MOVING TO ECONOMIC OPPORTUNITY: THE MIGRATION RESPONSE TO THE FRACKING BOOM

Riley Wilson*

Abstract

Exploiting positive labor market shocks from localized "fracking" booms, I estimate that fracking increased migration to impacted areas, but there is significant heterogeneity across both demographics and regions. Migrants to fracking areas were more likely to be male, unmarried, young, and less educated than movers more generally. These local booms increased in-migration rates to North Dakota fracking counties by nearly twice as much as other fracking areas. Differences across geography in labor market impacts, commuting behavior, initial population characteristics, or non-linearities only partially explained this gap. There is evidence that heterogeneous information flows might be playing a role.

Keywords: hydraulic fracturing, fracking, migration, mobility, North Dakota
JEL Classification Codes: J61, Q33, Q35, R11, R23

---

*Riley Wilson is an Assistant Professor of Economics at Brigham Young University, 435 Crabtree Building, Provo UT, 84602. (riley_wilson@byu.edu). The author acknowledges funding from the Economic Club of Washington DC through the Vernon E. Jordan Jr. Fellowship. The author is grateful to Melissa Kearney and Lesley Turner for their support and guidance, and to Judy Hellerstein, Phil Levine, Joe Price, Lucija Muehlenbachs, Cody Tuttle, Fernando Saltiel, Lucas Goodman, and Stephanie Rennane, as well as participants at the University of Maryland microeconomics workshop, Brigham Young University, the Southern Economic Association, and APPAM for useful comments; Lisa Boland and Michael Bender for research assistance; and DrillingInfo for access to the production data. The public data used in this article are available through the Census LEHD (https://lehd.ces.census.gov/data/) and IRS Statistics of Income (https://www.irs.gov/statistics/soi-tax-stats-migration-data). DrillingInfo data can be obtained through an academic license (https://www.enverus.com/blog/future-of-energy-drillinginfo-academia/). The author is willing to assist.

I. 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 to the lowest it has 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,[2] but it is unclear if we should expect a symmetric response to positive economic shocks. In this paper I document the migration response to positive local labor market shocks in the current context of low geographic mobility.

To do this, I exploit the large, positive local labor market shocks that have been generated by localized fracking booms over the past ten years. The geographic dispersion of these shocks was 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.

Consistent with previous work, I show that these localized fracking booms led to large gains in both potential earnings and employment (Feyrer et al., 2017). Among high intensity

[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).

[2] See Autor, Dorn, & Hanson, 2013; Autor et al., 2014; Hakobyan & McLaren, 2016; Cadena & Kovak, 2016; Foote et al., 2015; Monras, 2015; and Foote, 2016.

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 a significant in-migration response to these positive labor market shocks.

From individual-level data in the American Community Survey (ACS), migrants to frack areas 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. Fracking was also associated with an increase in out-migration. Out-migrants are demographically similar, suggesting that the increased in- and outflows are driven by the same types of people, and are likely due to increased churn and short term migration, with less evidence that different demographic groups are systematically sorting to and away from fracking areas. 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 migration response to localized fracking booms also varies across geography. 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. These geographic disparities still persist when scaling the migration response by the impact of fracking on earnings, to account for heterogeneous labor market impacts. 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.[3] This pattern is robust to a range of specifications and controlling for geographic spillovers or potential confounding changes in the housing market.

I explore several other potential explanations for this geographic disparity in the migration response. Accounting for potential differences in long distance commuting only widens the gap between North Dakota and elsewhere. The gap is only partially explained by differences in initial population characteristics across regions. Also, the geographic heterogeneity does not appear to be driven by potential non-linearities in the relationship between fracking production and in-migration. There is, however, geographic heterogeneity in the amount of information about each localized boom, with some fracking states– including North Dakota— receiving a large amount of media attention. 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.[4] I also show that the migration response to fracking is short-term in

---

[3] 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.

[4] A contemporaneous working paper by Vachon (wp 2015) uses net migration flows and adjusted 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 2017) 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.

nature and that many workers take advantage of the potential earnings gain through commuting. These findings also reveal compositional effects that are 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.

II. Background: The Decision to Migrate

A. A Simplified Migration Choice Model

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

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

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.[5] 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

$$\varepsilon_{io} - \varepsilon_{id} \leq (\mu_d - \mu_o) + \lambda_i'(\theta_d - \theta_o) - c_{iod}. \tag{2}$$

[5] 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).

The decision to move depends on earning differentials ($\mu_d - \mu_o$), the evaluation 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 $\varepsilon_o - \varepsilon_d$, the probability of individual i moving can be calculated as

$$\Pr(m_{iod} = 1 | \mu_o, \mu_d, \theta_o, \theta_d, \lambda_i, c_{iod}) = \Pr(\varepsilon_o - \varepsilon_d \leq (\mu_d - \mu_o) + \lambda_i'(\theta_d - \theta_o) - c_{iod}). \qquad (3)$$

This model is often used to conceptualize migrant self-selection, 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 propensity to move 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).[6]

B. 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.

[6] 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.

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, Kermit, & Sanders, 2002; 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. (2002) and Black et al. (2005) looking at the Appalachian Coal Boom in the 1970s and 1980s using coal reserves as an instrument for earnings. These authors find that for a one percent increase in earnings, the total population increased by approximately 0.16 percent. Both of these shocks occurred when migration levels were still relatively high, and 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 noted before, only two working papers have considered migration to fracking in a much more limited context and do not address important demographic and geographic differences (Bartik, wp 2017; Vachon, wp 2015).

A recent literature has explored 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, 2016). In response to negative shocks from the Great Recession, out-migration slightly 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, 2016).

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).[7] 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 a short run analysis, and I return to a discussion of this issue when I present the empirical approach.

Migration responses to fracking should be placed in the context of current migration in the US. Annual interstate migration rates are about half the level observed in the 1980s, with no current consensus on what has driven this change (Molloy et al., 2011).[8] 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

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

have to move to take advantage of wage gains (Kaplan & Schulhofer-Wohl, 2017; Molloy et al., 2011).[9]

III. 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 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 this fine 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.[10] 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. These combined technologies made extraction from shale both feasible and profitable. These technological innovations, combined with high prices, fueled localized fracking booms. Fracking was first used for natural gas extraction in the Barnett play in Texas. After success there, fracking spread to other shale plays and then was later adopted for use in oil extraction.[11] Prior to 2005, shale gas and tight

[9] 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.

[10] 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).

[11] Although oil and gas extraction do differ, Feyrer et al. (2017) find no evidence that oil and gas fracking effected labor markets differently.

oil production was almost non-existent (see Figure 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,[12] 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 believed 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., 2017; Maniloff & Mastromonaco, 2017). 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 & Mastromonaco, 2017). 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. 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). Recent predictions from both the Energy Information Administration (EIA) (2015) and independent researchers (Lasky, 2016) suggest long run

[12] These states include Arkansas, California, Colorado, Louisiana, Montana, New Mexico, North Dakota, Oklahoma, Pennsylvania, and Texas.

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.[13] Work looking at oil booms in the 1970s and 1980s found 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 found no evidence of this (Allcott & Keniston, 2014; Maniloff & Mastromonaco, 2017), perhaps because the fracking revolution was not only fueled by temporarily high prices, but by persistent technological innovation.

Importantly, recent working papers have also found that fracking might have other impacts, including high school students' graduation decisions (Cascio & Narayan, 2015), local public finance (Bartik et al., 2019; Newell & Raimi, 2015), and possibly crime rates (Bartik et al., 2019; 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. (2019) find that housing values increased by about 6 percent. To understand the relationship between fracking and migration, it will be important to consider fracking's impact on these other markets.

IV. Data

[13] 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.

To understand how labor demand shocks affect migration behavior, I will first estimate the impact of fracking on labor market measures (to show the labor market was affected) and then estimate the reduced form impact of fracking on migration. 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 (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. [14] 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 some inference 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

[14] 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.

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.[15]

V. Empirical Approach

I exploit county-level variation in fracking production to estimate the causal effect of local economic shocks on internal migration patterns. These localized fracking booms led to sudden, large increases in both local employment and earnings, allowing me to examine how improved labor market opportunities affect migration behavior. In theory, the migration response to increased earnings could be estimated using an instrumental variables strategy, where fracking production is used to instrument for local, average earnings. The exclusion restriction necessary for this strategy to estimate causal effects, is that fracking production only affects migration behavior through its impact on local, average earnings. As noted above, other markets that possibly enter the migration decisions might adjust to fracking as well, potentially violating this assumption. For this reason, my main analysis focuses on the reduced form relationship between fracking production and migration. For this reduced form approach to inform us about the relationship between labor market opportunities and migration, it must be true that fracking production primarily affects migration through its effect on labor market opportunities. As I show later, fracking has a strong impact on earnings and employment, while other potentially relevant markets are only marginally responsive in the short run. In order to fully explore geographic heterogeneity, I will impose more structure and estimate the two stage least squares equation to determine if

[15] 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.

differential migration responses are simply due to different "first stage" effects on labor markets. These instrumental variables specifications should be interpreted with these assumptions in mind, although I do provide evidence supporting the exclusion restriction – at least in the short run.

