Local Booms and Innovation*

Federica Coelli[a] and Paul Pelzl[b]

[a]EBRD & University of Zurich & CEPR

[b]NHH Norwegian School of Economics

December 2025

Abstract

Using oil and gas shocks as an exogenous source of booms and busts at the U.S. commuting zone level, we provide novel evidence that local booms increase local patenting, especially in non-metropolitan areas. This reflects agglomeration economies that make incumbent inventors more productive. In contrast to total patenting, innovation in oil and gas – the sector closest to the boom – is countercyclical, consistent with higher opportunity costs of innovation in a booming industry. Our findings shed new light on the spatial dimension of innovation, inform recent debates on place-based industrial policy, and help to reconcile mixed evidence on the cyclicality of innovation.

JEL codes: O12, O31, Q33, R11

Keywords: Innovation, patents, local economic booms, agglomeration

---

*We thank Hunt Allcott and Daniel Keniston for generously sharing data, and Knut Blind, Filippo Boeri, Ralph De Haas, Eugenie Dugoua, Marion Dumas, David Hemous, Hans Hvide, Josh Lerner, Ralf Martin, Yasusada Murata, Steven Poelhekke, Maria-Eugenia Sanin, André Seidel, Florian Seliger, Vincent Sterk, Jan Stuhler, Bram Timmermans, Zexi Zhou, seminar participants at Tinbergen Institute Amsterdam, NHH Norwegian School of Economics, University of Bergen, The Research Institute of Industrial Economics (IFN), University of Stavanger, Western Norway University of Applied Sciences, and conference participants at the 18th North American Meeting of the Urban Economics Association, the 13th European Meeting of the Urban Economics Association, the Venice Summer Institute 2024 on Globalization and Innovation, the Ninth ZEW/MaCCI Conference on the Economics of Innovation and Patenting, the 39th meeting of the European Economic Association, the 14th ifo Workshop on Regional Economics, the MaCCI Annual Conference 2023 and the NOeG Winter Workshop 2023 for valuable comments. Federica Coelli gratefully acknowledges the financial support of the European Commission under the ERC Starting Grant 805007 AUTOMATION.
Federica Coelli: European Bank for Reconstruction and Development (EBRD), 5 Bank Street, London E14 4BG, United Kingdom. Email: coellif@ebrd.com.
Paul Pelzl (corresponding author): NHH Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway. Email: paul.pelzl@nhh.no.

1 Introduction

American innovation is highly concentrated in a few large metropolitan areas. For example, Santa Clara County in Silicon Valley averaged 3,000 patents per year between 1970 and 2010, corresponding to a remarkable 2.2 patents per thousand residents. In contrast, Okmulgee County in Oklahoma produced only one patent annually, translating to 0.03 patents per thousand residents. These spatial disparities have sparked a vivid policy debate in recent years. On one side of the debate, Gruber and Johnson (2019) or Atkinson et al. (2019) advocate for a stronger geographic dispersion of research funding to promote local innovation clusters beyond superstar cities, to help "jump-start the American growth engine" (Gruber and Johnson, 2019, p.11). On the other side, such views have been challenged both on the grounds of efficiency (Moretti, 2021) and equity (Glaeser and Hausman, 2020).

This debate raises broader questions about the responsiveness of innovation to a local economic impetus. Does a rise in local economic activity promote the creation of new ideas? Does the impact vary across urban areas and more remote places, or across other location-specific characteristics? And what are the mechanisms at play? These questions are largely unanswered in the existing literature, although they are crucial for understanding local innovation dynamics and regional development.

A priori, the effect of local economic booms on innovation is not obvious. On the one hand, booms may relax financial constraints or lead to agglomeration economies, both of which have been shown to promote innovation (see e.g. Amore et al., 2013; Carlino and Kerr, 2015). On the other hand, theoretical work (e.g. Aghion and Saint-Paul, 1998) suggests that innovation is counteryclical due to higher opportunity costs of innovation in boom times, and some have argued that positive income shocks could reduce labor supply and thereby lower innovation (Glaeser and Hausman, 2020).

This paper makes two main contributions. First, we provide novel and plausibly causal evidence that innovation increases in response to local economic booms whose genesis is unrelated to innovation. We derive this finding from a long sample of US patenting activity at the commuting zone level covering over four decades (1969-2012). These findings are not driven by superstar cities; in contrast, we find the effects to be largest in commuting zones that are urban but non-metropolitan. Controlling for urbanization, we also find that commuting zones with a history of higher patenting per capita experience stronger innovation growth during booms, whereas higher levels of human capital or a larger college presence do not lead to similar effects. Further results indicate that (i) patent *quality* does not decline during local economic upswings; (ii) busts reduce patenting to a similar extent as booms increase it; (iii) booms raise patenting more strongly in technologies a commuting zone is historically familiar with, indicating path dependence; and (iv) our results are even observable at the firm level: a local boom makes multi-location firms patent more in that booming commuting zone relative to other zones where they are active, controlling for firm-time fixed effects. Using such within-firm regressions and a wide range of other specifications, we test for multiple mechanisms and present evidence indicating agglomeration economies as the key explanation for procyclical local innovation. Agglomeration in booming commuting zones has mixed consequences for the surrounding region: patenting rises in neighboring commuting zones, signaling wider agglomeration effects of the boom, while patenting declines in somewhat more distant areas, indicating displacement effects.

The second contribution of this paper is to highlight heterogeneous effects across industries which shed light on a longstanding puzzle in the cyclicality of innovation literature. Theory typically predicts that firms undertake productivity-improving activities such as innovation during recessions, because of lower opportunity costs (Davis and Haltiwanger, 1990; Hall, 1991; Aghion and Saint-Paul, 1998; Canton and Uhlig, 1999). However, empirical studies exploiting country- or industry-level data typically find that innovation is procyclical (Geroski and Walters, 1995; Comin and Gertler, 2006; Ouyang, 2011). While studies that address this puzzle typically take the opportunity cost argument as given and explore offsetting factors, our approach allows us to test the empirical relevance of opportunity costs in a novel way. We show that sectors more exposed to rising local demand, thereby facing a larger increase in the opportunity cost of innovation in boom times, keep their patenting activity unchanged or reduce it. In contrast, less exposed sectors significantly increase patenting. These findings provide direct empirical evidence that opportunity costs do affect firms' innovation decisions, although they are not sufficiently relevant to generate a reduction in overall innovation during local booms.

We measure local economic booms through exogenous oil and gas shocks, which provide a natural experiment for studying the impact of economic fluctuations on local innovation. We define oil and gas shocks as the interaction of a commuting zone's initial oil and gas endowment with time-series variation in national oil and gas employment, following Allcott and Keniston (2018). Our long sample period spans the oil boom of the late 1970s and early 1980s, the prolonged bust through the late 1990s, and the recent fracking boom of the 2000s. Each of these episodes added or eliminated hundreds of thousands of oil and gas jobs across the United States, providing substantial time-series variation in our analysis. The share variable in our shift-share design, initial oil and gas endowment, is a function of geology, and we show that it does not robustly correlate with other relevant commuting zone characteristics. This makes our shares more exogenous relative to most other shift-share approaches. Nonetheless, we account for recent methodological advances in this literature (Goldsmith-Pinkham et al., 2020).

The described shift-share interaction is a significant driver of various measures of economic activity, including population, employment, wages, personal income, GDP, and local government revenue (see also Allcott and Keniston, 2018). This makes the interaction an effective proxy for local economic booms. Given that the effect of economic booms on patenting may operate via multiple measures of economic activity, we apply a reduced-form approach rather than a 2SLS-IV strategy, regressing innovation outcomes on the shift-share variable. The results show that in a commuting zone with an initial oil and gas endowment of one standard deviation (approximately five million dollars per square mile), a 100-log-point increase in national oil and gas employment leads to an 8.3% increase in granted patents. For a commuting zone with median patenting activity, this corresponds to one additional patent. Note that while this estimate captures effects relative to less endowed commuting zones, estimating absolute effects net of spatial spillovers yields a similar magnitude (+6.9%).

The main result of procyclical local innovation masks substantial heterogeneity across different types of sectors. First, we examine oil and gas, the sector most directly affected by the boom. Our results reveal that local patenting in oil and gas technology is *counter*cyclical. This is consistent with theory, since the elevated profits for the sector in times of high national oil and gas employment raise local oil and gas producers' opportunity cost of innovation. In contrast, we find that patenting in non-oil and gas technology (which represents 98% of total patenting in our sample) is *pro*cyclical. To explore the role of opportunity costs further, we use novel techniques to disaggregate non-oil and gas patenting into different industries within manufacturing, which overall accounts for the great majority of patenting. Our industry distinction is guided by the differing implications that a rise in demand during a local boom has across industries. Firms in highly traded industries mainly sell outside their commuting zone and thus hardly benefit from higher local demand during a boom, leaving their opportunity cost of innovation largely unaffected. In contrast, firms in lowly-traded industries can raise both prices and sales upon higher local demand since they face lower import competition, raising their opportunity cost of innovation during booms. Consistent with this variation in opportunity costs, we find that the rise in non-oil and gas patenting during booms is driven by highly traded goods producers, while patenting in lowly-traded industries does not significantly rise.[1]

We test various mechanisms to understand why local non-oil and gas patenting is procyclical overall. We start by testing for a finance channel, based on evidence that relaxed credit constraints can help explain procyclical innovation (Ouyang, 2011; Aghion et al., 2012). Our results do not support a credit constraints narrative: (i) industries that depend more on external finance (see Rajan and Zingales, 1998) do not raise patenting by more than other industries during booms; (ii) listed firms – which appear less dependent on finance created by local booms – do not raise patenting by less than other firms; and (iii) the result that multi-location firms raise patenting by more in booming commuting zones than others speaks against a finance channel, since these firms could channel additional funds to other zones and raise innovation there, rather than use them in the booming zone. Second, we show that firms upstream to oil and gas do not raise patenting by more than others during booms, which rules out supply chain linkages as an important mechanism. Additional findings jointly suggest a key channel explaining our results: agglomeration effects which make incumbent inventors more productive. First, innovation is procyclical due to increased patenting by inventors who had patented in the same commuting zone before, rather than by in-moving or new inventors. Second, local booms raise the count of college graduates and creative class workers (which we attribute to migration), likely benefiting local inventors. Third, local booms not only raise the number of patents, but also patents per capita. Fourth, the mentioned within-firm level results suggest that local

[1] Note that upward wage pressure from the booming oil and gas sector might induce highly traded goods producers to shed some production workers, but Allcott and Keniston (2018) only find weak evidence for this. More importantly, such crowding-out effects are unlikely to occur in the labor market for workers involved in innovation (and thereby negatively affect highly traded industries' patenting activity), given our evidence that oil and gas patenting decreases during booms.

factors which are not transferrable to other zones are driving the increase in patenting, such as precisely local agglomeration economies. Two other findings further support the agglomeration narrative. First, as discussed above, patenting responds most in urban yet non-metropolitan commuting zones. This aligns with the finding of Carlino et al. (2007) that the benefits of agglomeration for innovation are strongest in moderately sized areas. Second, our results on geographic spillovers (increased patenting nearby, decreased patenting somewhat further away) would be difficult to rationalize if factors other than agglomeration were the main driver of our results.

