2005. "The Economic Impact of the Coal Boom and Bust," *The Economic Journal*, 115 (April):449-476.

[10] Blanchard, Olivier and Lawrence F. Katz. 1992. "Regional Evolutions," *Brookings Papers on Economic Activity*.

[11] Borjas, George. 1987. "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77(4): 531-553.

[12] Borjas, George. 1999. "The Economic Analysis of Immigration," Chapter 28 in Handbook of Labor Economics, Volume 3, Edited by O. Ashenfelter and D. Card. Elsevier Science B.V.

[13] Borjas, George J., Stephen G. Bronars, and Stephen J. Trejo. 1992. "Self-selection and Internal Migration in the United States," *Journal of Urban Economics*, 32: 159-185.

[14] Boslett, Andrew, Todd Guilfoos, and Corey Lang. 2016. "Valuation of expectation: A hedonic study of shale gas development and New York's moratorium," *Journal of Environmental Economics and Management*, 77: 14-30.

[15] Bottan, Nicolas and Ricardo Perez-Truglia. 2017. "Choosing Your Pond: Revealed-Preference Estimates of Relative Income Concerns," Working Paper No. 23615, National Bureau of Economic Research, Cambridge, MA.

[16] Bound, John and Harry Holzer. 2000. "Demand Shifts, Population Adjustments, and Labor Market Outcomes During the 1980s," *Journal of Labor Economics*, 18(1): 20-54.

[17] Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2014. "Underinvestment in a Profitable Technology: the Case of Seasonal Migration in Bangladesh," *Econometrica*, 82(5): 1671-1748.

[18] Brooks, Arthur C. 2016. "How to Get Americans Moving Again," *The New York Times*, May 20, 2016. https://www.nytimes.com/2016/05/21/opinion/how-to-get-americans-moving-again.html.

[19] Bureau of Labor Statistics, U.S. Department of Labor. 2015. The Economics Daily, State employment-population ratios in 2014 on the Internet at https://www.bls.gov/opub /ted/2015/state-employment-population-ratios-in-2014.html (visited July 18, 2017).

[20] Cadena, Brian and Brian Kovak. 2016. "Immigrants Equilibrate Local Labor Markets: Evidence from the Great Recession," *American Economic Journal: Applied Economics*, 8(1): 257-290.

[21] Carrington, William. 1996. "The Alaskan Labor Market during the Pipeline Era," *Journal of Political Economy*, 104(1): 186-218.

[22] Cascio, Elizabeth and Ayushi Narayan. 2015. "Who Needs a Fracking Education? The Educational Response to Low-Skill Biased Technological Change," Working Paper No. 21359, National Bureau of Economic Research, Cambridge, MA.

[23] Chetty, Raj and Emmanuel Saez. 2013. "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients," *American Economic Journal: Applied Economics*, 5(1): 1-31.

[24] Chetty Raj, John N. Friedman, and Emmanuel Saez. 2013. "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American Economic Review*, 103(7): 2683-2721.

[25] Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2015. "The Effect of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *American Economic Review*, 106 (4): 855-902.

[26] Chetty, Raj, Nathaniel Hendren. 2016. "The Impact of Neighborhoods on Intergenerational Mobility II: County-Level Estimates," Working Paper No. 23002, National Bureau of Economic Research, Cambridge, MA.

[27] Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. "The Association Between Income and Life Expectancy in the United States, 2001-2014," *Journal of the American Medical Association*, 315 (16): 1750-1766.

[28] Chong, Alberto and Eliana La Ferrara. 2009, "Television and Divorce: Evidence from Brazilian "Novelas"," *Journal of the European Economic Association*, 7(2/3): 458-468.

[29] Cohen, Patricia. 2016. "Fewer Americans Strike Out for New Jobs, Crimping the Recovery," *The New York Times*, May 24, 2016. https://www.nytimes.com/2016/05/21/opinion/how-to-get-americans-moving-again.html.

[30] Conley, Timothy. 1999. "GMM estimation with cross sectional dependence," Journal of Econometrics, 92(1): 1-45.