To estimate the reduced form impact of fracking on labor markets and migration, one could exploit variation in production from new wells as a local shock to oil and gas production. However, oil and gas extraction firms might choose to drill more in counties with more favorable labor markets 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 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, current 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, to construct the simulated measures I take the sample of counties with shale play and estimate

$$\ln(new\ production_{ct} + 1) = \alpha_c + \sum_{\tau} \sum_{j=1}^{J} \theta_{\tau j} I\{county\ c\ over\ shale\ play\ j\} * I\{year = \tau\} + \nu_{ct} \quad (4)$$

where $new\ production_{ct}$ 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 production plus one as the outcome in equation (4), allows me to include non-producing counties in the simulation and isolate exogenous variation along both the extensive and intensive margin of production.[16]

The identifying variation comes from differences across place in the presence of shale plays, and differences across time in technology and world prices. The vector of coefficients $\theta_{\tau j}$ estimates the average effect of being in shale play *j* on new production in each year. This is done by interacting year indicators with an indicator that equals one if county c intersects shale play j. This set of interactions will capture the average productivity for shale play *j* each year, and allows this productivity to change from year to year as world prices and technology change (as the boom progresses). 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. Because shale play productivity is determined by many counties, these coefficients capture the inherent productivity of the shale play, rather than the local economic conditions in the county that might factor into drilling intensity. 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 (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

[16] One alternative to using the log of production plus one is to take the inverses hyperbolic sine of production. Both measures yield similar migration results, with the predicted values having a correlation coefficient of 0.996.

interactions in equation (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.[17] As seen in Figure 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 might vary by state, counties might not be comparable 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. I first explore the effects of simulated new production on labor markets to understand the "first stage" labor market effects, and motivate my main analysis which explores the reduced form. This relationship is estimated as follows

$$Y_{ct} = \alpha_c + \beta_1 Sim.\ New\ Prod_{ct} + \phi_{st} + \varepsilon_{ct} \tag{5}$$

where Y_{ct} is a measure of the labor market (e.g., log average earnings) in county *c* in year *t*. *Sim. New Prod.* is the simulate production from new wells measured in tens of millions of 2010 dollars. As such, β_1 can be interepreted as the percent effect of an additional ten million dollars of simulated production from new wells. 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

[17] In Table A.6 I re-estimate the 2SLS estimates using actual new production rather than simulated production as the instrument and find similar results.

correlated within a county over time, so I adjust the standard errors to account for clustering at the county level.[18] In all estimation, I only include 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.[19]

I estimate the reduced form impact of simulated new production on migration rates using the following specification

$$Y_{ct} = \alpha_c + \beta_1 Sim.\ New\ Prod._{ct-1} + \phi_{st} + \varepsilon_{ct}. \tag{6}$$

This equation is similar to equation (5) above, but looks at migration rates into and out of county *c* in year *t* as a function of one year lagged simulated new production. As individuals likely observe earnings or employment in t-1 when making migration decisions in period t, I will look at the impact of lagged production on current migration. This specification also includes county and state-by-year fixed effects and corrects the standard errors for clustering at the county level.

The estimation in equations (5) and (6) compare counties in the same state and year with different levels of simulated new production. However, this 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

[18] 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.

[19] 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.

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.

VI. Results

A. Summary Statistics and Trends

In Table 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. Relative to non-fracking counties in the same state, fracking counties exhibit slightly lower levels of both in-and out-migration in 2000, suggesting that if anything, these labor markets are more negatively selected (which would bias the estimates downward). Fracking counties were also slightly more white and less educated, but otherwise the data suggest that fracking and non-fracking counties were quite similar before fracking began.

Any pre-fracking level differences from Table 1 are not inherently problematic for identification, but become so if they result in differential trends. I next explore changes over time to see if fracking and non-fracking counties in the same state followed similar trends prior to fracking. This also provides initial reduced form estimates of the impact of fracking on migration. To do this I estimate the following equation

$$Inmigration\,rate_{ct} = \alpha_c + \sum_{\tau}\sum_{s=1}^{S} \delta_{\tau s} I\{fracking\ county\ in\ state\ s\}_c * I\{year = \tau\} + \phi_{st} + \nu_{ct} \quad (7)$$

where the county annual in-migration rate is regressed on the set of interactions between an indicator that equals one if a county is a fracking county in state s and an indicator that equals one if the year is τ, as well as a county fixed effect and state-by-year fixed effects. As such this becomes a comparison between fracking counties and non-fracking counties in the same state each year. When doing this, I omit the 2003 year effects, just prior to the start of the fracking boom, to serve as the reference year. This allows me to trace out the effect of being a fracking county on in-migration rates over time, relative to non-fracking counties in the same state. To show these trends, I plot the percentage point difference for in-migration rates in Figure 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 mark the early transition years as fracking began.

Before 2003, the differences between fracking and non-fracking counties in the same state are flat and insignificant, suggesting counties that would later be affected by fracking were not on different in-migration trends. Starting around 2005, in-migration to North Dakota fracking counties increased, and 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 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).

B. 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 report the reduced form impact of simulated production on earnings from equation (5) in Table 2. For reference, the average simulated production from new wells in 2013 was \$13 million (2010\$) (see Table A.1 for more details about the distribution of simulated production). 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, to construct a measure of consumption earnings that adjusts for the cost of living (Blanchard & Katz, 1992), I follow the method of Ganong and Shoag (2017), and subtract five percent of the average house price from average earnings. 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 combines the effects on earnings and employment and looks at average earnings per capita. Ten million dollars of production increased average earnings per capita by two percent. 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 often the technical tasks, such as monitoring equipment, do not required advanced degrees. As seen in Appendix Table A.2, the largest labor market improvements are observed among men without a college degree. More educated men and women experienced smaller earnings and employment gains. These patterns are consistent with spillovers into complementary occupations and industries such as engineering, services, and hospitality highlighted in the previous literature (Feyrer et al., 2017; Maniloff & Mastromonaco, 2017).[20]

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. As seen in Figure (3) migration behavior in North Dakota is quite different, so I include North Dakota as a separate fifth group and will test for differences across regions. I then estimate

$$Y_{ct} = \alpha_c + \sum_{r}^{R} \beta_r Sim.New\ Prod._{ct} * 1\{region_c = r\} + \phi_{st} + \varepsilon_{ct} \tag{8}$$

where r equals *North Dakota*, *West*, *South*, *Northeast*, or *Midwest*. Throughout my sample, 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.

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.

[20] The imputation process used to assign educational attainment in the QWI could also result in significant estimates for the college educated. Census tries to minimize these errors and it likely only has a modest impact on the estimates. For more information on education imputation in the QWI see the data appendix.

C. Reduced Form Impact of Fracking on Migration

I next estimate equation (6) to explore the reduced form impacts of simulated new production on migration. 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 baseline population, arrived in the county.[21]

Migration impacts are reported in Table 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.[22] 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 population 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

[21] The number of migrants could also be measured in logs, so that β_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 7 I show that regional differences are robust to differences in initial population.

[22] Consistent with this being a causal effect, the effect of simulated production on migration from 10 year prior is small and insignificant (see Appendix Table A.3).

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. In the most productive counties in North Dakota the implied total population growth from fracking 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.[23]

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 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 and be counted). Understanding the role of these two channels also has implications for future population and labor market dynamics. On the one hand, certain groups systematically sorting away from fracking (such as the wealthy, more educated, or politically progressive) might have real effects on local governance and public good provision. On

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

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

$$Y_j = \alpha_{s_{-1}} + X_j'\Gamma + \phi_t + \varepsilon_j. \quad (9)$$

Where Y_j is an indicator for whether the individual moved to a fracking area, and 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 migrate. Cells are weighted by the summed individual weights provided by the ACS to be population representative. These demographic results are provided in Table 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.[24] 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 less educated men 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.[25] Such prevalent short term migration would suggest that monetary costs

[24] Marriage decisions could potentially adjust to fracking, although this does not seem to be the case (Kearney & Wilson, 2018).
[25] 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

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., 2015), and warrants further exploration in the future.