1.1 Related Literature

Our paper contributes to several strands of literature. First, we add to the literature on local innovation. This body of work has studied the impact of socio-economic conditions on local patenting (Crescenzi and Rodríguez-Pose, 2013; Hasan et al., 2020), agglomeration and local innovation spillovers (Jaffe et al., 1993; Audretsch and Feldman, 1996; Carlino et al., 2007; Matray, 2021), effects of spatial proximity (Roche, 2020; Xiao et al., 2021), place-based policies (Moretti and Wilson, 2014; Tian and Xu, 2022), or the benefits of local innovation for regional development (Akcigit et al., 2017). Little attention has been paid to the cyclicality of innovation at the local geographic level. Considering the large body of work on innovation's cyclicality at the industry level, this appears surprising; however, it is likely explained by data constraints and the difficulty of identifying exogenous local economic shocks. We address these issues and contribute to the literature by showing that innovation rises upon increased local economic activity, by highlighting heterogeneous effects across types of regions and sectors, and by identifying mechanisms.

Second, we contribute to the literature on the cyclicality of R&D and innovation. We do so by adding to a body of work seeking to reconcile the divergence between theoretical predictions and empirical evidence. Aghion et al. (2012) find that for financially constrained firms, there is a positive correlation between firm-level sales and R&D investment. Bernstein et al. (2021) show that negative wealth shocks made inventors less productive during the Great Recession. Assuming that inventors quickly commercialize R&D, Barlevy (2007) theorizes that R&D is less profitable during recessions, and benefits competitors in the next boom.

Consistent with this theory, Fabrizio and Tsolmon (2014) present empirical evidence that R&D and patenting are more procyclical in industries with faster obsolescence or a greater threat of imitation. The above studies thus typically explain procyclical innovation by introducing a factor that may outweigh the impact of opportunity costs, and test for the factor's relevance by exploiting its heterogeneity across individuals, firms, or industries. In our paper, we instead test the opportunity cost theory directly by exploiting heterogeneity in the opportunity cost of innovation across different industries within a given boom and locality. While Barlevy (2007) asks, "is the simple opportunity cost model inappropriate when it comes to R&D?" (p.1131), our patent-based analysis provides plausibly causal evidence against this hypothesis. This offers novel empirical support for a longstanding theory dating back to Schumpeter (1939), thereby considerably strengthening the argument that the "cyclicality puzzle" is explained by pro-innovation forces *outweighing* higher opportunity costs during booms.

Our paper also adds to the ongoing discussion about the relationship between natural resources and economic development, which has been studied both theoretically (Corden and Neary, 1982; van Wijnbergen, 1984) and empirically (e.g. Sachs and Warner, 2001; Aragón and Rud, 2013; Caselli and Michaels, 2013; Allcott and Keniston, 2018; De Haas and Poelhekke, 2019; Pelzl and Poelhekke, 2021). This literature has largely focused on relatively short-run movements in employment, revenue, and population, rather than the drivers of long-run growth (except for total factor productivity). Our contribution is to show that natural resource booms lead to an increase in local patenting, providing evidence against a resource curse in innovation.

Finally, we contribute to the literature on the economic responses to local shocks (Blanchard and Katz, 1992; Mian and Sufi, 2014; Giroud and Mueller, 2017). By using oil and gas shocks as a source of local economic booms, our approach is related to Feyrer et al. (2017), who exploit local fracking booms to study the geographic dispersion of local economic shocks.

The remainder of the paper is structured as follows. Section 2 outlines the empirical strategy; Section 3 describes the data sources and the construction of key variables; Section 4 reports the main results, examines heterogeneity across regions and sectors, analyzes underlying mechanisms and geographic spillovers, and presents robustness checks; and Section 5 concludes.

2 Empirical Strategy

We set out to estimate plausibly causal effects of variation in economic activity on innovation at the US commuting zone level. Our empirical strategy is to exploit nationwide oil and gas shocks as an exogenous source of local booms and busts. The idea is that these nationwide shocks are more relevant for commuting zones with larger oil and gas endowment, such that the co-occurrence of a large aggregate shock and large local endowment captures a local economic boom that is unrelated to local developments. We implement this approach via a shift-share design in the spirit of Bartik (1991). Following Allcott and Keniston (2018), we define local oil and gas booms as the interaction between time series variation in oil and gas employment at the national level (the "shift") and cross-sectional variation in initial (1960) oil and gas reserves per square mile at the local level (the "share"). National oil and gas employment fluctuates substantially across our sample period, reflecting the booms of the late 1970s and the 2000s and a long bust period in between (see Figure 1).[2] At the start of Section 4, we show that the chosen oil and gas boom measure has a statistically significant and positive impact on various measures of local economic activity, such as population, employment, or wages. This makes the shift-share interaction an effective proxy for local booms and busts. As Equation (1) below illustrates, we do not use the interaction as an instrument in a 2SLS-IV setup, but instead use it directly as our main explanatory variable, thereby applying a reduced-form approach. We do so because the effect of an economic boom on patenting likely operates via multiple measures of economic activity, rather than only via population, for instance. Therefore, the exclusion restriction would likely be violated in any 2SLS-IV setup.

The local endowment measure is largely a function of geology. This holds especially because the measure includes "undiscovered" reserves, which are unrelated to exploration efforts (see Section 3). In line, we find that initial oil and gas endowment does not robustly correlate with other relevant commuting zone characteristics (see Table A2 in the Appendix). This eases identification concerns, given the evidence of Goldsmith-Pinkham et al. (2020) that exogeneity

[2] An alternative shift variable would be oil and gas prices, but these are not as good a proxy of US oil and gas booms. This is because declining gas prices at the end of our sample period (see Figure A4 in the Appendix) do not indicate a gas bust, but rather reflect increased gas supply during the fracking boom. That said, Figure A3 shows that national oil and gas employment and the *oil* price follow a very similar trajectory during our sample period, and our results are robust to using the oil price as a shift variable (see Table 9, column 6).

of the shares (conditional on controls, which we add in a robustness check) is sufficient to make a shift-share approach a valid identification strategy.[3] We further discuss identification assumptions and how we deal with them after presenting our empirical specification next.

Our main dependent variables are local patent counts. Therefore, we use a Poisson specification. We estimate the following model, where the unit of observation is a commuting zone:

$$
\begin{aligned}
Y_{c,\tau} = exp\Big(&\beta_1[\mathit{Oil\&Gas\ Endowment}_{c,T} \times ln(\mathit{National\ Oil\&Gas\ Employment}_{\tau})] \\
&+ \delta_{c,T} + \delta_{c,T} * \tau + \gamma_{s,\tau}\Big) + \epsilon_{c,\tau}
\end{aligned} \tag{1}
$$

We define $Y_{c,\tau}$ as the number of granted patent applications in commuting zone c during period τ. Depending on the specification, Y refers to either the universe of patents or a specific subset. c refers to the commuting zone where the inventor resides at the time of application. τ denotes three-year periods, which we use to account for lagged responses of innovation to changes in local economic activity. This choice is supported by prior literature showing that innovation typically reacts to shocks with a delay of 2–3 years (Popp, 2002; Dugoua, 2023; Dechezleprêtre et al., 2025). Online Appendix Table OA2 shows that our results are robust to using shorter and longer period lengths, and that the strongest patenting response occurs 2-3 years after the boom, aligning with previous literature and corroborating our choice of three years as baseline period length.

National Oil&Gas Employment$_\tau$ is computed as the average over the three-year period. *Oil&Gas Endowment*$_{c,T}$ equals oil and gas reserves as of 1960 that are economically recoverable in time period T, divided by commuting zone area to account for size. Following Allcott and Keniston (2018), in period $T = \{1969, 2000\}$, the endowment measure only includes reserves that are economically recoverable using "conventional" extraction techniques, while in $T = \{2001, 2012\}$, endowment also includes reserves that are economically recoverable using hydraulic fracturing ("fracking") techniques (see Section 3 for details). To ease the interpretation of $\hat{\beta}_1$, we scale oil and gas endowment by its standard deviation (based on the 1969–2000 measure) across commuting zones, which equals around five million dollars per square mile.

[3] While Goldsmith-Pinkham et al. (2020) derive their results in an instrumental variables setting, they explicitly note that "the insights of this paper still apply when Bartik is used in the reduced form" (p.2589).

$\gamma_{s,\tau}$ are state[4] times (three-year) period fixed effects, and $\delta_{c,T}$ are commuting zone times period T fixed effects. The inclusion of $\delta_{c,T}$ implies that we have two dummies per commuting zone rather than one: one for 1969-2000 and one for 2001-2012. Intuitively, we thus demean the data separately over the pre-fracking period and over the fracking period, rather than over the entire sample period. This allows us to estimate a single regression over the full sample period rather than splitting the sample, and yet considers the endowment that is relevant in the respective time frame (1969-2000 versus 2001-2012).[5] Besides that, the described fixed effects structure implies that any local patenting in fracking technology that makes local shale oil and gas endowment economically recoverable (thereby raising total endowment above conventional endowment) would not make our model suffer from reverse causality at the turn of the century.[6] $\delta_{c,T} * \tau$ stands for commuting zone-specific linear trends, where for each zone we estimate one linear slope coefficient for the pre-fracking period and one for the fracking period.

β_1 indicates the effect of a local oil and gas boom on a commuting zone's patent count, relative to commuting zones that experience a smaller or no boom (see Section 4.4 for *absolute* effects). $\hat{\beta}_1$ is an unbiased estimator of this effect if, conditional on our rich fixed effects structure, the shift-share interaction does not correlate with other factors that influence local innovation. While this appears unlikely, one might be concerned that oil and gas endowment correlates with other local characteristics such as income per capita, and changes in national oil and gas employment might in turn affect richer commuting zones differently than others. To address such concerns, as a first step we evaluate the correlation of local oil and gas endowment with multiple commuting zone characteristics, using conventional versus total endowment in different specifications. The results (see Table A2) show that none of the included char-

[4] Some commuting zones span across multiple states. Following previous literature (Autor and Dorn, 2013), in these cases we assign a commuting zone to the state containing the largest share of the commuting zone's population in 1969.