[31] Cooke, Thomas. 2011. "It is not Just the Economy: Declining Migration and the Rise of Secular Rootedness," *Population, Space, and Place*, 17(3): 193-203.

[32] Dahl, Gordon. 2002. "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 70(6): 2367-2420.

[33] Dao, Mai, Davide Furceri, and Prakash Loungani. 2017. "Regional Labor Market Adjustments in the United States: Trend and Cycle," *the Review of Economics and Statistics*, 99(2): 243-257.

[34] Davis, James A, Tom W. Smith, Robert W. Hodge, Keiko Nakao, and Judith Treas. 1991. "Occupational Prestige Ratings from the 1989 General Social Survey." Chicago, IL: National Opinion Research Center [producer]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1991. http://doi.org/10.3886/ICPSR09593.v1

[35] DellaVigna, Stefano and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting," *Quarterly Journal of Economics*, 122(3): 1187-1234.

[36] Dionnes, Georges and Scott E. Harrington. 1991. An Introduction to Insurance Economics, in Foundations of Insurance Economics: *Readings in Economics and Finance*, Georges Dionnes and Scott Harrington eds. (Boston, Mass: Kluwer Academic, 1991).

[37] Edin, Kathryn and Laura Lein. 1997. "Making Ends Meet: How Single Mothers Survive Welfare and Low-Wage Work." *Russell Sage Foundation* New York, NY.

[38] Eissa, Nada and Hillary Hoynes. 2004. "Taxes and the labor market participation of married couples: The Earned Income Tax Credit," *Journal of Public Economics*, 88 (9): 1931-58.

[39] Eissa, Nada and Jeffrey Liebman. 1996. "Labor supply response to the earned income tax credit," *The Quarterly Journal of Economics*, 111 (2):605-37.

[40] Eliason, Paul. 2014. "Measuring the Employment Impacts of Shale Gas Development," Working Paper.

[41] Farre, Lidia and Francesco Fasani. 2013. "Media exposure and internal migration – Evidence from Indonesia," *Journal of Development Economics*, 102: 48-61.

[42] Fetzer, Thiemo. 2014. “Fracking Growth,” Working Paper.

[43] Feyrer, James, Erin Mansur, and Bruce Sacerdote. 2017. “Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution,” *American Economic Review*, 107(4): 1313-1334.

[44] Fletcher, Michael. 2010. “Few in U.S. move for new jobs, fueling fear the economy might get stuck, too,” *Washington Post*, July 30, 2010 p.A1.

[45] Flood, Sarah, Miriam King, Steven Ruggles, and J. Robert Warren. 2015. Integrated Public Use Microdata Series, Current Population Survey: Version 4.0. [dataset]. Minneapolis: University of Minnesota, http://doi.org/10.18128/D030.V4.0.

[46] Foote, Andrew, Michel Grosz, and Ann Stevens. 2015. “Locate Your Nearest Exit: Mass Layoffs and Local Labor Market Response,” Working Paper No. 21618, National Bureau of Economic Research, Cambridge, MA.

[47] Foote, Andrew. 2016. “The effect of negative house price changes on migration: Evidence across U.S. housing downturns,” *Regional Science and Urban Economics*, 60: 292-299.

[48] Ganong, Peter and Daniel Shoag. 2017. “Why has Regional Income Convergence in the U.S. Declined?” Working Paper No. 23609, National Bureau of Economic Research, Cambridge, MA.

[49] Garthwaite, Craig and Timothy J. Moore. 2012. “Can Celebrity Endorsements Affect Political Outcomes? Evidence from the 2008 US Democratic Presidential Primary,” *The Journal of Law, Economics, & Organization*, 29(2): 355-384.

[50] Garthwaite, Craig. 2014. “Demand Spillovers, Combative Advertising, and Celebrity Endorsements,” *American Economic Journal: Applied Economics*, 6(2): 76-104.

[51] Gelbach, Jonah. 2004. “Migration, the Life Cycle, and State Benefits: How Low Is the Bottom?” Journal of Political Economy, 112(5): 1091-1130.

[52] Gentzkow, Matthew. 2006. “Television and Voter Turnout,” *Quarterly Journal of Economics*, 121(3): 931-972.