VII. Explaining Geographic Heterogeneity

In Table 3 there are stark geographic differences in migration responses to local fracking booms. For an additional ten million dollars of simulated production, in-migration rates in North Dakota increased by nearly five times as much as in the West, fifteen times as much as in the South, and twice as much as in the Northeast. Even when allowing this relationship to vary for each state, rather than by region, the reduced form in-migration response is significantly larger in North Dakota than in all other states (see Appendix Table A.4, column (1)). I next explore five potential explanations for this geographic disparity in an attempt to unpack individuals' migration decisions.[26]

A. Heterogeneous Labor Market Effects

Not only is there regional heterogeneity in the migration response, but as seen in Table 2, there is also regional variation in the "first stage" effect of fracking on labor markets. If individuals consider potential earnings when making migration decisions –rather than fracking production – the regional heterogeneity in migration responses might simply be due to heterogeneous effects on labor markets. To test this, I impose more structure and estimate the two stage least squares

---

simulated production at the fracking destination is higher, suggesting that return migration increased with fracking production.

[26] The explanations I explore relate to characteristics of the fracking destination. Characteristics and conditions at the origin could also lead to heterogeneous effects. In Appendix Figure A.1 I map the average annual migration flows from 2008 to 2012 to fracking state by origin county, net of average annual migration flows from 2000 to 2003 (prior to the fracking boom).

equation relating log average earnings to in-migration rates, using simulated new production to instrument for log average earnings as described by the following first and second stage equations

$$\begin{aligned} \ln Ave.Earnings_{ct-1} &= \alpha_c + \beta_1 Sim.New\ Prod._{ct-1} + \phi_{st} + \varepsilon_{ct} \\ Inmigration\ rate_{ct} &= \alpha_c + \gamma_1 \ln Ave.\widehat{Earnings}_{ct-1} + \phi_{st} + \eta_{ct}. \end{aligned} \tag{10}$$

To the extent that equation (10) identifies the elasticity of in-migration with respect to earnings, the coefficient γ_1 is of more general interest to understand how local labor market conditions affect migration. For this analysis 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.[27] 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 A.5, 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.[28] 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

[27] Simulated new production is also relevant, it is highly predictive of average earnings, with an F-statistic over 27 (see Table 2).

[28] One reason housing prices might not have risen much is because localized fracking booms also led to large inflows of commuters (see Section VII.A.). Because these commuters live somewhere else they will not affect the housing market in the same way they affect the labor market. This measure is constructed from the Federal Housing Finance Agency housing price index and converted to real dollars as explained in the data appendix. This measure is constructed from repeat sales, so new builds will not be captured. For this reason I also examine HUD median rental rates in Appendix Table A.6.

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.

First, I will directly control for housing prices in the equation.[29] It should be noted 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 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

[29] Ideally, I would like to instrument for housing prices. However, as seen in Appendix Table A.5, many of the measures that could be used to identify exogenous variation in housing supply or prices are only weakly related. This weak relationship is not entirely unexpected as many fracking areas are rural and sprawling, with elastic housing supplies.

years of production and restrict my analysis to the short run.[30] For robustness I also consider an even shorter period, and find similar patterns.

In practice, I estimate a variation of equation (10) by interacting both simulated new production and average earnings with the set of region indicators, to estimate the regions specific relationship between earnings and in-migration rates. These estimates are reported in Table 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. When accounting for heterogeneous labor market effects, the migration response to similarly sized increases in earnings is quite varied across regions, with a particularly large response in North Dakota.

[30] 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.

As seen in Appendix Table A.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 (4), or simulated new wells as the instrument, measuring simulated production in per capita terms to allow large and small counties to be affected differently by production, or controlling for median two bedroom rental rates (from the Department of Housing and Urban Development) rather than house prices to account for changes in the housing market. The geographic disparity also persists when I allow the relationship to vary for each state, rather than by region, and when I use other proxies for labor market opportunities (see Appendix Table A.4), suggesting this is not simply capturing responses to different labor market treatments.

B. Commuting as a Response to Potential Earnings Gains

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. Alternatively, it might be more costly to migrate to isolated locations (like North Dakota) so we might observe more commuters to those regions.

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. It is worth noting that this work arrangement might be particularly relevant in the oil and gas extraction industry where many workers will work on two weeks, off two weeks, and travel home during this extended break.[31]

In Table 6 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 labor market improvements in North Dakota than elsewhere.[32]

C. 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, a tight labor market with no additional labor supply, or a lack of workers with the appropriate skill set. However, in other parts of the country, there were

[31] As such, the commuting results might not generalize to other economic shocks or industries.

[32] 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.

counties with similar initial labor market conditions 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 7.[33] 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 fifth 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 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. Finally, I re-weight counties to resemble the percent of the population employed in oil and gas extraction in North Dakota in 2000. Some counties in Texas and Oklahoma have a strong tradition in oil and gas extraction, and the existing population might have the necessary skills to supply the needed labor for this new economic shock, meaning new migrants are not needed. When putting more weight on counties similar to North Dakota, the coefficients in the West and South fall slightly, suggesting the geographic disparity is not due to differences in initial oil and gas involvement. Although initial

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

population characteristics explain some of the regional gap, there still appear to be regional differences in responsiveness that are unexplained by initial population size, density and attachment to oil and gas extraction.[34]

D. Non-linear Relationship between Fracking Production and Migration

Another alternative explanation for the heterogeneous migration estimates is that the relationship between the economic shock (fracking production) 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 economic improvement is sufficiently large. Perhaps fracking counties in North Dakota experienced large enough labor market gains that justify moving, while other regions did not. Non-linearities could also arise if fracking counties in North Dakota uniformly experienced the largest economic shocks, 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 new fracking productivity. To do this I first construct the residual simulated new production, after accounting for county and state-by-year fixed effects. I then truncate my sample to county/year observations below 20 million dollars of residual simulated new production, which excludes four county/year observations outside of North Dakota, and one observation from North Dakota. Residual simulated new production is 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 Figure 4). For reference I also plot the OLS linear relationship between residual simulated

[34] 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.

production and residual in-migration rates for each region and report the coefficients. Even when restricting the sample to counties that experienced a similar economic shock the relationship in North Dakota is three times as large, and statistically different than elsewhere. As seen in Figure 4, these relationships appear fairly linear, and the geographic disparity persists when applying more restrictive truncation. 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 economic shocks migrants were more likely to select North Dakota.[35]

E. Geographic Heterogeneity in Information

A fifth 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.[36] Among domestic newspapers articles from LexisNexis which reference both fracking and a state's name, only Pennsylvania and Texas received more out of state attention than North Dakota in 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

[35] The relationship between residual simulated new production and residual log average earnings also appears to be approximately linear. Regression estimates of the first stage effect on earnings, reduced form migration effect, and the two stage least squares elasticity estimates yield similar results when a quadratic of simulated production is included.
[36] See for example, Edwin Dobb's National Geographic article (2013), Konigsberg's New Yorker article (2011), or Davey's 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.

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 8 the data suggest that an additional 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) newspaper 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 slightly smaller, when we exclude North Dakota or include a state of origin fixed effect to control for changing characteristics at the origin. Although information is likely not the only mechanism at play, information in the news about fracking appears to explain some of the difference between North Dakota and the other states, but it also appears to explain some of the variation across the other states as well.

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 explore the causality of this relationship by exploiting 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 (Wilson, 2019).

VIII. Conclusion

Internal migration rates in the US are historically low (Molloy 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, 2016; 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. This shock has been unique in that it is one of the few large, positive labor demand shocks that has affected less-educated men. However, not only have these localized fracking booms improved employment opportunities in oil and gas, but they have generated spillover effects to other demographic groups in other industries (Feyrer et al., 2017).

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 in-migration 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, which is significant and robust to changes in the housing market, geographic spillovers, and a range of other specifications. Even when accounting for differential labor market impacts of fracking, 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. A small part of this gap can be explained by commuting behavior, differences in initial population and labor market characteristics, and non-linear effects of fracking production on migration, but a geographic disparity still remains, 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. I find suggestive evidence that 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.

For comparison, both Carrington (1996) and Black et al., (2005) suggested that during the seventies and eighties, a ten percent increase in earnings was associated with a 1.6 percent increase in population. My estimates would imply that a ten percent increase in earnings from localized fracking booms, increased the population through positive net migration by 1.1 percent on average. Previous work looking at negative shocks from the Great Recession find estimates comparable in magnitude to the response in the West and South but with the opposite sign. However, these comparisons are less informative given the documented large geographic heterogeneity. As fracking affected oil and gas extraction most directly, the migration response might not fully generalize to labor demand shocks to other industries or demographic groups. However, given the cross industry spillovers generated by fracking, the evidence presented here would suggest that workers do still respond to large positive labor demand shocks by moving. Further work is needed to understand heterogeneity in 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.

REFERENCES

Allcot, Hunt & Daniel Keniston. 2014. “Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America,” Working Paper No. 20508, National Bureau of Economic Research, Cambridge, MA.