[5] Demeaning over the whole sample period would mean that in every three-year period we subtract, from the current realization of our key interaction term, the average value across 1969-2012 of *endowment* $\times$ *national oil and gas employment*. Since this average is driven upwards by the inclusion of fracking reserves starting from 2001, this would imply that in the pre-fracking period 1969-2000, we subtract a value that is influenced by a later change in endowment which is completely irrelevant at the time.

[6] Suppose we only included δ_c rather than $\delta_{c,T}$ into Equation (1), implying only one fixed effect (dummy) per commuting zone. In that case, our model would allow for an increase in local patenting in fracking technology (on the left-hand side) to lead to an increase in endowment (on the right-hand side) between the last pre-fracking and the first fracking three-year period. However, including $\delta_{c,T}$ implies estimating our results only based on changes *within* the periods $T = \{1969, 2000\}$ and $T = \{2001, 2012\}$, but not across them.

acteristics significantly correlate with endowment in all specifications. While this eases the above-described identification concern, we further address the concern through a robustness check in which we augment Equation (1) with interactions of national oil and gas employment and the commuting zone characteristics used in Table A2. This approach is motivated by the fact that "[t]he exogenous shares assumption (...) is about the exogeneity *conditional on observables*" (Goldsmith-Pinkham et al., 2020, p.2598, emphasis in original). The results are robust to using the augmented specification (see Table 9, column 2).

3 Data

In this section, we describe our sample, as well as the data on oil and gas endowment and patenting. Additional data sources are described in Section 4 as they become relevant. Further details on the data used in the analysis are provided in Online Appendix Section OA3.

Our sample covers all US states except Alaska and Hawaii. We start our analysis in 1969 since data on most variables are not available for earlier years, and we end in 2012. We aggregate all data from the county to the commuting zone level because commuting zones better represent local labor markets. Several counties split or merged over our sample period, and the Bureau of Economic Analysis (BEA) merges selected counties in its regional data reporting. We address this by defining commuting zones that are consistent over 1969-2012, of which there are 759 in our sample. As of 1969, the average commuting zone had 260,000 residents, while the median population was 75,000. Table A1 reports descriptive statistics.

3.1 Oil and gas endowment

For our shift-share interaction, we need commuting zone-level data on oil and gas reserves at the beginning of our sample period. While such disaggregated data are not publicly available, Hunt Allcott and Daniel Keniston (Allcott and Keniston, 2018) generously shared with us county-level data on oil and gas reserves in 1960, which we aggregate to the commuting zone level. To account for the fact that combining horizontal drilling with hydraulic fracturing made large amounts of existing reserves economically recoverable in the early 2000s, we follow Allcott and Keniston (2018) in using two distinct reserve measures. "Conventional reserves" are those that are in the ground in 1960 and economically recoverable using conventional (i.e., non-fracking) extraction methods, while "total reserves" equal 1960 reserves that are economically recoverable using either conventional or fracking techniques. Intuitively, the way to think about this distinction is that reserves in 1960 consist of two parts: conventional reserves, which are economically recoverable throughout our entire sample period; and fracking reserves, which were obviously already under the ground in 1960 but only "switch on" in 2001, as new technology makes this portion of 1960 reserves economically recoverable. Both reserve measures include "undiscovered reserves", which make up 35% of total reserves at the national level. Undiscovered reserves are "postulated from geologic knowledge and theory to exist outside of known fields" (Schmoker and Klett, 1999, p.1), making them more closely related to geology and less related to exploration effort.

Since local oil and gas reserves were not comprehensively assessed as early as in 1960, Allcott and Keniston (2018) compute both conventional and total reserves by combining post-1960 reserves data and historical production data. This is done in the following way:

$$Oil\&Gas\ Reserves^{conventional}_{c,1960} = Proven\ Reserves_{c,1995} + Undisc.\ Res._{c,1995} + \sum_{t=1960}^{1995} Production_{c,t}$$

$$Oil\&Gas\ Reserves^{total}_{c,1960} = Proven\ Reserves_{c,2011} + Undisc.\ Res._{c,2011} + \sum_{t=1960}^{2011} Production_{c,t}$$

To add up oil and gas units in the above two formulas, we follow Allcott and Keniston (2018) and transform physical to dollar values using average (real 2010) prices over 1960-2011.

Scaling reserves by commuting zone area (including both land and water area) to account for variation in size yields the following distinct endowment measures, which we use separately for the time periods 1969-2000 and 2001-2012 in our empirical analysis:[7]

$$Oil\&Gas\ Endowment_{c,T=\{1969,2000\}} = \frac{Oil\&Gas\ Reserves^{conventional}_{c,1960}}{Area\ in\ Square\ Miles_c}$$

[7] After scaling by the standard deviation of conventional endowment (see Section 2), the two endowment measures jointly correspond to $Oil\&Gas\ Endowment_{c,T}$ in Equation (1). Note that in computing the two endowment measures, we make the same choices as Allcott and Keniston (2018), but write down the two measures in a different way so that the notation is equivalent to our chosen notation in Section 2.

$$Oil\&Gas\ Endowment_{c,T=\{2001,2012\}} = \frac{Oil\&Gas\ Reserves^{total}_{c,1960}}{Area\ in\ Square\ Miles_c} \tag{2}$$

595 of 759 commuting zones have non-zero conventional endowment, while 613 commuting zones have non-zero total endowment. The average conventional endowment across all commuting zones equals \$1.5 million per square mile, and the standard deviation equals \$4.6 million. For total endowment, the average and standard deviation equal \$2.9 and \$7.0 million, respectively. Figure 2 maps conventional endowment (Panel I), total endowment (Panel II), and the difference between the two (Panel III). Panels I and II illustrate that oil and gas endowment is widely spread across the United States and does not only occur in rural areas of the country. More specifically, among the 379 commuting zones with above-median (below-median) conventional endowment per square mile, 38 (34) are metropolitan, 234 (262) are urban non-metropolitan, and 107 (83) are rural.[8] In terms of population density, zones with above-median endowment had 50 residents per square mile in 1969, while below-median endowment zones had 75.

3.2 Patents

We measure innovation using patent data. Patents are typically highly correlated with R&D expenditure and other indirect measures of innovation (Griliches, 1990), and the richness of patent data allow us to study local innovation dynamics over a long period of time. Our main data source is the Worldwide Patent Statistical Database (PATSTAT), which contains the population of all patents filed globally since the mid-1960s. PATSTAT collects a wide range of information for each patent application, including bibliographic data—such as filing,

[8] Examples of metropolitan commuting zones with above-median oil and gas endowment are the commuting zones including Austin, New Orleans, Los Angeles, or Tulsa, respectively. In terms of methodology, we bring the Rural-Urban Continuum Codes, published by the US government's *Economic Research Service* (ERS), to the commuting zone level. In the original data, each county is assigned a value from 0 to 9, ranging from "Central county of metro areas of 1 million population or more" (Code=0) to "Completely rural or less than 2,500 urban population, not adjacent to a metro area" (see Online Appendix Figure OA1 for details). Counties with value 4 to 9 are classified as "non-metropolitan". For our purposes, we compute the average value across all counties within a commuting zone, based on the 1974 edition of the data (earlier data are unavailable). We define commuting zones with an average value of 4 or greater as non-metropolitan, which is true for 90% of commuting zones. We further distinguish non-metropolitan commuting zones into urban (rural-urban code greater than or equal to 4, and less than 8) and rural (rural-urban code greater than or equal to 8).

grant, and publication dates, as well as inventor and applicant details—and additional data on technology fields, family links, and citations. We merge these data with information on disambiguated inventors' addresses from PatentsView (1976-2012) and HistPat (1969-1975). Our sample consists of all patents filed in the period 1969-2012 and granted by the USPTO, with at least one inventor based in the United States and non-missing information on the patent's technological field. This amounts to roughly 2.5 million patents.

We leverage these rich data to construct a novel panel dataset listing the number of granted patents in technology j filed in year t by inventors located in commuting zone c. To construct this dataset, we first identify the research team behind a patent and the inventors' location at the time of patent filing. We use the inventor address rather than the applicant address, as the latter (e.g., the address of the firm's headquarter) may not reflect where the invention has taken place (OECD, 2009). Second, we observe the filing and granting dates for each patent office where a patent is submitted. These data enable us to date a patent based on its earliest filing date (priority date), which most closely approximates when the innovation project was conducted. Third, we observe technology classes, which are categorized by the patent office based on the patent's technical characteristics following the International Patent Classification (IPC) scheme. When a patent comprises several technology classes, we count it fractionally, with a weight proportional to the frequency of each technology class. Similarly, if the inventors on a patent are located in multiple commuting zones, we count the patent fractionally.[9] Further details on the construction of patent variables are provided in Section OA2.

Patent counts vary greatly across commuting zones, with 5% of commuting zones accounting for nearly three quarters of all patenting activity throughout 1969-2012. The average commuting zone produces 74 patents per year, while the median patent count is four. 18% of commuting zone-years have zero patents; by aggregating to three-year periods (see Section 2), the fraction of cells with zero patents falls to 9%. Patents per 100,000 residents average to nine per year. Patenting activity is similar across oil and gas-rich and -poor commuting zones: the former (defined as zones with above-median conventional endowment) average 77 patents per year,

[9] For example, suppose that the patent office assigns technology classes A01B and A01C to patent P; then we count 0.5 patents in technology A01B and 0.5 patents in technology A01C. Suppose further that there are four inventors, three residing in commuting zone A and one residing in commuting zone B; then we have 0.5×0.75 patents in technology A01B and 0.5×0.75 in technology A01C in commuting zone A, and 0.5×0.25 patents in technology A01B and 0.5×0.25 patents in technology A01C in commuting zone B.

while the latter average 71.

Given our setting, it is important to distinguish patents in oil and gas technology from non-oil and gas patents. Following the Derwent World Patents Index (Clarivate Analytics, 2020), we define the IPC classes C10G, C10L, C10M, and E21B – which jointly cover all aspects of the oil and gas industry – as oil and gas classes.[10] We define a patent as an oil and gas patent if it contains at least one of the four classes. These patents make up 2% of all patents over 1969-2012. While our analysis centers on non-oil and gas patents, our data reveal that oil and gas patenting is geographically widespread and increases with local oil and gas supply, creating variation to study local oil and gas patenting during booms (see Online Appendix Section OA1.6).