[53] Gentzkow, Matthew, and Jesse Shapiro. 2010. “What Drives Media Slant? Evidence from U.S. Daily Newspapers,” *Econometrica*, 70(1): 35-71.

[54] Gold, Russell. 2014. "The Boom: How Fracking Ignited the American Energy Revolution and Changed the World," Simon & Schuster Paperbacks. New York , NY.

[55] Goodman, Lucas. 2016. "The Effect of the Affordable Care Act Medicaid Expansion on Migration," *Journal of Policy Analysis and Management*, 36(1): 211-238.

[56] Goodman-Bacon, Andrew and Leslie McGranahan. 2008. "How do EITC recipients spend their refunds?" *Economic Perspectives*, 32 (2): 17-32.

[57] Gopalakrishnan, Sathya and H. Allen Klaiber. 2014. "Is the Shale Boom a Bust for Nearby Residents? Evidence from Housing Values in Pennsylvania," *American Journal of Agricultural Economics*, 96(1): 43-66.

[58] Greenwood, Michael J. 1975. "Research on Internal Migration in the United States: A Survey," *Journal of Economic Literature*, 13(2): 397-433.

[59] Greenwood, Michael J. 1985. "Human Migration: Theory, Models, and Empirical Studies," *Journal of Regional Science*, 25(4): 521-544.

[60] Hakobyan, Shushanik and John McLaren. 2016. "Looking for Local Labor Market Effects of NAFTA," *The Review of Economics and Statistics*, 98(4): 728-741.

[61] Hanson, Gordon and Craig McIntosh. 2010. "The Great Mexican emigration," *the Review of Economics and Statistics*, 92(4): 798-810.

[62] Hicks, John. 1932. "The Theory of Wages," Macmillian, New York.

[63] Hotz, V. Joseph, Charles Mullin, and John Karl Scholz. 2005. "Examining the effect of the Earned Income Tax Credit and labor market participation of families on welfare," Working Paper.

[64] Hotz, V. Joseph and John Karl Scholz. 2003. "The Earned Income Tax Credit" In *Means-tested transfer programs in the United States*, ed. Robert A. Moffitt, 141-97. Chicago: University of Chicago Press.

[65] Hoynes, Hilary, Doug Miller, and David Simon. 2015. "Income, the Earned Income Tax Credit, and Infant Health," *American Economic Journal: Economic Policy*, 7 (1): 172-211.

[66] Hoynes, Hilary and Ankur Patel. 2016. "Effective Policy for Reducing Poverty and Inequality? The Earned Income Tax Credit and the Distribution of Income," *Journal of Human Resources*, forthcoming.

[67] Hughes, J. David. 2013. "Drill, Baby, Drill: Can Unconventional Fuels Usher in a New Era of Energy Abundance?" Post Carbon Institute, Santa Rosa, California.

[68] Jacobsen, Grant D. and Dominic P. Parker. 2014. "The Economic Aftermath of Resource Booms: Evidence from Boomtowns in the American West," *The Economic Journal*, 126(593): 1092-1128.

[69] James, Alexander and Brock Smith. 2014. "There Will Be Blood: Crime Rates in Shale-Rich U.S. Counties," Working Paper.

[70] Jensen, Robert and Emily Oster. 2009. "The Power of TV: Cable Television and Women's Status in India," *Quarterly Journal of Economics*, 124(3): 1057-1094.

[71] Internal Revenue Service (IRS). 2017. "Statistics for Tax Returns with EITC." Eitc and other refundable credits. Calendar Year Report, January 2017. https://www.eitc.irs.gov/EITC-Central/eitcstats. Accessed March 23, 2017.

[72] Kaplan, Greg and Sam Schulhofer-Wohl. 2011. "Interstate Migration has Fallen Less than You Think: Consequences of Hot Deck Imputation in the Current Population Survey," Working Paper 681, Federal Reserve Bank of Minneapolis.

[73] Kaplan, Greg and Sam Schulhofer-Wohl. 2017. "Understanding the Long-Run Decline in Interstate Migration," *International Economic Review*, 58(1): 57-94.