Autor, David H., David Dorn, & Gordon H. Hanson. 2013. “The China Syndrome: Local Labor Market Effects of Import Competition in the United States,” *American Economic Review,* 103(6): 2121-2168.

Autor, David H., David Dorn, Gordon H. Hanson, & Jae Song. 2014. “Trade Adjustment: Worker-level Evidence,” *Quarterly Journal of Economics,* 129(4): 1799-1860.

Bartik, Alexander. 2017. “Worker Adjustment to Changes in Labor Demand: Evidence from Longitudinal Census Data.” Working paper.

Bartik, Alexander, Janet Currie, Michael Greenstone, and Christopher Knittel. 2019. “The Local Economic and Welfare Consequences of Hydraulic Fracturing,” *American Economic Journal: Applied Economics*, 11(4): 105-155.

Black, Dan, Kermit Daniel, & Seth Sanders. 2002. “The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust,” *The American Economic Review*, 92(1): 27-50.

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

Blanchard, Olivier. & Lawrence F. Katz. 1992. “Regional Evolutions,” *Brookings Papers on Economic Activity.*

Borjas, George. 1987. “Self-Selection and the Earnings of Immigrants,” *The American Economic Review*, 77(4): 531-553.

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.

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

Boslett, Andrew, Todd Guilfoos, & 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.

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

Brooks, Arthur. 2016. “How to Get Americans Moving Again,” New York Times, May 20, 2016.

Cadena, Brian & Brian Kovak. 2016. “Immigrants Equilibrate Local Labor Markets: Evidence from the Great Recession,” *American Economic Journal: Applied Economics,* 8(1): 257-290.

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

Cascio, Elizabeth & 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.

Cohen, Patricia. 2016. “Fewer Americans Strike Out for New Jobs, Crimping the Recovery,” New York Times, May 24, 2016.

Conley, Timothy, 1999. “GMM estimation with cross sectional dependence,” *Journal of Econometrics,* 92(1): 1-45.

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.

Dahl, Gordon. 2002. "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica* 70(6): 2367-2420.
Eliason, Paul. 2014. "Measuring the Employment Impacts of Shale Gas Development," Working Paper.
Fetzer, Thiemo. 2014. "Fracking Growth," Working Paper.
Feyrer, James, Erin Mansur, & Bruce Sacerdote. 2017. "Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution," *American Economic Review,* 107(4): 1313-1334.
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.
Foote, Andrew, Michel Grosz, & 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.
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.
Ganong, Peter & Daniel Shoag. 2017. "Why has Regional Income Convergence in the U.S. Declined?" Working Paper No. 23609, National Bureau of Economic Research, Cambridge, MA.
Gelbach, Jonah. 2004. "Migration, the Life Cycle, and State Benefits: How Low Is the Bottom?" *Journal of Political Economy,* 112(5): 1091-1130.
Gold, Russell. 2014. "The Boom: How Fracking Ignited the American Energy Revolution and Changed the World." Simon & Schuster Paperbacks. New York , NY.
Goodman, Lucas. 2016. "The Effect of the Affordable Care Act Medicaid Expansion on Migration." *Journal of Policy Analysis and Management.* 36(1): 211-238.
Gopalakrishnan, Sathya & 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.
Hakobyan, Shushanik & John McLaren. (2016) "Looking for Local Labor Market Effects of NAFTA," *The Review of Economics and Statistics,* 98(4): 728-741.
Hicks, John. 1932. "The Theory of Wages," Macmillian, New York.
Hughes, J. David. 2013. "Drill, Baby, Drill: Can Unconventional Fuels Usher in a New Era of Energy Abundance?" Post Carbon Institute, Santa Rosa, California.
Jacobsen, Grant D. & Dominic P. Parker. 2014. "The Economic Aftermath of Resource Booms: Evidence from Boomtowns in the American West," *the Economic Journal,* 126(593): 1092-1128.
James, Alexander & Brock Smith. 2014. "There Will Be Blood: Crime Rates in Shale-Rich U.S. Counties," Working paper.
Kaplan, Greg & 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.
Kaplan, Greg & Sam Schulhofer-Wohl. 2017. "Understanding the Long-Run Decline in Interstate Migration," *International Economic Review,* 58(1): 57-94.
Kearney, Melissa & Riley Wilson. 2018. "Male Earnings, Marriageable Men, and NonMarital Fertility: Evidence from the Fracking Boom," *Review of Economics and Statistics* 100(4): 678-690.
Kennan John & James Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions," *Econometrica,* 79(1): 211-251.
Kling, Jeffrey, Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects," *Econometrica,* 75(1): 83-119.
Kotkin, Joel. 2009. "There's no place like home." Newsweek Oct. 9, 2009.

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

Ludwig, Jens & Steven Raphael. 2010. “The Mobility Bank: Increasing Residential Mobility to Boost Economic Mobility,” *The Hamilton Project,* Discussion Paper, October 2010.

Malamud, Ofer & Abigail Wozniak. 2010. “The Impact of College Education on Geographic Mobility: Identifying Education Using Multiple Components of Vietnam Draft Risk,” Working Paper No. 16463, National Bureau of Economic Research, Cambridge, MA.

Maniloff, Peter & Ralph Mastromonaco. 2017. “The Local Economic Impacts of Fracking: A National Study,” *Resource and Energy Economics,* 49(2017): 62-85.

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

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

Molloy, Raven, Christopher Smith, Riccardo Trezzi, & Abigail Wozniak. 2016. “Understanding declining fluidity in the U.S. Labor Market,” *Brookings Papers on Economic Activity.*

Molloy, Raven, Christopher Smith, & Abigail Wozniak. 2011. “Internal Migration in the United States,” *Journal of Economic Perspectives,* 25(2): 1-42.

Monras, Joan. 2015. “Economic Shocks and Internal Migration,” IZA Discussion Paper No. 8840.

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

Muehlenbachs, Lucija. Elisheba Spiller, & Christopher Timmins. 2015. “The Housing Market Impacts of Shale Gas Development,” *American Economic Review,* 105(12): 3633-3659.

Newell, Richard & 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.

Notowidigdo, Matthew. 2013. “The Incidence of Local Labor Demand Shocks,” Working Paper No. 17167, National Bureau of Economic Research, Cambridge, MA.

Partridge, Mark, Dan Rickman, M. Rose Olfert, & 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.

Roback, Jennifer. 1982. “Wages, Rents and the Quality of Life.” *The Journal of Political Economy,* vol. 90(6): 1257-78.

Rosen, Sherwin. 1974. “Hedonic Prices and Implicit Markets: Product differentiation in pure competition.” *The Journal of Political Economy*, vol. 82(1): 34-55.

Ruggles, Steve, Katie Genadek, Ronald Goeken, Josiah Grover, & Matthew Sobek. 2015. *Integrated Public Use Microdata Series: Version 6.0.* Minneapolis: University of Minnesota.

Sjaastad, Larry. 1962. “The costs and returns to human migration,” *Journal of Political Economy,* 70(supplemental): 80-93.

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.

U.S. Census Bureau (Census) 2014. Quarterly Workforce Indicators Data. Longitudinal-Employer Household Dynamics Program http://lehd.ces.census.gov/data/#qwi.

U.S. Census Bureau (Census). 2015. LEHD Origin-Destination Employment Statistics. Longitudinal-Employer Household Dynamics Program http://lehd.ces.census.gov/data/#lodes.

U.S. Energy Administration (EIA). 2015. "Annual Energy Outlook 2015 with projections to 2040."
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.
Wozniak, Abigail. 2010. "Are College Graduates More Responsive to Distant Labor Market Opportunities?" *Journal of Human Resources* 45(4): 944-970.
Wilson, Riley. 2019. "Moving to Jobs: The Role of Information in Migration Decisions," Unpublished Manuscript, Brigham Young University.