97% of patents in our sample are owned by companies, rather than individuals or institutions such as universities. Some of these patents might reflect innovation by single inventors that transfer ownership to their one-person company, but the data do not allow to identify the share of such patents.[11] 53% of all patents are owned by firms included in the Compustat database, all of which are publicly listed.[12] For such companies, one might wonder to what extent they centralize their R&D in large labs or instead spread their R&D across multiple labs and commuting zones, and what this means for identification. Our data reveals that Compustat firms in our sample (which produce an average of 20 patents per year) carry out patented innovation in an average of 3.3 commuting zones in a given year (median=2). On average, the commuting zone where a Compustat firm produces the most patents accounts for

[10] See Online Appendix Section OA3 for details on the individual classes. The DWPI classification has been frequently used in the academic literature (see e.g. Duch-Brown and Costa-Campi, 2015) and yields oil and gas patent counts that are comparable to those reported in non-academic outlets. For example, Reuters reports 2,188 oil and gas patents filed in the United States in 2013 (see `https://www.reuters.com/article/us-energy-shale-research-idUSKBN0F411B20140629`), while applying the DWPI definition to our data results in 2,513 oil and gas patents in the same year.

[11] The patent data do not contain relevant firm-specific information such as the number of employees. However, they do contain information on the type of applicant: 57% of patents in our sample (1969-2012) were applied for by a company, 32% by individuals, 2% by universities, and 2% by government non-profit organizations (for 7% of patents, the applicant type is unknown). To the extent that single inventors apply for a patent themselves rather than through their one-person company, 32% can be interpreted as an upper bound of the share of single-inventor patents. However, the actual share is most likely well below this upper bound, since even among the sample of firms included in the Compustat database – which are often very large – 12% of patents are applied for by an individual rather than the company itself.

[12] We map patents to Compustat firms using the crosswalk from Autor et al. (2020). Due to data availability in this crosswalk, all descriptive statistics and regressions involving "Compustat status" are based on the sample period 1975-2006.

71% of that firm's total patents during our sample period (median=70%). For large firms such as Compustat firms, innovating in multiple commuting zones might imply that the oil and gas shock in one of those commuting zones (call it A) has a smaller relevance on the firm's patenting in A compared to other firms or individual inventors that only operate in A. In fact, this point on the relevance of local shocks applies more broadly: for example, the patenting activity of firms that innovate in only one commuting zone (call it B), but generally operate in multiple zones, may be influenced not only by the oil and gas shock in B, but also by shocks in other zones. From an identification perspective, such "network factors" create measurement error in the de-facto shock exposure of multi-location firms in our sample, implying that our estimates should be regarded as a lower bound of the actual effect size. We come back to this in Section 4.3.1, when analyzing patenting by Compustat firms in the context of financial constraints as a potential mechanism.

4 Results

We structure the presentation of our results as follows. First, we show that our shift-share interaction is a significant driver of various indicators of local economic activity, which lends empirical support to the chosen boom measure. Then, we present our main results. In Section 4.1, we discuss the boom effects on local patenting activity; Section 4.2 documents heterogeneous effects across different types of commuting zones and across other dimensions; Section 4.3 tests for multiple mechanisms; Section 4.4 examines geographic spillovers; Section 4.5 explores the role of varying opportunity costs of innovation across industries during booms; and Section 4.6 presents our main robustness checks. Online Appendix Section OA1 presents further robustness checks and additional results such as the impact of local booms on patent quality as measured by citations.

Oil and gas booms stimulate local economic activity

In Table 1, we analyze the impact of oil and gas booms on various measures of economic activity, at the commuting zone level. We do so by estimating Equation (1) using OLS, and measuring the dependent variables as the average over the respective three-year period. The results show that in the commuting zone with an initial oil and gas endowment equal to one standard deviation (approximately five million dollars per square mile), a 100-log-point increase in national oil and gas employment between two three-year periods leads to a rise in population by 1.9%, employment by 3.7%, earnings per worker by 2.2%, personal income per capita by 1.8%, and GDP per capita by 4.6% (see columns 1-5). Moreover, column 6 shows a 6.2% rise in own-source local government revenue (thus excluding state and federal transfers, but including property tax revenue, for instance) during booms.[13] One can think of these results as evidence on a "conceptual first stage" of our reduced-form approach.

4.1 Main Results

Table 2 reports estimates on the effect of oil and gas booms on local innovation. Column 1 shows that in a commuting zone with a $5 million per square mile oil and gas endowment, a 100-log-point increase in national oil and gas employment leads to an increase in the number of patents by about 8.3%. Column 2 shows a significant increase of similar magnitude for non-oil and gas patents. In column 3, we study innovation in the oil and gas sector and find that the number of oil and gas patents significantly *decreases* during oil and gas booms.[14] While we discuss reasons for this decrease in Section 4.5, Sections 4.2-4.4 focus on non-oil and gas patenting as outcome variable, as this category represents 98% of total patenting during our sample period.

In column 4 of Table 2, we complement our baseline analysis at the commuting zone level with a within-*firm* level analysis. Taking this firm perspective is not only interesting per se but also helps us to understand mechanisms behind our results, as we discuss in Section 4.3. We restrict our sample to firms that patent in more than one commuting zone, and test whether

[13] Data for the variables used in columns 1-5 are obtained from the Regional Economic Accounts database provided by the Bureau of Economic Analysis (BEA). For the analysis in column 6, we source five-yearly data from the Census of Governments. We aggregate the data across all county governments in the commuting zone and adjust Equation (1) to account for the different data frequency, measuring the dependent variable and national oil and gas employment in the years when Census of Governments data are available.

[14] Online Appendix Table OA3 shows that these effects are not significantly different across the pre-fracking period 1969-2000 and the fracking period 2001-2012. Further note that in Table 2, the number of observations in column 3 is smaller than in columns 1-2. This is because the command *ppmlhdfe* (Correia et al., 2020), which we use to estimate our Poisson regressions, drops separated observations to avoid statistical separation issues, and these omissions occur more frequently in the presence of many zeros in the dependent variable.

a boom in a given commuting zone makes the firm patent more in that zone relative to other zones where the firm is active (in terms of patenting). The specification looks as follows:

$$
\begin{aligned}
Y_{i,c,\tau} = exp\Big(&\beta_1[Oil\&Gas\ Endowment_{c,T} \times ln(National\ Oil\&Gas\ Employment_\tau)] \\
&+ \delta_{c,T} + \delta_{c,T} * \tau + \gamma_{s,\tau} + \omega_{i,\tau}\Big) + \epsilon_{i,c,\tau}
\end{aligned} \tag{3}
$$

where $Y_{i,c,\tau}$ is the total number of patents by firm i in commuting zone c in three-year period τ; and $\omega_{i,\tau}$ are firm times three-year period fixed effects which account for drivers of the firm's overall patenting activity in period τ. Our sample includes 48,000 firms that patent in multiple commuting zones in at least one three-year period, of which around 7,000 are Compustat firms. For each of these firms, in order to focus on the empirically relevant commuting zones, we drop zones where the firm has zero patents throughout our entire sample period, and create a balanced panel with the remaining zones.[15] The results show that a local boom makes multi-location firms patent by 2% more in a booming commuting zone compared to non-booming zones where the firm also patents during our sample period. This result shows that local booms raise patenting activity even through the lens of one particular firm, accounting for the firm's general patenting activity at the time. Note that the magnitude of this column 4-coefficient is not directly comparable to the coefficient in column 1. This is (i) because of the different unit of observation (for example, if a boom in one commuting zone promotes a firm's patenting activity also in other zones, then such spillover effects would drive down the column-4 coefficient compared to the column-1 coefficient); (ii) because column 4 only considers firms patenting in multiple commuting zones; and (iii) due to a slightly different sample period due to data availability (see Section 3).

4.2 Heterogeneous effects

An important question is whether our results are driven by a particular type of commuting zone. We therefore examine heterogeneity in innovation responses across several dimensions, including urbanization status, baseline patenting intensity, human capital availability, and college

[15] Our results are robust to a more restrictive analysis where we build a balanced panel based on only those commuting zones where the firm patented in the very first period in which the firm enters our sample (i.e. the period in which the firm first patented in one or more zones, during our sample period).

presence. Moreover, we study heterogeneous effects across well-established and novel technologies within a commuting zone, and across booms and busts. The results are reported in Table 3. Column 1 repeats the baseline results from column 2 of Table 2. In column 2, we restrict the sample to commuting zones with a below-median patents per capita ratio over 1969-2012.[16] The results are similar for this subsample, indicating that a high patenting activity is not a prerequisite for experiencing positive effects on innovation during economic upswings. In column 3, we test for heterogeneous effects across more or less urban commuting zones.[17] The results show that compared to metropolitan commuting zones, non-oil and gas patenting rises by about twice as much in non-metropolitan commuting zones during local economic booms. In column 4, we distinguish non-metropolitan commuting zones into more versus less urban ones. The results show that the innovation effects of a boom are largest overall for commuting zones that are non-metropolitan but urban, rather than rural or metropolitan. In column 5, starting from the specification of column 4, we add interaction terms with several other variables, which we measure at the beginning of our sample period whenever possible: patenting intensity, computed as total patents over 1960-1969 divided by population in 1969; human capital, measured as the share of population aged 25 or older with at least one year of college education, as of the year 1970; and college density, which we define as the number of employees in colleges, universities, and professional schools divided by total population, measured in 2018 since earlier data are unavailable.[18,19] All three variables are scaled by their standard deviation. The results show that commuting zones with a higher initial patenting intensity experience a significantly larger rise in patenting during boom times. This suggests that booms reinforce existing innovation

[16] We compute this ratio as total patents in 1969-2012 divided by the sum of all annual population counts of the commuting zone over 1969-2012. The median equals six patents per 100,000 residents, per year.

[17] See footnote 8 for details on this classification. During our sample period, the 73 metropolitan commuting zones produced an average of 1,687 patents per three-year period (or 75% of total patents over 1969-2012); the 496 urban non-metropolitan zones produced 80 (24%); and the 190 rural zones produced 6 (1%).

[18] County-level education data, which we aggregate to the commuting zone level, are available from the ERS. We use the measure described in the text because data on the number of residents who actually obtained a college degree are not available prior to 1990. Our definition of college density is inspired by Valero and Van Reenen (2019), who use a region's number of universities per capita in their analysis. Since we have information on a college's number of employees, we are able to account for heterogeneity in college size across commuting zones, thereby refining their measure. Data are only available for the academic year 2018-19, and are obtained from the *Homeland Infrastructure Foundation-Level Data* (HIFLD).