[74] Kearney, Melissa and Phillip B. Levine. 2015a. "Early Childhood Education by MOOC: Lessons from Sesame Street," Working Paper No. 21229, National Bureau of Economic Research, Cambridge, MA.

[75] Kearney, Melissa and Phillip B. Levine. 2015b. "Impact of MTV's *16 and Pregant* on Teen Childbearing," *American Economic Review*, 105(12): 3597-3632.

[76] Kearney, Melissa S. and Phillip B. Levine. 2015. "Investigating recent trends in the U.S. teen birth rate," *Journal of Health Economics*, 41: 15-29.

[77] Kearney, Melissa and Riley Wilson. 2016. "Male Earnings, Marriageable Men, and NonMarital Fertility: Evidence from the Fracking Boom," Working Paper No. 23408, National Bureau of Economic Research, Cambridge, MA.

[78] Kennan, John and James Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79(1): 211-251.

[79] Kling, Jeffrey, Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75(1): 83-119.

[80] Kotkin, Joel. 2009. "There's no place like home," *Newsweek*, Oct. 9, 2009.

[81] La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. "Soap Operas and Fertility: Evidence from Brazil," *American Economic Journal: Applied Economics*, 4(4): 1-31.

[82] LaLumia, Sara. 2013. "The EITC, Tax Refunds, and Unemployment Spells," *American Economic Journal: Economic Policy*, 5 (2): 188-221.

[83] Langley, Paul C. 1974. "The Spatial Allocation of Migrants in England and Wales: 1961-66," *Scottish Journal of Political Economy*, 21: 259-277.

[84] Lasky, Mark. 2016. "The Outlook for U.S. Production of Shale Oil," Congressional Budget Office, Working Paper Series, Washington, DC. Working Paper 2016-01.

[85] Lefgren, Lars and Brigitte Madrian. 2000. "An Approach to Longitudinal Matching Current Population Survey (CPS) Respondents," *Journal of Economic and Social Measurement*, 26 (1): 31-62.

[86] Leigh, Andrew. 2010. "Who benefits from the Earned Income Tax Credit? Incidence among recipients, coworkers, and firms," *The B.E. Journal of Economic Analysis and Policy*, 10(1).

[87] Liebman, Jeffrey. 1998. "The impact of the Earned Income Tax Credit on incentives and income distribution," In *Tax Policy and the Economy*, ed. James Poterba, 83-119. Cambridge: MIT Press.

[88] Liebman, Jeffrey, and Richard Zeckhauser. 2004. "Schmeduling," https://sites.hks.harvard.edu/ jeffreyliebman/schmeduling.pdf.

[89] Looney, Adam and Day Manoli. 2013. "Are there returns to experience at low-skill jobs? Evidence from single mothers in the United States over the 1990s," Unpublished working paper.

[90] Ludwig, Jens and Steven Raphael. 2010. "The Mobility Bank: Increasing Residential Mobility to Boost Economic Mobility," *The Hamilton Project*, Discussion Paper, October 2010.

[91] Malamud, Ofer and Abigail Wozniak. 2012. "The Impact of College Migration: Evidence from the Vietnam Generation," *Journal of Human Resources*, 47(4): 913-950.

[92] Maniloff, Peter and Ralph Mastromonaco. 2014. "The Local Economic Impacts of Hydraulic Fracturing and Determinants of Dutch Disease," Division of Economic and Business Working Paper Series, Colorado School of Mines.

[93] Martini, Alberto. 2002. "Seam Effect, Recall Bias, and the Estimation of Labor Force Transition Rates from SIPP," *Mathematica Policy Research*.

[94] McKenzie, David, and Hillel Rapoport. 2007. "Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence From Mexico," *Journal of Development Economics*, 84(1): 1-24.

[95] McKenzie, David and Hillel Rapoport. 2010. "Self-Selection patterns in Mexico-U.S. migration: the role of migration networks," *the Review of Economics and Statistics*, 92(4): 811-821.

[96] McKinnish, Terra. 2005. "Importing the Poor: Welfare Magnetism and Cross-Border Welfare Migration," *Journal of Human Resources*, 15(1): 57-76.