Table 1. Pre-fracking 2000 County Population and Labor Market Summary Statistics

	Mean Values		Within State Differences
	Non-Fracking Counties	Fracking Counties	
	(1)	(2)	(3)
Net Migration Rate	-0.09	-0.14	0.03
In Migration Rate	5.19	4.64	-0.34***
Out Migration Rate	5.28	4.78	-0.37***
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. Reduced Form Impact of Simulated Production on Local Labor Market Measures

	County Labor Market Measure in t-1				
	Log Average Earnings (1)	Log Average Non-O&G Earnings (2)	Log Earnings Adjusted for Housing Price (3)	Log Jobs to Pop. Ratio (4)	Log Average Earnings per capita (5)
Sim. New Prod. Value in Cty$_{t-1}$ (10 Millions 2010$)	0.010***	0.006***	0.011***	0.010***	0.020***
	(0.002)	(0.001)	(0.002)	(0.003)	(0.005)
	Regional Heterogeneity				
*Sim. New Prod. Value in Cty$_{t-1}$ (10 Millions 2010$)*North Dakota*	0.025***	0.016***	0.027***	0.029***	0.054***
	(0.003)	(0.001)	(0.003)	(0.004)	(0.006)
*Sim. New Prod. Value in Cty$_{t-1}$ (10 Millions 2010$)*West*	0.009***	0.005***	0.010***	0.006***	0.015***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
*Sim. New Prod. Value in Cty$_{t-1}$ (10 Millions 2010$)*South*	0.004**	0.002	0.004**	0.003	0.007*
	(0.002)	(0.001)	(0.002)	(0.003)	(0.004)
Sim. New Prod. Value in Cty$_{t-1}$ (10 Millions 2010$) Northeast*	0.103***	0.068***	0.105***	0.101***	0.205***
	(0.015)	(0.022)	(0.018)	(0.024)	(0.035)
Sim. New Prod. Value in Cty$_{t-1}$ (10 Millions 2010$) Midwest*	0.103	0.051	0.083	0.046	0.150
	(0.081)	(0.056)	(0.081)	(0.099)	(0.150)
F-statistic	27.28	18.45	28.16	13.01	19.47
Dependent Mean	34,248	33,848	28,450	0.538	19,210
Observations	31,153	31,157	31,151	31,139	31,139

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 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 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 from Fracking*100		Move from Bakken*100
Sample	Full Adult Pop.	All Migrants	Migrants to Fracking	Full Adult Pop.	All Migrants	Migrants to Fracking
	(1)	(2)	(3)	(4)	(5)	(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 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$_{t-1}$ *North Dakota*	38.02***	40.35***	35.03***	36.47***	38.40***
	(5.82)	(6.32)	(5.25)	(5.68)	(6.11)
*Log Average Earnings$_{t-1}$ *West*	24.20***	24.53***	20.59***	25.55***	24.93***
	(3.81)	(3.72)	(3.29)	(4.41)	(3.71)
*Log Average Earnings$_{t-1}$ *South*	15.67**	16.15**	14.47**	12.53	13.77*
	(7.14)	(7.60)	(6.63)	(9.79)	(8.27)
*Log Average Earnings$_{t-1}$ *Northeast*	4.71***	4.69***	4.60***	5.56***	5.09**
	(1.61)	(1.65)	(1.68)	(1.97)	(2.01)
*Log Average Earnings$_{t-1}$ *Midwest*	3.65	3.96	4.52	-1.17	9.52
	(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 6. Impact of Simulated New Production 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)
Sim. New Prod. Value in Cty$_{t-1}$	2.56***	3.05***	4.01***
*(10 Millions 2010$)*North Dakota*	(0.20)	(0.16)	(0.15)
Sim. New Prod. Value in Cty$_{t-1}$	0.29***	0.36***	0.53***
*(10 Millions 2010$)*West*	(0.08)	(0.12)	(0.17)
Sim. New Prod. Value in Cty$_{t-1}$	0.20***	0.30***	0.36***
*(10 Millions 2010$)*South*	(0.07)	(0.11)	(0.11)
Sim. New Prod. Value in Cty$_{t-1}$	0.88*	1.30	1.82
(10 Millions 2010$) Northeast*	(0.50)	(1.15)	(1.15)
Sim. New Prod. Value in Cty$_{t-1}$	-0.68	-3.42	-2.94
(10 Millions 2010$) Midwest*	(1.03)	(3.26)	(3.20)
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	0.129	0.058
P-value North Dakota equals Midwest	<0.01	0.047	0.030
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 7. 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)	Percent Population in Oil and Gas Extraction (6)
Sim. New Prod. Value in Cty$_{t-1}$	0.95***	0.95***	0.95***	0.95***	0.95***	0.95***
*(10 Millions 2010$)*North Dakota*	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Sim. New Prod. Value in Cty$_{t-1}$	0.21***	0.24***	0.24***	0.26***	0.23***	0.18***
*(10 Millions 2010$)*West*	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)
Sim. New Prod. Value in Cty$_{t-1}$	0.06***	0.13***	0.13***	0.07***	0.19***	0.07***
*(10 Millions 2010$)*South*	(0.01)	(0.03)	(0.03)	(0.01)	(0.04)	(0.02)
Sim. New Prod. Value in Cty$_{t-1}$	0.48***	0.72**	0.74**	0.51***	1.47***	0.48***
(10 Millions 2010$) Northeast*	(0.13)	(0.30)	(0.30)	(0.13)	(0.35)	(0.13)
Sim. New Prod. Value in Cty$_{t-1}$	0.38	0.98	0.86	0.08	8.99	0.38
(10 Millions 2010$) Midwest*	(0.64)	(1.22)	(1.21)	(0.60)	(5.71)	(0.64)
P-value North Dakota equals West	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
P-value North Dakota equals South	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
P-value North Dakota equals Northeast	<0.01	0.45	0.49	<0.01	0.15	<0.01
P-value North Dakota equals Midwest	0.37	0.98	0.94	0.15	0.16	0.38
Observations	31,157	31,157	31,157	31,157	31,157	31,157

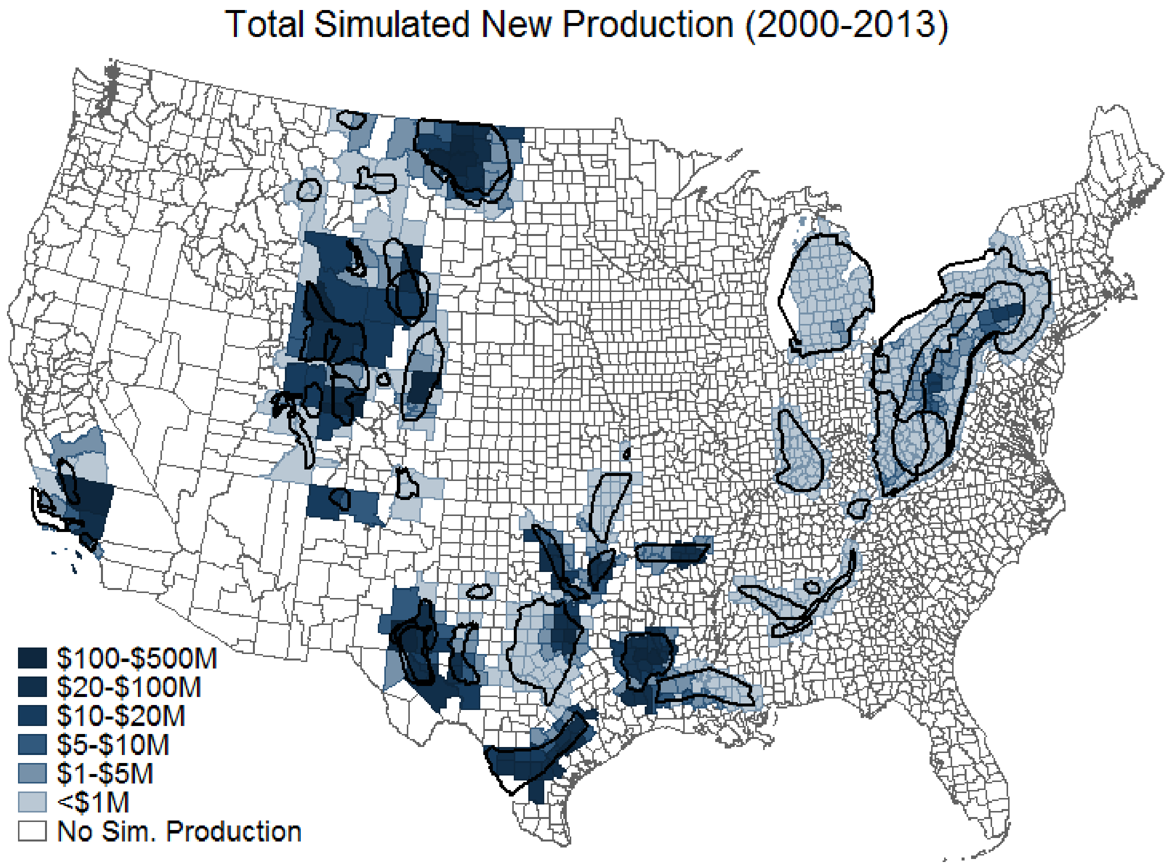
Notes: Data compiled from the IRS SOI, QWI, 2000 Census, and DrillingInfo. The impact across regions are estimated jointly to test for differences. The p-values provided are from the test of equality across the regions. Column (1) provides the baseline results from Table 3. Columns (2) through (6) re-weight 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 8. Potential Mediating Role 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