[19] In order to interpret the coefficient on *ln(National oil&gas employment)* $\times$ *endowment* in column 5 as the effect in the average commuting zone (across the five sources of heterogeneity we test for), we demean the rural-urban dummies and the three additional variables before performing the regression.

capacity rather than create it from scratch. The coefficients on the human capital and college density interactions are not statistically significant.[20]

Next, we study whether the rise in non-oil and gas patenting reflects more innovation in historically familiar or novel technologies, from the perspective of a specific commuting zone. To do so, we first calculate a commuting zone's relative historical patenting experience in each 2-digit technology class. We compute this as the ratio of patents in the 2-digit IPC class to total patents, both measured as the sum over 1960-1969. We then proceed by estimating Equation (1) at the 2-digit IPC class by commuting zone level, including an interaction of the boom variable with the technology's historical share. Besides state times three-year period fixed effects and the trends in Equation (1), we include 2-digit IPC class times commuting zone times century fixed effects and 2-digit IPC class times three-year period fixed effects. The results (see column 6 of Table 3) indicate that while the rise in non-oil and gas patenting is not exclusively driven by technology classes the commuting zone has historically patented in, classes that are historically more prominent do experience a comparatively larger rise in patenting. We interpret this as evidence of path dependence in innovation, consistent with Aghion et al. (2016) and Manso et al. (2023).

Finally, we test whether the baseline effects are symmetric across booms and busts. As Figure 1 shows, national oil and gas employment continually rose between the three-year periods 1969-1971 and 1981-1983; then continually declined until the end of the century; and then continually rose from the 2001-03 period to the 2010-12 period. Therefore, we define the intermediate time frame 1984-2000 as bust period, and interact our boom measure with a 1984-2000 "bust dummy". The coefficient on this extra interaction term is close to zero and statistically insignificant (see column 7 of Table 3), suggesting that busts reduce innovation in the same magnitude as booms increase innovation.

4.3 Mechanisms

In this section, we test various mechanisms that may explain the rise in non-oil and gas innovation during economic upswings. Based on previous literature and our specific setting, we

[20] While recent literature shows that local corporate innovation is positively affected by local universities' innovation (Hausman, 2022), our results suggest that local universities do not mediate the impact of local economic booms on patenting.

evaluate the relevance of relaxed financial constraints, input-output linkages, and agglomeration effects. Online Appendix Section OA1.5 analyzes further mechanisms that appear less relevant a priori or have been contested by previous studies in other contexts, namely public finance, strategic timing of patent filings, and inventor-level wealth effects.

4.3.1 Relaxed financial constraints?

Several studies show that credit constraints have a negative impact on innovation (Ouyang, 2011; Aghion et al., 2012; Amore et al., 2013; Gorodnichenko and Schnitzer, 2013; Nanda and Nicholas, 2014; Akcigit et al., 2017). Moreover, there is evidence that oil and gas windfalls can alleviate credit constraints, for example by raising bank deposits (Gilje et al., 2016; Gilje, 2019). Given these findings, the rise in non-oil and gas patenting may be explained by relaxed credit constraints in boom times. We test for this potential channel in two ways. First, we test whether our baseline results are driven by industries that typically depend more on external finance (see Rajan and Zingales, 1998), and may thus benefit more from increased credit availability during booms. This requires us to map patents – and thus technologies – to industries, which is challenging because existing techniques often struggle to distinguish between an industry's development and its use of a technology.[21] We overcome this limitation through a two-step approach. First, we use the fact that for each patent in our dataset, we know whether it is owned by a firm included in Compustat, and if so, in which 4-digit SIC industry the firm operates. By examining Compustat firm patents, we can then infer the association between certain manufacturing industries and specific technologies. For instance, if firms in a particular industry tend to file patents in technologies X and Y, we can map that industry to technologies X and Y. Using this method, we subsequently map all patents in our analysis to industries, irrespective of whether the patent is owned by a Compustat firm or not. We then compute a measure of dependence on external finance for each of the 20 2-digit SIC manufacturing industries following the procedure of Rajan and Zingales (1998),[22] and correlate it with the

[21] For example, some of these techniques are based on identifying text similarity between patent technology and industry descriptions (see e.g. Lybbert and Zolas, 2014).

[22] External finance dependence is computed as $\frac{Capital\ expenditure–cash\ flow\ from\ operations}{Capital\ expenditure}$. We first compute this ratio at the firm level, averaging both the numerator and the denominator over 1971-2012 (earlier data are unavailable) and then taking the ratio of the two averages. Thereafter, we define the industry-wide ratio as the median observation across all firms in the industry.

industry's patenting activity across local booms and busts. We do so by estimating Equation (1) at the commuting zone – (2-digit) industry level, replacing commuting zone times century fixed effects with commuting zone times century times industry fixed effects, and additionally including industry times three-year period fixed effects. In column 1 of Table 4, we estimate this specification without interacting the boom variable with the external finance dependence measure, for comparison. In column 2, we include the interaction. To ease the interpretation of our results in column 2, we scale the values of external finance dependence by the variable's standard deviation across all 20 industries. The results suggest that industries that depend more on external finance do not raise patenting by more than others during boom times, which is not supportive of the finance channel hypothesis.

Since the validity of this conclusion hinges on the quality of our technology-industry match, we perform a second test. This test exploits that all firms in Compustat are publicly listed and are thus typically less financially constrained than other firms (Saunders and Steffen, 2011), implying that a smaller rise in patenting by Compustat firms would be consistent with a finance channel. To account for the possibility that Compustat firms tend to patent in technology classes that are differently affected by local booms, we estimate a specification at the commuting zone – 2-digit technology class – firm type level and include appropriate fixed effects.[23] In column 3 of Table 4 we estimate this specification without an extra interaction, while in column 4 we add an interaction of our boom variable with a dummy indicating patenting by Compustat firms.[24] The results of column 4 suggest that patenting by Compustat firms rises by less in boom times, but the interaction coefficient is small and not statistically significant. Together with the results in column 2, this evidence suggests that a finance channel is unlikely to (fully) explain our results. Further weakening the case for a finance-based explanation, the negative

[23] This means that for each commuting zone, 2-digit technology class, and three-year period "cell", we have two observations: the number of patents owned by Compustat firms, and the number of patents owned by other entities. In terms of fixed effects, besides state times three-year period fixed effects and the trends in Equation (1), we include 2-digit IPC class times commuting zone times century fixed effects and 2-digit IPC class times three-year period fixed effects.

[24] Besides their role as reference for columns 2 and 4, the specifications in columns 1 and 3 of Table 4 also address the potential identification concern that changes in national oil and gas employment might be correlated with shocks to specific sectors or technology classes. If this were the case, *and* these specific sectors or technology classes were over- or under-represented in oil and gas-rich commuting zones, then our baseline results would be biased. However, given that the specifications in columns 1 and 3 include sector times period and technology class times period fixed effects, respectively, and produce similar results as our baseline specification (Table 2, column 2), such concerns are invalidated.

(albeit insignificant) coefficient on the Compustat interaction may be driven by other differences across Compustat and other firms, which are unrelated to finance. For example, Compustat firms are more likely to operate in multiple commuting zones, which could imply that a local boom has a smaller effect on their local patenting compared to other firms.

Finally, we perform a third test for the financial constraint channel. We do so by estimating the firm-level regression of column 4 of Table 2 for Compustat firms only, implying that all firms in the sample are both listed and patent in more than one commuting zone. We hypothesize that if booms weakened financial constraints for innovative firms located in the booming commuting zone, then for multi-location Compustat firms, we should not observe a stronger patenting rise in the booming zone relative to other zones where the firm patents. The reason is that large, listed firms typically have a well-functioning internal capital market, enabling them to channel funds to other locations where the funding may be of better use (see Stein, 1997, for theory on such within-firm transactions).[25] However, the results (see column 5 of Table 4) show that listed multi-location firms do increase non-oil and gas patenting in booming commuting zones relative to other, less-booming commuting zones where they operate in. This piece of evidence is inconsistent with finance as the (main) driver of our results. To the contrary, the result suggests a driving factor which is specific to the booming zone in the sense that it is non-transferrable to other areas – such as precisely agglomeration economies. We test this channel more directly further below.

4.3.2 Input-output linkages

In this section, we examine whether the rise in non-oil and gas patenting during local booms is driven by supply chain linkages to the oil and gas sector. We do so by testing whether patents in industries that are upstream to the oil and gas sector increase patenting by more than others. The key idea is that “the amount of invention is governed by the extent of the market” (Schmookler, 1966, p.104), such that an expanding oil and gas sector that demands more intermediate inputs may benefit innovative local input suppliers.

Using BEA Input-Output tables, we first calculate the direct and indirect share of a 4-digit

[25] Note that even if new external finance from the booming commuting zone is tied to being used locally (local banks might impose such requirements), this likely frees up internal funds which can then be transferred to other locations within the spatial firm network.

industry's output that is sold to the oil and gas sector.[26] We define an industry as upstream if the sum of the direct and indirect output share is larger than 0.1%. To isolate upstream effects, we exclude downstream industries (defined as those with a direct oil and gas *input* share exceeding 0.1%) from this group.[27] For comparison, we also identify "non-linked" industries, which are neither upstream nor downstream. We then aggregate non-oil and gas patents into patents by industries that are upstream (but not downstream) versus patents by non-linked industries, and use these as dependent variables in Equation (1). The results (see columns 6-7 of Table 4) suggest that supply chain linkages are unlikely to explain the local rise in non-oil and gas innovation: we find a positive and significant innovation response in upstream industries, but the response is even larger in non-linked industries.

4.3.3 Agglomeration effects

The preceding sections evaluated finance and supply chain linkages as potential mechanisms and found no evidence that either of them explains the observed increase in innovation. Turning to an alternative mechanism, several studies highlight the positive impact of agglomeration on innovation (Carlino and Kerr, 2015; Moretti, 2021; Andrews, 2023), suggesting agglomeration economies as a possible channel. A priori, at least two types of agglomeration effects may explain our results, which we test separately next.

Testing for in-migration of inventors into booming commuting zones

First, it could be that an economic boom induces inventors to migrate to a commuting zone and continue to patent there. Considering the results of Moretti (2021), such in-migration could also make incumbent inventors more productive. We test for an inventor migration channel by classifying each patent in our data into three mutually exclusive and exhaustive categories: patents by incumbent inventors, patents by in-moving inventors, and patents by new inventors. We classify a patent in year t as an incumbent inventor's patent if the inventor's physical address on her most recent pre-t patent is in the same commuting zone as the address on the

[26] The direct output share is the share of an industry's output purchased by the oil and gas sector. The indirect share is the share of output purchased by the oil and gas sector through an intermediate industry. We do not consider higher distance linkages.