[97] Mead, Lawrence. 2014. "Overselling the Earned Income Tax Credit," *National Affairs*, 21:20-33.

[98] Meyer, Bruce and Dan T. Rosenbaum. 2001. "Welfare, the Earned Income Tax Credit, and the labor supply of single mothers," *The Quarterly Journal of Economics*, 116 (3): 1063-14.

[99] Moffitt, Robert. 1992. "Incentive Effects of the U.S. Welfare System: A Review," *Journal of Economic Literature*, 30(March): 1-61.

[100] Molloy, Raven, Christopher Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States," *Journal of Economic Perspectives*, 25(2): 1-42.

[101] Molloy, Raven, Christopher Smith, Riccardo Trezzi, and Abigail Wozniak. 2016. "Understanding declining fluidity in the U.S. Labor Market," *:*Brookings Paper in Economic Activity, Spring 2016.

[102] Monras, Joan. 2015. "Economic Shocks and Internal Migration," *IZA DP No.* 8840.

[103] Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2015. Commuting, Migration, and Local Employment Elasticities," Working Paper No. 21706, National Bureau of Economic Research, Cambridge, MA.

[104] Muehlenbachs, Lucija. Elisheba Spiller, and Christopher Timmins. 2015. "The Housing Market Impacts of Shale Gas Development," *American Economic Review*, 105(12): 3633-3659.

[105] Munshi, Kaivan. 2003. "Networks in the modern economy: Mexican migrants in the US labor market," *Quarterly Journal of Economics*, 118, 549-599.

[106] Moulton, Jeremy G., Alexandra Graddy-Reed, and Lauren Lanahan. 2016. "Beyond the EITC: The Effect of Reducing the Earned Income Tax Credit on Labor Force Participation," *National Tax Journal*, 69(2): 261-284.

[107] Newell, Richard and Daniel Raimi. 2015. "Shale Public Finance: Local Government Revenues and Costs Associated with Oil and Gas Development," Working Paper No. 21542, National Bureau of Economic Research, Cambridge, MA.

[108] Nichols, Austin, and Jesse Rothstein. 2015. "The Earned Income Tax Credit (EITC)," Working Paper No. 21211, National Bureau of Economic Research, Cambridge, MA.

[109] Notowidigdo, Matthew. 2013. "The Incidence of Local Labor Demand Shocks," Working Paper No. 17167, National Bureau of Economic Research, Cambridge, MA.

[110] O'Connell Paul G. J. 1997. "Migration Under Uncertainty: 'Try Your Luck' or 'Wait and See'," *Journal of Regional Science*, 37(2): 331-347.

[111] Partridge, Mark, Dan Rickman, M. Rose Olfert, and Kamar Ali. 2012. "Dwindling U.S. internal migration: Evidence of spatial equilibrium or structural shifts in local labor markets?" *Regional Science and Urban Economics*, 42: 375-388.

[112] Pew Research Center, Washington D.C. 2013a. "As Gas Prices Pinch, Support for Oil and Gas Production Grows, Those Aware of Fracking Favor Its Use," (March, 19, 2012). http://www.people-press.org/2012/03/19/as-gas-prices-pinch-support-for-oil-and-gas-production-grows/.

[113] Pew Research Center, Washington D.C. 2013b. "Amid Criticism, Support for Media's 'Watchdog' Role Stands Out," (August, 8, 2013). http://www.people-press.org/2013/08/08/ amid-criticism-support-for-medias-watchdog-role-stands-out/2-main-source-for-news/.

[114] Piketty, Thomas and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913-1998," *The Quarter Journal of Economics*, 118 (1): 1-41.

[115] Rivera Drew, Julia, Sarah Flood, and John Robert Warren. 2014. "Making full use of the longitudinal design of the Current Population Survey: Methods for linking records across 16 months," *Journal of Economic and Social Measurement*, 39 (1): 121-44.

[116] Roback, Jennifer. 1982. "Wages, Rents and the Quality of Life," *The Journal of Political Economy*, vol. 90(6): 1257-78.