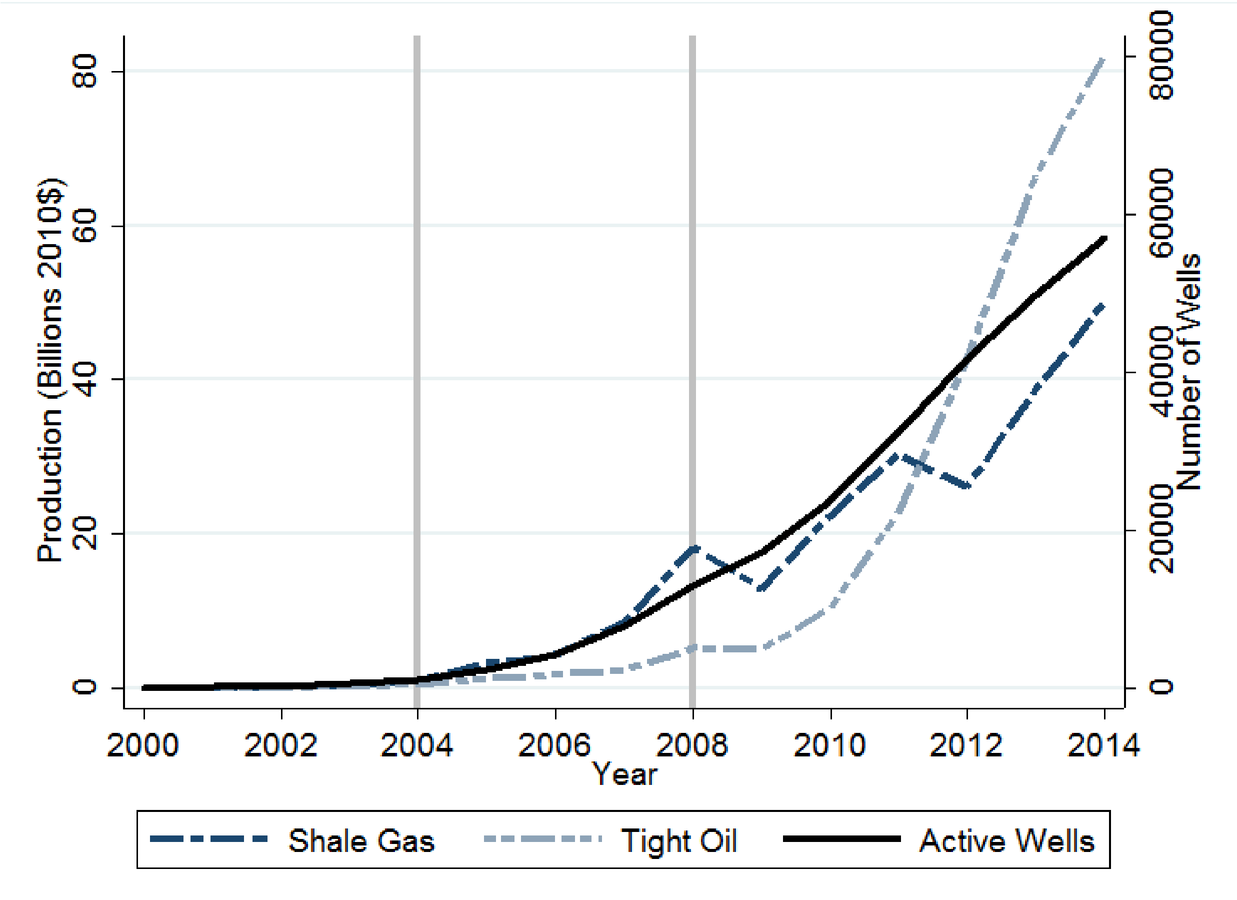
Notes: Articles were collected from LexisNexis and combined with data from the IRS SOI and DrillingInfo. Observation at the county by year by state of origin level, and capture the annual county migration inflow from each state. "Articles" is the number of news articles that reference the fracking county's 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 (6) 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 *.

Figure 1. Geographic Variation in Fracking Feasibility and Simulated Production



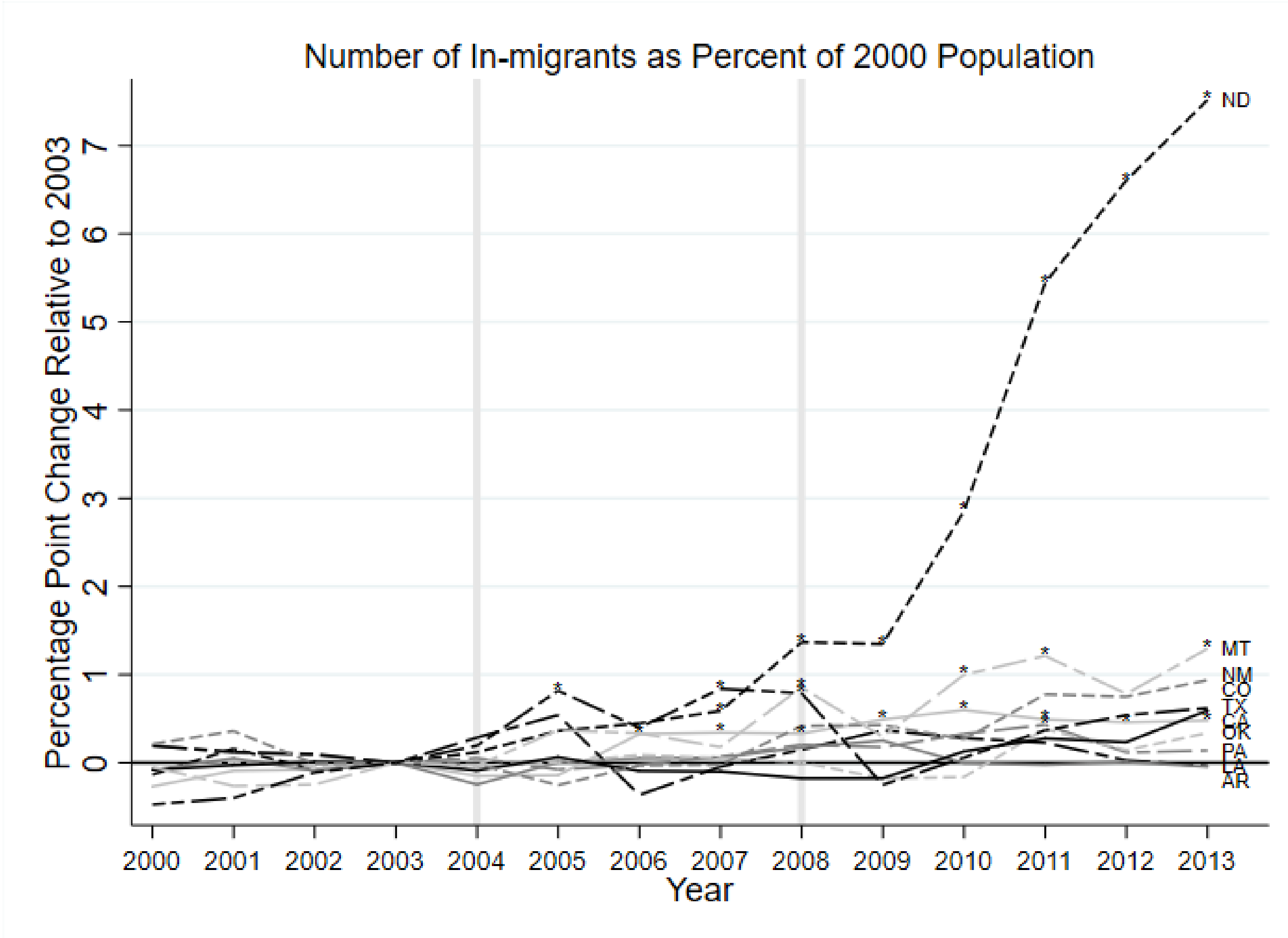
Notes: Black outlines indicate the location of shale plays. Simulated new production estimates the production value from new wells in each county as a function geology and time (see equation 4). Source: Author's calculations from DrillingInfo well level reports. Shale play boundaries obtained from the Energy Information Administration.

Figure 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.