[27] Around 30% of industries in our sample are upstream and not downstream. Note that we do not add indirect input shares to the downstream measure because they largely reflect a sector's electricity intensity, given that the US electricity generation sector uses large amounts of natural gas.

year t patent. Similarly, a patent in year t is classified as an in-mover's patent if the address on the inventor's most recent pre-t patent belongs to a different commuting zone. Patents by new inventors (as of year t) are patents for which the inventor appears in our entire dataset (spanning all commuting zones and years) for the first time in year t. In the average commuting zone, 58% of patents are filed by incumbents, 5% by in-moving inventors, and 37% by new inventors.[28] We then regress each of these patent counts separately on our boom variable. Since the described classification of patents requires a level of detail that is only available from 1976 onwards, our sample period is 1976-2012. We perform all regressions at the annual level rather than at the three-year period level, because we can only meaningfully classify patents into the three described categories at the annual level.[29]

Table 5 presents the results, showing that local booms do not significantly increase patenting by in-moving inventors compared to non-boom periods (see column 3). This suggests that temporary local economic booms, such as those driven by oil and gas shocks, are not a major factor in inventor migration – consistent with prior literature emphasizing taxation (Akcigit et al., 2016; Moretti and Wilson, 2017; Akcigit et al., 2022) and technology- or industry-related determinants (Moretti, 2021). We also find no conclusive evidence that patenting by new inventors (see column 4) drives our baseline results. Instead, the main driver of the overall rise in non-oil and gas patenting (see column 1) appears to be increased patenting by incumbent inventors (column 2).

Agglomeration economies that raise the productivity of incumbent inventors

Why do incumbent inventors produce more patents during local economic booms? The reason may be agglomeration economies that take a different shape than the in-migration of inventors. For example, Carlino et al. (2007) show that a rise in US city size is positively associated with a

[28] Same as in our baseline analysis, we count patents fractionally. Thus, when a patent involves two inventors located in two different commuting zones, we treat the patent as two observations, assigning half of the patent to each zone and classifying the geographic status (incumbent, in-mover, or new) separately for each "half-patent". Note that our fractional patent count is based on the priority year only, rather than also on subsequent re-filings.

[29] For illustration, consider the three-year period 2001-2003, and an inventor who in 2000 files a patent in commuting zone A, in 2001 files a patent in commuting zone B, and in 2002 files a patent in commuting zone A. If we were to classify patents into the above three categories from a three-year period perspective, we would for instance need to classify the 2002 patent as an incumbent's patent, which would make little sense because the inventor patented in another commuting zone in the year before, i.e. in 2001.

rise in the city's patents per capita, and attribute this link to positive agglomeration economies. In light of this evidence, our results of increased population, employment, and patenting activity by *incumbent* rather than other inventors are consistent with a channel through which more people and jobs in a booming area make incumbent inventors more productive – be it through enhanced labor market matching, input sharing, or knowledge spillovers (Carlino and Kerr, 2015). Consistent with this idea, Hunt and Gauthier-Loiselle (2010) argue that "[e]ven immigrants who do not patent themselves may increase patenting by providing complementary skills to inventors, such as entrepreneurship." (p.32) Further support for such agglomeration economies is provided by a finding of Carlino et al. (2007), combined with evidence of our own. Specifically, the positive association between city size and patent intensity in their data does not hold for cities with a 1990-population above 500,000 or 750,000 (depending on the specification), leading Carlino et al. (2007) to conclude that "the benefits of urban scale are realized for cities of moderate size" (p.401). This aligns with our finding that boom effects are largest in urban non-metropolitan commuting zones (90% of which had less than 500,000 residents in 1990), signaling agglomeration as a key driver of our results – especially in mid-sized areas where limiting factors such as congestion are less relevant.[30]

We proceed with several empirical exercises to test the agglomeration hypothesis more directly. We hypothesize that from the perspective of incumbent inventors, the above-listed types of agglomeration economies (matching, sharing, spillovers) increase with in-migrants' skill level and the extent to which they work in creative occupations. Therefore, we analyze the impact of local booms on different population segments, starting with a distinction of college- versus non-college-educated residents. Local-level data on educational attainment are only available for the years 1990 and 2000 from the Population Census, and thereafter as moving five-year averages from the annual American Community Survey (ACS), starting with 2006-2010. We therefore measure both the dependent variables and our key interaction term in 1990, 2000, as

[30] This discussion resonates with a literature showing diminishing returns of agglomeration outside an innovation context, which has been reviewed by Combes and Gobillon (2015).

the average over 2006-10, and as the 2011-2015 average.[31] The results (see Table 6) show that oil and gas booms raise not only total adult population (column 1), but also both college- and non-college-educated populations (columns 2-3). Because of the low data frequency, a possible interpretation is that the observed increase in college graduates during boom periods may reflect educational choices by incumbent residents rather than in-migration. However, column 4 of Table 6 shows that the results on total adult population are robust to measuring booms annually, and the literature typically finds that natural resource-driven booms *negatively* affect residents' schooling (Mosquera, 2022; Kovalenko, 2023). Therefore, we mainly attribute our findings to in-migration.

We continue by testing the boom impact on the number of workers in creative occupations, exploiting the "creative class" concept introduced by Florida (2002). This exercise addresses the concern that most college graduates may work in occupations that are not closely related to the innovation process. We use a classification of occupations into the creative class (and corresponding employment data) provided by the ERS, which excludes from Florida's original measure "many occupations with low creativity requirements and those involved primarily in services to the residential community". Included occupations are, for instance, Computer and Mathematical, Architecture and Engineering, or Life, Physical, and Social Science occupations, but also Management occupations. The ERS provides county-level data for 1990, 2000, and the average over the 2007-11 ACS rounds, which we aggregate to the commuting zone level. On average, 17% of workers are employed in the creative class during our sample period, with a median of 16%. The results (see column 5 of Table 6) show that local booms lead to a statistically significant increase in the number of creative class workers. This suggests that during booms, more people – measured in absolute terms – are employed in occupations linked to local innovation processes. For instance, local booms might raise the supply of qualified lab assistants, raising the productivity of incumbent inventors. This productivity-centered

[31] We aggregate county-level data to the commuting zone level. The census data are made available by the ERS, and the ACS data are obtained from `data.census.gov`. The ACS has insufficient coverage for a reliable county estimate in any given year, but the five-year averages are representative and can be compared to the census data (see e.g. Weber, 2014). Regarding the used specification, we use total endowment for the entire sample period rather than only from 2001 onwards. We do so to make the results comparable to the results on creative class workers (see the next paragraph), for which we only have one observation per commuting zone post-2000, prohibiting us to update endowment in 2001 and separately demean the data over 1969-2000 and 2001-2012. Column 5 of Table 9 shows that our baseline results are robust to using total endowment throughout our entire sample period.

interpretation is further supported by our finding that, in addition to the *number* of non-oil and gas patents, the number of patents *per capita* also rises during booms (see column 6 of Table 6).[32]

4.4 Geographic spillovers and absolute versus relative effects

In this section, we estimate geographic spillovers from booming commuting zones to other areas. With some additional assumptions, which we empirically test, this exercise will also enable us to estimate absolute effects of local economic booms on patenting, as opposed to the relative effects estimated via Equation (1).

Negative geographic spillovers can occur as the inflow of people into booming commuting zones mirrors emigration from other areas; positive spillovers could for instance arise through increased business demand originating from the booming commuting zone. The nature of migration spillovers depends on migration patterns: the more a boom induces in-migration from all over the country to a fairly balanced degree, the less we expect (substantial) negative spillovers in any non-booming areas, and vice versa. We therefore precede our spillover analysis by examining county-level migration data from the *Internal Revenue Service* (IRS). These data allow us to understand the impact of our boom measure (now defined at the county level) on in-migration from counties in the same state versus from other states or abroad. The results (see Online Appendix Section OA1.7) reveal that out of ten migrants moving into a booming county, eight come from within the same state. Although this finding does not directly inform us about migration by individuals involved in innovation, the evidence that the great majority of migrants move from nearby rather than more distant areas suggests that local booms may harm innovation in surrounding regions.

We proceed by directly testing for geographic spillovers. In terms of outcome variables, we focus on adult population (paralleling the specification in column 4 of Table 6) and non-oil and gas patenting (paralleling the specification in column 2 of Table 2). Based on the migration

[32] The dependent variable in column 6 is the number of non-oil and gas patents divided by working-age population (age 15-64). We use OLS rather than Poisson in this regression. The results show that in the commuting zone with endowment equal to one standard deviation, an increase in national oil and gas employment by 100 log points leads to one more patent in non-oil and gas technology per 100,000 residents of working age.

results, we first examine how outcomes in commuting zone c are affected by booms within the same state. Thereafter, we test for boom spillovers on closer areas, by studying the effects of booms (i) within various distance radii of up to 400 miles and (ii) in neighboring commuting zones.

Spillovers from booms in the same state

We first test the impact of same-state booms on commuting zone c's outcomes, controlling for c's "own boom" measure as per Equation (1). Same-state booms are defined as the interaction of national oil and gas employment and oil and gas endowment per square mile when considering all commuting zones in the same state, except for the "home zone" c, as one geographic unit.[33] Columns 1 and 2 of Table 7 show that both for population and patenting, the coefficients on the same-state boom interaction are not statistically significant. This could be because positive and negative spillovers cancel each other out, or because geographic spillovers occur in closer proximity than within an entire state.

Spillovers from booms at varying distances

Next, we test whether and how booms that occur within certain distance intervals affect outcomes in commuting zone c. This way of testing for geographic spillovers is analogous to the approach of Miguel and Kremer (2004), Allcott and Keniston (2018) and James and Smith (2020). We define four "doughnuts" around commuting zone c. The most proximate doughnut consists of all commuting zones whose centroid is less than 100 miles away from the centroid of c, and the remaining three doughnuts include commuting zones whose centroid is 100-200 miles, 200-300 miles, and 300-400 miles away, respectively. For each doughnut, we sum the endowment of all included commuting zones and divide by their joint total area in square miles. The specification looks as follows:

[33] To ease the interpretation of results, we scale this endowment by the standard deviation of conventional endowment at the commuting zone level (which equals about five million dollars per square mile), same as in our baseline analysis. Note that we also scale all other endowment variables that we use in Section 4.4 by the standard deviation of conventional endowment. Further note that since the same-state boom interaction is collinear with state times three-year period fixed effects, we replace the latter by census region times three-year period fixed effects. Except for this modification and the extra interaction term, the specification is identical to Equation (1).

$$
\begin{aligned}
Y_{c,\tau} = {} & \beta_1[OGendow_{c,T} \times ln(NatOGempl_\tau)] + \gamma_1[OGendow0to100_{c,T} \times ln(NatOGempl_\tau)] \\
& + \gamma_2[OGendow100to200_{c,T} \times ln(NatOGempl_\tau)] + \gamma_3[OGendow200to300_{c,T} \times ln(NatOGempl_\tau)] \\
& + \gamma_4[OGendow300to400_{c,T} \times ln(NatOGempl_\tau)] + \delta_{c,T} + \delta_{c,T} * \tau + \gamma_{s,\tau} + \epsilon_{c,\tau} \qquad (4)
\end{aligned}
$$

γ_1 to γ_4 capture the spillovers from booms occurring in the four different distance intervals. Clarke (2017) shows that if there are no spillovers beyond the maximum specified radius, then β_1 in Equation (4) reflects an absolute effect net of spillovers, as opposed to the relative effect estimated in column 2 of Table 2. Clarke suggests a data-driven, iterative approach for selecting the maximum radius, whereby an additional doughnut should be included in the specification only if its coefficient is statistically significant. However, we directly include doughnuts up to a 400 mile radius rather than applying Clarke's procedure, because positive and negative spillovers might cancel out within intermediate distance intervals and thereby produce insignificant coefficients. That said, we do apply Clarke's procedure in Online Appendix Table OA8 to show that extending Equation (4) to even larger distances is not necessary.