[117] Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product differentiation in pure competition," *The Journal of Political Economy*, vol. 82(1): 34-55.

[118] Rothschild, Michael, and Joseph Stiglitz. 1970. "Increasing risk: I. A definition," *Journal of Economic Theory*, 2(3): 225-243.

[119] Rothstein, Jesse. 2008. "The unintended consequences of encouraging work: Tax incidence and the EITC," Princeton University working paper.

[120] Ruggles, Steve, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. *Integrated Public Use Microdata Series: Version 6.0.* Minneapolis: University of Minnesota.

[121] Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 2 (August): 180-212.

[122] Schochet, Peter and Ann Rangarajan. 2004. "Characteristics of Low-Wage Workers and Their Labor Market Experiences: Evidence from the Mid- to Late 1990s," Mathematica Policy Research, Inc. Final Report.

[123] Short, Kathleen. 2011. "The Research Supplemental Poverty Measure: 2010," U.S. Census Bureau Department of Commerce. Washington D.C. November.

[124] Shultz, Theodore W. 1961. "Investment in Human Capital," *American Economic Review*, 51 (1):1-17.

[125] Sjaastad, Larry. 1962. "The costs and returns to human migration," *Journal of Political Economy*, 70(supplemental): 80-93.

[126] Smeeding, Timothy, Katherin Ross Phillips, and Michael O'Connor. 2000. "The EITC: Expectations, Knowledge, Use, and Economic and Social Mobility," *National Tax Journal*, 53 (4): 1187-1210.

[127] Tax Policy Center. 2015. "EITC Parameters 1975 to 2015." http://www.taxpolicycenter.org/statistics/eitc-parameters Accessed August 27, 2015.

[128] Tobin, James. 1965. "Money and Economic Growth," *Econometrica*, 33(4): 671-684.

[129] Todaro, Michael P. 1969. "A Model of Labor Migration and Urban Unemployment in Less Developed Countries," *American Economic Review*, 59(1): 138-148.

[130] Topel, Robert. 1991. "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," *The Journal of Political Economy*, 99 (1): 145-76.

[131] U.S. Census Bureau (Census). 2008. "A Compass for Understanding and Using American Community Survey Data: What General Data Users Need to Know," U.S. Government Printing Office, Washington, DC.

[132] U.S. Census Bureau (Census) 2014. Quarterly Workforce Indicators Data. Longitudinal-Employer Household Dynamics Program http://lehd.ces.census.gov/data/qwi.

[133] U.S. Census Bureau (Census). 2015. LEHD Origin-Destination Employment Statistics. Longitudinal-Employer Household Dynamics Program http://lehd.ces.census.gov/data/lodes.

[134] U.S. Energy Administration (EIA). 2015. "Annual Energy Outlook 2015 with projections to 2040."

[135] Vachon, Mallory. 2015. "The Impact of Local Labor Market Conditions on Migration: Evidence from the Bakken Oil Boom," Working Paper. Available at SSRN: https://ssrn.com/abstract=2727380.

[136] Wilson, Riley. 2017. "Moving to Economic Opportunity: the Migration Response to the Fracking Boom," Available at SSRN: https://ssrn.com/abstract=2814147.

[137] Wilson, Riley. 2017. Moving to Jobs: The Role of Information in Migration Decisions, Available at SSRN: https://ssrn.com/abstract=30711104.

[138] Winters, Paul, Alain de Janvry, and Elsabeth Sadoulet. 2001. "Family and Community Networks in Mexico-U.S. Migration," *the Journal of Human Resources*, 36(1): 159-184.

[139] Wiswall, Matthew and Basit Zafar. 2015. "How Do College Students Respond to Public Information about Earnings," *Journal of Human Capital*, 9(2): 117-169.

[140] Wozniak, Abigail. 2010. "Are College Graduates More Responsive to Distant Labor Market Opportunities?" *Journal of Human Resources*, 45(4): 944-970.

[141] Yucong, Jiao. 2016. "The Impacts of the Earned Income Tax Credit and Welfare Reform on Work Entry and Exit," working paper, University of Illinois at Chicago.