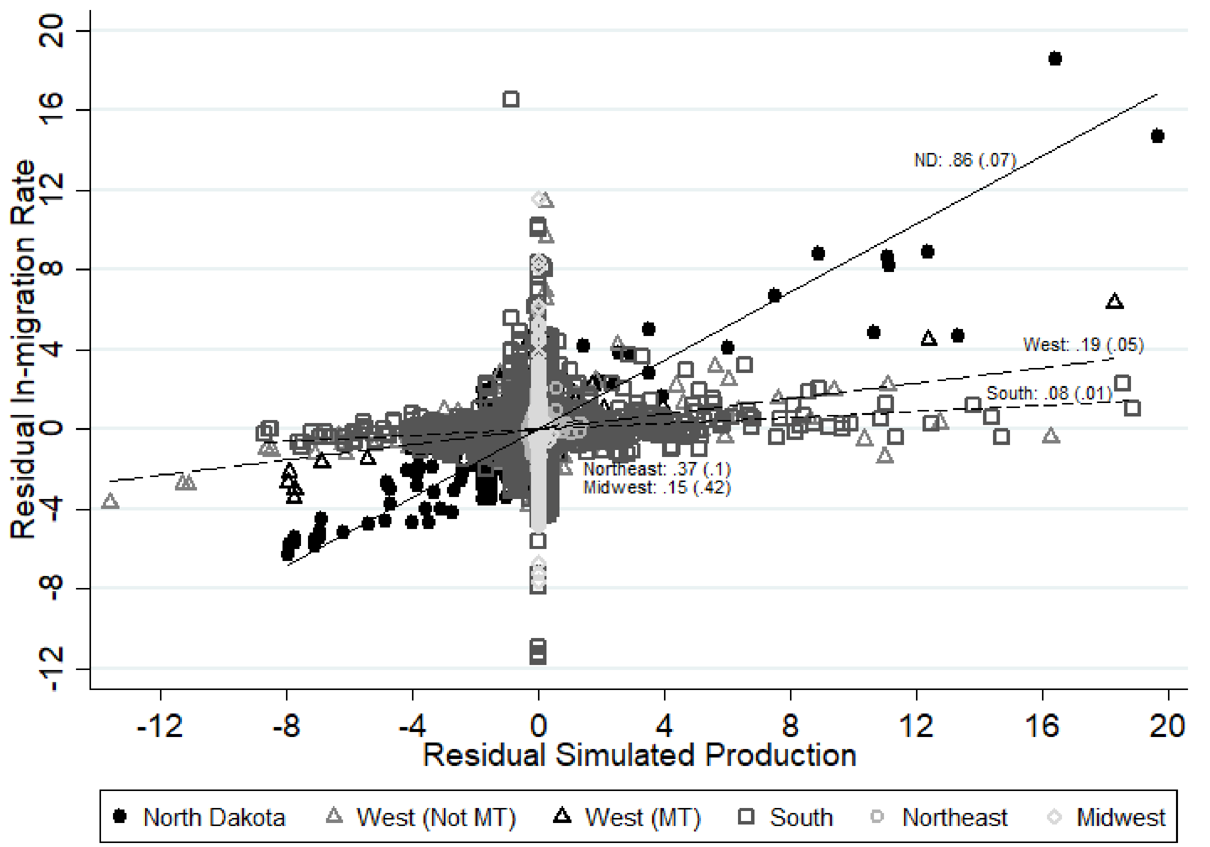
Figure 3. Trends in In-migration by State



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 4. Estimated Relationship by Region for Counties that Experienced Similar Earnings Increases



Notes: Residual simulated production that accounts for county and state by year fixed effects are 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 at 20 million dollars of residual simulated production, which excludes four county year observations outside of North Dakota, and one county-year observation from North Dakota. 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 the West, South, and Northeast. As many fracking counties in Montana also lay over the Bakken shale play, observations from Montana are indicated with a black triangle. The OLS linear fits are similar if the sample is truncated at lower values of simulated production.

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

Appendix A. Additional Tables and Figures

Appendix Table A.1. Simulated Production Summary Statistics for Fracking Counties

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
# Counties with Simulated New Production	368	413	413	489	521	614	631	681	726	720
	Simulated Annual New Production Value by Percentile (Millions 2010$)									
10th	0.02	0.01	0.02	0	0.1	0	0.06	0.03	0.09	0.12
50th	0.66	0.6	0.61	0.75	0.86	0.36	0.62	0.76	0.73	0.92
90th	15.56	18.06	16.91	17.75	24.95	14.82	23.52	27.71	25.88	26.19
	Average Simulated Annual New Production Value (Millions 2010$)									
Mean	7	7.81	7.93	7.9	11.13	7.17	9.79	11.42	11.65	12.88
Mean Among Top 150 Counties	15.05	19.02	19.25	22.94	34.37	26.39	36.74	46.25	50.27	54.81
Mean Among Top 100 Counties	21.71	27.32	27.57	32.87	49.26	37.73	52.38	65.85	71.36	77.7
Mean Among Top 50 Counties	37.97	47.47	47.93	57.04	85.89	65.42	88.8	109.64	119.13	131.86

Notes: The statistics are for all counties with any simulated production in the given year. Simulated production is reported in millions of dollars, deflated to 2010 dollars.

Appendix Table A.2. 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 (1)	College Degree (2)	No College Degree (3)	College Degree (4)	No College Degree (5)	College Degree	No College Degree	College Degree
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 *.

Appendix Table A.3. Falsification Test: Effect of Simulated Production on Migration Year T-10

	Number of In-Migrants, as Percent of 2000 Population		
Year of migration outcome	Current	Current	10 Years Prior
	(1)	(2)	(3)
Sim. New Prod. Value in Cty$_{t-1}$	0.30***	0.32***	0.01
(10 Millions 2010$)	(0.09)	(0.10)	(0.01)
	Regional Heterogeneity		
Sim. New Prod. Value in Cty$_{t-1}$	0.95***	0.96***	0.03*
*(10 Millions 2010$)*North Dakota*	(0.06)	(0.07)	(0.02)
Sim. New Prod. Value in Cty$_{t-1}$	0.21***	0.16**	0.01
*(10 Millions 2010$)*West*	(0.05)	(0.07)	(0.03)
Sim. New Prod. Value in Cty$_{t-1}$	0.06***	0.07***	0.00
*(10 Millions 2010$)*South*	(0.01)	(0.01)	(0.01)
Sim. New Prod. Value in Cty$_{t-1}$	0.48***	0.48***	-0.47***
(10 Millions 2010$) Northeast*	(0.13)	(0.11)	(0.15)
Sim. New Prod. Value in Cty$_{t-1}$	0.38	0.37	-0.19
(10 Millions 2010$) Midwest*	(0.64)	(0.43)	(0.52)
Years in Sample	1999-2013	2005-2013	2005-2013
Dependent Mean	5.17	5.20	5.07
Observations	31,157	20,850	20,840

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. All regressions include county and state by year fixed effects, which make this a comparison between counties in the same state. Column (1) presents the baseline estimates from Column (2) of Table 3. Column (2) restricts the baseline estimate to the same years available for the falsification test. Column (3) assigns the annual in-migration rate for the period ten years earlier. For example, this regression considers how simulated production in 2013 affects migration in 2003. Standard errors are corrected for clustering at the county level. $p<0.01$ ***, $p<0.05$ **, $p<0.1$ *.

Appendix Table A.4. State Specific Migration Response to Earnings

	Outcome: Number of In-migrants as a Percent of 2000 Population					
Labor Market Measure	Reduced Form: Sim. Prod. (1)	Average Earnings (2)	Average Earnings Controlling for Housing Price (3)	Housing Adjusted Earnings (4)	Jobs to Population Ratio (5)	Average Earnings per capita (6)
Log Measure$_{t-1}$	0.95***	38.02***	40.35***	35.03***	32.96***	17.54***
	(0.06)	(5.82)	(6.32)	(5.25)	(5.40)	(2.69)
			Western States			
*Log Measure$_{t-1}$*MT*	-0.60***	-8.53	-10.76	-9.11	7.95	-0.61
	(0.06)	(6.21)	(6.68)	(5.62)	(7.76)	(3.04)
*Log Measure$_{t-1}$*NM*	-0.75***	-27.27***	-29.01***	-25.59***	-17.51***	-11.17***
	(0.08)	(5.97)	(6.47)	(5.38)	(6.68)	(2.86)
*Log Measure$_{t-1}$*CO*	-0.87***	-7.73	-3.66	-11.65	-362.4	15.77
	(0.08)	(24.74)	(20.80)	(14.69)	(3,469)	(60.96)
*Log Measure$_{t-1}$*CA*	-0.88***	-22.63	-25.13	-47.95***	-19.73*	-10.46
	(0.07)	(20.19)	(20.46)	(18.36)	(10.53)	(6.58)
			Southern States			
*Log Measure$_{t-1}$*TX*	-0.88***	-17.96	-18.64	-16.30	-15.91	-8.41
	(0.06)	(12.88)	(14.67)	(11.97)	(16.33)	(7.02)
*Log Measure$_{t-1}$*OK*	-0.93***	-31.91***	-33.88***	-29.38***	-26.52*	-14.43**
	(0.06)	(11.60)	(12.87)	(10.64)	(15.87)	(6.34)
*Log Measure$_{t-1}$*AR*	-0.90***	-34.08***	-36.50***	-31.71***	-17.83	-14.42***
	(0.06)	(6.12)	(6.59)	(5.49)	(18.64)	(2.87)
*Log Measure$_{t-1}$*LA*	-0.93***	-26.33	-28.70	-23.37	-43.99**	-1,047
	(0.06)	(25.62)	(27.85)	(27.79)	(21.96)	(143,500)
			Northeastern States			
*Log Measure$_{t-1}$*PA*	-0.42***	-32.81***	-35.01***	-30.25***	-27.86***	-14.98***
	(0.14)	(6.06)	(6.57)	(5.50)	(5.74)	(2.83)
			Other States			
*Log Measure$_{t-1}$*Other*	-0.79***	-12.53*	-14.58**	-16.60***	-3.73	-3.96
	(0.06)	(6.40)	(6.80)	(6.05)	(7.42)	(3.23)
Independent Mean	0.178	34,516	34,516	28,688	0.538	19,363
Observations	31,143	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 *.