Estimating Equation (4) reveals that both population and patenting activity respond positively to booms in other commuting zones that are less than 100 miles away (see columns 3-4 of Table 7). These results indicate positive spillover effects from very proximate booms. In contrast, commuting zones that lie beyond 100 miles from a boom do not experience positive spillovers; and for commuting zones that are 300-400 miles away, the coefficients are in fact negative and – for population – statistically significant.[34] These results are consistent with the hypothesis that while positive business demand spillovers weaken with distance, areas within 400 miles are still close enough to supply workers to the booming commuting zone (and/or to areas more proximate to the boom). The coefficient on a commuting zone's "own boom" is statistically significant and indicates an absolute boom effect (that is, after accounting for geographic spillovers) that is similar in size to the relative effect in column 2 of Table 2.

Spillovers from booms in first- and second-order neighboring commuting zones

[34] This evidence broadly aligns with previous findings on geographic spillovers of oil and gas-driven economic booms on outcomes such as employment and wages (Feyrer et al., 2017; Allcott and Keniston, 2018; James and Smith, 2020).

In columns 5 and 6 of Table 7, we think of proximity in terms of sharing a common border rather than in terms of distance, similar to Weber (2014) or Weinstein et al. (2018). We do so to overcome two disadvantages of our distance approach in columns 3 and 4: (i) for some commuting zones, there are no other zones within 100 miles (centroid to centroid); and relatedly, (ii) any choice of doughnut size is somewhat arbitrary. The results show that neighbors of booming commuting zones experience an increase in population and patenting, while "neighbors of neighbors" actually face a statistically significant *decrease* in patenting.[35] While neighbors of neighbors' population appears unaffected, these commuting zones might still experience out-migration of individuals in innovation-related occupations, although unreported regressions do not show a statistically significant decline in college graduates or creative class workers.

In sum, the spillover results indicate that booms are beneficial for innovation in very proximate areas, which could be interpreted as evidence of wider agglomeration effects during a boom. In contrast, innovation appears to suffer in areas of intermediate distance, which appear close enough to lose workers to the extended boom area but too distant to substantially benefit from positive spillovers such as increased business demand.

4.5 Different sectors, different opportunity costs of innovation

Thus far we have presented evidence of procyclical local innovation and identified agglomeration economies as the main mechanism. Does that mean that theories suggesting countercyclical innovation through variation in opportunity costs of innovation are empirically irrelevant? Our setting allows us to directly test this hypothesis. Specifically, in this section we examine heterogeneous effects across industries that differ in the extent to which their opportunity costs of innovation rise during local booms. We start by discussing why oil and gas patenting declines during oil and gas-driven local booms (see Section 4.1). Figures A3-A5 provide suggestive evidence that times of high national oil and gas employment are times of high opportunity costs of innovation in the oil and gas sector. Figure A3 shows that national oil and gas employment

[35] Neighbors of neighbors of commuting zone c are defined as commuting zones that are contiguous to commuting zones which share a border with c (but are not contiguous to c themselves). Across all commuting zones, the average distance between the centroid of c and the centroid of such "second-order neighbors" is 135 miles, while the maximum distance is 444 miles. See for example Acemoglu et al. (2015) or Ladino et al. (2021) for other papers studying neighbors of neighbors in the context of spillovers.

correlates positively with the oil price during 1969-2012, and Figure A5 shows that higher oil prices have historically implied higher profits in the petroleum sector. Figure A4 shows that national oil and gas employment also closely co-moves with the natural gas price, until the early 2000s. Thereafter, the increase in American gas supply due to the fracking revolution depressed gas prices, but the reduced cost of gas production – which had initially triggered the record-level oil and gas employment in the early 2010s – led to rising profits. Times of high oil and gas employment have thus generally been times of high profits in the oil and gas sector over 1969-2012, leading local producers to prioritize extraction and sales – as evidenced by a rise in local oil and gas activity during boom times (see Online Appendix Table OA4) – rather than innovation. In this context, the results of Table 2 provide strong empirical evidence supporting the "opportunity cost hypothesis": the oil and gas sector, facing the most apparent rise in the opportunity cost of innovation, reduces patenting during booms, while other sectors, which are less central to the boom, increase patenting.

In the remainder of this section, we provide an additional test of the opportunity cost hypothesis. Specifically, we study heterogeneous boom effects on non-oil and gas patenting across highly- and lowly traded goods producers. The underlying motivation is that these sectors vary in their exposure to higher local demand during the boom, resulting in heterogeneous changes to these sectors' opportunity cost of innovation. Highly traded goods producers mainly sell outside of the local market, which implies that the demand for their products and thus their opportunity cost of innovation is largely unchanged during local booms. In contrast, lowly traded goods producers primarily serve the local market and can raise both prices and sales upon higher local demand, as they face lower import competition. This implies a rise in their opportunity costs of innovation during local booms. We bring this reasoning to the data by first classifying a 4-digit SIC manufacturing sector as relatively highly traded if it has a below-median distance elasticity to trade. The latter equals the percentage change in trade volume as distance increases by one percent, as calculated by Holmes and Stevens (2014) using data on industries' average shipment distance. Ready-mixed concrete and ice have the highest distance elasticity, while 14 industries including watches, x-ray equipment, and aircraft parts have the lowest. Having classified manufacturing industries into highly- and lowly traded, we map technology classes to industries as described in Section 4.3.1, and allocate patents in manufacturing industries into the two tradedness categories. The results (see Table 8, column 2) show that lowly traded goods producers do not significantly raise patenting in boom times. This is consistent with two opposing forces offsetting each other: a negative effect on innovation due to larger opportunity costs, but a positive effect which most likely operates through agglomeration economies (see Section 4.3). In contrast, and consistent with theory, highly traded goods producers significantly raise patenting in boom times (see column 3; column 1 reports the average effect across all manufacturing industries).

As a robustness check, we divide industries into tradedness terciles. If the results in columns 2-3 can indeed be explained by differences in opportunity costs arising from varying exposure to local demand, then we should see the weakest impact on the least traded tercile, stronger effects on the intermediate, and the strongest impact on the most traded tercile. This is exactly what we observe in columns 4-6 of Table 8.

Finally, we note that upward wage pressure from the booming oil and gas sector might lead highly traded goods producers – which can hardly respond to higher local wages by raising prices – to shed some production workers. However, Allcott and Keniston (2018) only find weak evidence for this. More importantly, such crowding-out effects are unlikely to occur in the labor market for workers involved in the innovation process, and thereby negatively affect highly traded industries' patenting activity. The reason is that oil and gas patenting decreases during booms, indicating that the sector does not raise its demand for inventors.

4.6 Robustness Checks

In Table 9 we carry out several robustness checks. Panel *I* reports the results on non-oil and gas patenting, while Panel *II* reports the results on the same robustness checks for oil and gas patenting. Column 1 repeats our baseline results (see columns 2-3 of Table 2), for comparison. In column 2, we add interactions of national oil and gas employment with commuting zone-level initial personal income per capita, human capital, and rural-urban dummy variables, respectively – although these variables do not robustly correlate with oil and gas endowment (see Table A2). We do so to address recent advances in the shift-share literature (Goldsmith-Pinkham et al., 2020) showing that a shift-share setup provides a valid identification strategy if the shares are exogenous conditional on controls. In column 3, we control for local coal booms by adding an interaction of local coal endowment in 1960 with national coal employment. In column 4, we use conventional endowment as our endowment measure for the entire sample period, while in column 5, we use total endowment (including fracking reserves) throughout 1969-2012. In column 6, we use the oil price rather than national oil and gas employment as shift variable. The results show that our baseline findings are robust to all of these modifications.[36]

5 Conclusion

This paper studies the responsiveness of innovation to changes in economic activity at the local geographic level. Exploiting nationwide oil and gas shocks as an exogenous source of local booms and busts in oil and gas-endowed commuting zones, we provide novel evidence that local economic booms lead to an increase in local patenting. The results are strongest in non-metropolitan but urban areas, showing that innovation responds to increased economic activity also in regions far from the current hotspots of American innovation. However, controlling for urbanization, boom effects are larger in areas that have historically patented more, indicating that local areas require a pre-existing innovation infrastructure to fully exploit the benefits of a boom. This evidence on heterogeneous effects informs the recent academic and policy debate on the spatial distribution of innovation across the United States. For example, while Gruber and Johnson (2019) emphasize the importance of a skilled local workforce for promoting local innovation, our results suggest that prior patenting experience is more important than human capital availability for translating local economic growth into increased innovative output.

We test multiple mechanisms that may explain the rise in local patenting, and show evidence consistent with agglomeration economies that enhance the productivity of incumbent inventors. Our geographic spillover results, however, indicate that positive local effects come at the expense of reduced innovation in areas at intermediate distance from the boom. These findings support the intuition that smart people are in limited supply, and more talent would need to be trained or attracted from abroad to substantially raise aggregate innovation (van Reenen, 2022).

Does our result of procyclical local innovation imply that higher opportunity costs of R&D

[36] Note that the coefficient in column 6 is smaller in magnitude compared to column 1, arguably reflecting that doubling the oil price is a smaller shock than doubling national oil and gas employment, as suggested by Figure A3.

and experimentation in boom times do not dampen innovation in such periods? We provide empirical evidence against this hypothesis, by showing that sectors experiencing the largest increases in the opportunity cost of innovation during local booms do not raise patenting or even reduce it. These novel findings show that opportunity costs of innovation are empirically relevant but are overall outweighed by pro-innovation factors in boom times, which helps to reconcile a long-standing puzzle in the literature on the cylicality of innovation.