Appendix Table A.5. Reduced Form Effect of Simulated Production on Housing Prices

		Characteristic		
	Baseline	Share Vacant in 2000	Geography Constraint	Share Own Water in 2000
	(1)	(2)	(3)	(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

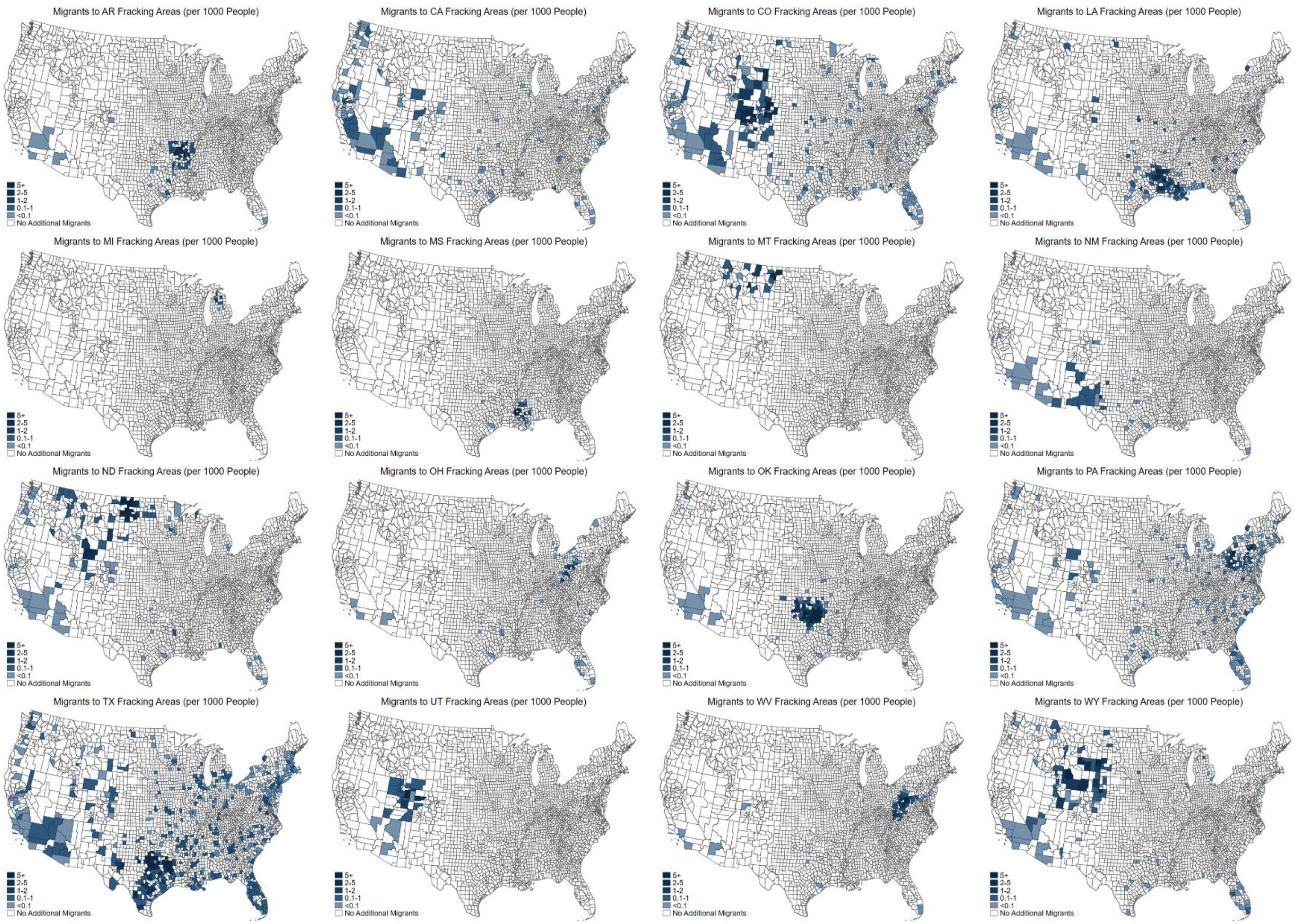
Notes: Housing price constructed from the housing price index provided by the Federal Housing Finance Agency and converted to dollars using county median house prices in 2000. Simulated production is interacted with a binary indicator for each of the five regions. The impact across regions are estimated 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 *.

Appendix Table A.6. Robustness of Regional Migration Elasticities

	Outcome: Number of In-migrants as a Percent of 2000 Population							
Specification:	Baseline (1)	Weighted by 2000 Population (2)	Shorter Sample (≤2011) (3)	Actual Prod. as Instrument (4)	Play by Year Interacts as Instruments (5)	Sim. New Wells as Instrument (6)	Sim. Prod. Per Capita as Instrument (7)	Control for Median 2 Bedroom Rent (8)
Log Average Earnings$_{t-1}$	38.02***	37.14***	28.69***	40.81***	36.51***	35.45***	31.06***	37.66***
**North Dakota*	(5.82)	(3.09)	(1.78)	(7.14)	(6.37)	(5.85)	(7.58)	(6.70)
Log Average Earnings$_{t-1}$	24.20***	20.19	20.14***	21.44***	0.74	19.05**	24.85***	16.80**
**West*	(3.81)	(30.82)	(4.08)	(4.48)	(2.15)	(7.76)	(2.65)	(6.56)
Log Average Earnings$_{t-1}$	15.67**	8.77	17.17	10.83*	2.86	14.47**	10.42**	12.04***
**South*	(7.14)	(14.00)	(15.35)	(6.40)	(1.88)	(7.24)	(5.01)	(4.45)
Log Average Earnings$_{t-1}$	4.71***	3.17*	17.01**	5.31***	5.03	4.33***	2.02	5.13***
**Northeast*	(1.61)	(1.63)	(7.59)	(1.78)	(3.62)	(1.58)	(1.73)	(1.67)
Log Average Earnings$_{t-1}$	3.65	-6.99	28.60	-5.89	3.62**	17.76	3.14	10.28
**Midwest*	(7.04)	(22.85)	(44.64)	(6.00)	(1.81)	(33.54)	(5.90)	(13.28)
P-values:								
North Dakota equals West	0.05	0.58	0.05	0.02	<0.01	0.09	0.44	0.03
North Dakota equals South	0.02	0.05	0.46	<0.01	<0.01	0.02	0.02	<0.01
North Dakota equals Northeast	<0.01	<0.01	0.13	<0.01	<0.01	<0.01	<0.01	<0.01
North Dakota equals Midwest	<0.01	0.06	0.99	<0.01	<0.01	0.60	<0.01	0.07
Observations	31,157	31,157	26,533	31,157	31,157	31,157	31,155	26,249

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 *

Appendix Figure A.1. Additional Migrants to Fracking Areas by Origin County



Notes: For each fracking state, the average annual number of migrants from 2008-2012 minus the average annual number of migrants from 2000-2003 per 1,000 people at the origin county is plotted. The number of migrants is aggregated over all fracking counties in the state, meaning the migration flows to Texas are capturing the flows to more counties than the flows to other states.

Appendix B. Data Appendix

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

I 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.[37] 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-

[37] 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.

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 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. 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.

II 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.

III U.S. Census Bureau 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 (LEHD) micro-level data collected from unemployment insurance earnings data from participating states and several other sources.[38] The QWI is aggregated to the county level, and can be tabulated by firm characteristics (industry, size) or worker characteristics (gender, age, education).[39] When tabulating by worker characteristics, only two levels of tabulation are feasible (gender by age or gender by education). The educational attainment measures in the LEHD are imputed based on a state-specific logistic regression among individuals from the 2010 Census long form and uses individual measures (such as age, earnings, and industry) to predict education level (LEHD, 2005). Because of this potential measurement error, I focus on earnings and employment across the entire population in my main analysis, rather than separately by education. 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 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

[38] 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.

[39] 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.

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.

IV 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.

V 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.

VI 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.