Tables and Figures

Table 1: The impact of oil&gas booms on local economic activity (=“conceptual first stage”)

Dependent Variable → ln()...	Population	Employment	Earnings per worker	Personal income per capita	GDP per capita	Local government revenue
	(1)	(2)	(3)	(4)	(5)	(6)
ln(National oil&gas employment) × endowment	0.019***	0.037***	0.022***	0.018***	0.046*	0.062***
	(0.006)	(0.007)	(0.003)	(0.002)	(0.023)	(0.011)
Observations	11,355	11,355	11,355	11,355	3,028	6,800
Sample period	69-12	69-12	69-12	69-12	01-12	72-12

Notes: In this table we evaluate the quality of our local economic boom proxy, by analyzing the impact of oil and gas booms on various measures of economic activity at the commuting zone level. In all columns we estimate Equation (1), or slightly modified versions of it (see below), using OLS. *endowment* equals initial oil and gas reserves per square mile at the commuting zone level and is scaled by the standard deviation of this variable across all commuting zones. From 2001 onwards, *endowment* is updated to include shale oil and gas (fracking) reserves. For each dependent variable, we use the average realization over the three-year period (see Section 2), except for *Local government revenue*, which we measure in the years for which data are available. *Personal income* includes wages and salaries and other income sources such as rental or interest income; see Online Appendix Section OA3 for details. Local-level GDP data are only available from 2001 onwards. *Local government revenue* equals total revenue collected by the commuting zone's county governments from its own sources, and excludes state and federal transfers. The variable is available at five-yearly intervals: we use data from the years 1972, 1977, 1982, 1987, 1992, 1997, 2002, 2007, and 2012. Given this frequency pattern, we simply drop all other years and evaluate also the key interaction term based on the years stated above. All columns include state times three-year period fixed effects. Columns 1-4 and 6 further include commuting zone times century (1969-2000 versus 2001-2012) fixed effects and control for commuting zone-specific patenting trends over the pre-fracking period (1969-2000) and the fracking period (2001-2012). Given that GDP data are only available from 2001 onwards, in column 5 we instead include commuting zone fixed effects and one linear time trend per commuting zone. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table 2: Local economic booms and patenting activity

Dependent Variable →	# Patents	# Non-oil&gas patents	# Oil&gas patents	# Patents
Unit of observation →	Commuting zone			Firm
	(1)	(2)	(3)	(4)
ln(National oil&gas employment) × endowment	0.083***	0.089***	-0.085***	0.020***
	(0.028)	(0.034)	(0.033)	(0.005)
Observations	11,112	11,112	7,767	920,880
Sample period	69-12	69-12	69-12	76-06

Notes: In this table we analyze the impact of oil and gas booms – which serve as a proxy for local economic booms (see Table 1) – on patenting activity. *endowment* equals initial oil and gas reserves per square mile and is scaled by the standard deviation of this variable across all commuting zones. From 2001 onwards, *endowment* is updated to include shale oil and gas (fracking) reserves. In columns 1-3 we estimate Equation (1); the unit of observation is a commuting zone. In column 1 we consider patents in all technology classes; in column 2 we exclude patents in oil and gas technologies; and in column 3 we focus exclusively on oil and gas patents. In column 4 we estimate Equation (3); the unit of observation is a firm (which patents in more than one commuting zone). We build the sample of column 4 as described in Section 4.1. The regression serves to test whether multi-location firms increase patenting in booming commuting zones relative to other commuting zones where they are active. In all columns, we aggregate the number of patents over a period of three years, starting with the period 1969-1971 and ending with the period 2010-2012 (the last period before updating endowment in 2001 consists of only two years, 1999-2000). In all columns we carry out Poisson pseudomaximum likelihood regressions. All regressions include commuting zone times century (1969-2000 versus 2001-2012) fixed effects and state times three-year period fixed effects, and control for commuting zone-specific patenting trends over the pre-fracking period (1969-2000) and the fracking period (2001-2012). In addition, column 4 includes firm times three-year period fixed effects. The sample period in column 4 starts with the three-year period 1975-1977 and ends with the period 2004-2006 due to data availability in the crosswalk from patents to Compustat firms. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table 3: Heterogeneous effects

Dependent Variable →	# Non-oil&gas patents						
Sample of commuting zones →	All	Low patenting activity	All				
Unit of observation →	Commuting zone					C-zone – tech class	Comm-zone
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln(National oil&gas employment) × endowment	0.089***	0.070***	0.072**	0.072**	0.110**	0.050***	0.084***
	(0.034)	(0.010)	(0.028)	(0.028)	(0.055)	(0.017)	(0.030)
ln(Nat. O&G empl.) × endow. × Non-metropolitan			0.070***				
			(0.027)				
ln(Nat. O&G empl.) × endow. × Urban non-metro				0.070***	0.100***		
				(0.027)	(0.034)		
ln(Nat. O&G empl.) × endow. × Rural non-metro				0.029	0.063		
				(0.136)	(0.181)		
ln(Nat. O&G empl.) × endow. × Initial patenting intensity					0.099***		
					(0.036)		
ln(Nat. O&G empl.) × endow. × Human capital					-0.008		
					(0.034)		
ln(Nat. O&G empl.) × endow. × College density					0.010		
					(0.024)		
ln(N. O&G e.) × end. × IPC2's historical share in c-zone pat.						0.170***	
						(0.045)	
ln(Nat. O&G empl.) × endow. × Bust period 1984-2000							0.002
							(0.002)
Observations	11,112	5,387	11,112	11,112	11,112	180,910	11,112
Sample period	69-12	69-12	69-12	69-12	69-12	69-12	69-12
Marginal effect on non-metropolitan commuting zones			0.142***				
			(0.048)				
Marginal effect on urban non-metropolitan commuting zones				0.142***			
				(0.048)			
Marginal effect on rural non-metropolitan commuting zones				0.102			
				(0.142)			

Notes: In this table we analyze heterogeneous effects across different types of commuting zones (columns 2–5), across technology classes depending on their historical relevance within a commuting zone (column 6), and across booms and busts (column 7). In all columns we focus on non-oil and gas patenting. In columns 1-2 we estimate Equation (1). Column 1 repeats the results of column 2 of Table 2, while in column 2 we restrict the sample to commuting zones with a below-median patents per capita ratio (we compute this ratio as total patents in 1969-2012 divided by the sum of all annual population counts of the commuting zone over 1969-2012; the median equals six patents per 100,000 residents per year). In columns 3-5 we depart from Equation (1) by adding one or more interactions of the boom interaction with a commuting zone-specific variable. In column 6 we estimate an alternative specification at the commuting zone-technology class (2-digit) level, as described in Section 4.2. In column 7 we estimate Equation (1) including an additional interaction term, where *Bust period 1984-2000* is a dummy variable that equals one for all three-year periods within the time frame 1984-2000. *Non-metropolitan*, *Urban non-metro* and *Rural non-metro* are defined based on the Rural-Urban Continuum Codes published by the ERS; see Section 4.2 for details. *Initial patenting intensity* equals the total number of patents over 1960-69 divided by population in 1969, scaled by the variable's standard deviation. *Human capital* equals the share of the commuting zone's population aged 25+ with at least one year of college education, as of 1970, scaled by the variable's standard deviation. *College density* equals the fraction of residents employed in "colleges, universities, and professional schools" in the commuting zone, as of 2018, scaled by the variable's standard deviation. *IPC2's historical share in c-zone patenting* equals the share of the 2-digit technology class in total patenting of the commuting zone over 1960-69 (to compute this share, we remove oil and gas patents from both the numerator and the denominator). In column 5, before performing the regressions, we demean the commuting zone characteristics that feature in the triple interaction terms. This allows us to interpret the boom coefficient in the first row as the effect in the average commuting zone across the sources of heterogeneity we test for in column 5. In all columns we estimate Poisson pseudomaximum likelihood regressions. In columns 1-5 and 7, we include commuting zone times century (1969-2000 versus 2001-2012) fixed effects and state times three-year period fixed effects, and control for commuting zone-specific patenting trends over the pre-fracking period (1969-2000) and the fracking period (2001-2012). In column 6 we replace commuting zone times century fixed effects with 2-digit IPC class times commuting zone times century fixed effects, and further include 2-digit IPC class times three-year period fixed effects. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table 4: Mechanisms I: finance, input-output linkages

Dependent Variable: # Non-oil&gas patents by... →	All industries					Up-stream industries	Non-linked industries
Unit of observation →	Commuting zone – industry		Commuting zone – technology class – firm type		Compustat firm	Commuting zone	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln(National oil&gas employment) × endowment	0.065**	0.064**	0.057**	0.058**	0.032***	0.067***	0.090***
	(0.028)	(0.026)	(0.024)	(0.023)	(0.007)	(0.024)	(0.033)
ln(N. O&G e.) × end. × Ext. finance dependence		0.001					
		(0.003)					
ln(N. O&G e.) × end. × Pat. by Compustat firms				-0.004			
				(0.003)			
Observations	221,684	221,684	254,752	254,752	470,887	11,112	11,112
Sample period	69-12	69-12	75-06	75-06	75-06	69-12	69-12
Marginal effect on patenting by Compustat firms				0.055**			
				(0.026)			

Notes: In this table we test various mechanisms that might explain the baseline results in column 2 of Table 2. In column 2 we test whether 2-digit SIC industries that depend more on external finance (Rajan and Zingales, 1998) are differently affected by local booms in terms of their patenting behavior. In column 4 we test whether within a given 2-digit IPC technology class, patenting by firms that are included in Compustat (all of which are publicly listed and therefore, on average, less financially constrained) is affected differently. We do so by including two observations per commuting zone – three-year period – 2-digit technology class: the number of patents in this cell that are owned by Compustat firms (for this observation, the dummy variable *Pat. by Compustat firms* equals one), and the number of non-Compustat patents (dummy=0). In columns 1 and 3 we estimate the specifications of columns 2 and 4, respectively, with the difference that we only include our boom interaction. In column 5, we estimate the firm-level Equation (3) using the sample of Compustat firms, and build the sample in the same way as described in Section 4.1, just applied to non-oil and gas rather than total patents. We thereby test whether Compustat firms increase non-oil and gas patenting in booming commuting zones relative to other commuting zones where they are active. The sample period in columns 3-5 starts with the three-year period 1975-1977 and ends with the period 2004-2006 due to data availability in the crosswalk from patents to Compustat firms. In column 6 we estimate Equation (1) using, as dependent variable, the number of patents in industries that are upstream (and not downstream) to the oil and gas sector. In column 7 we do the same for patents in industries that are neither upstream nor downstream. In all columns we employ Poisson pseudomaximum likelihood regressions. The regressions in columns 1-2 contain commuting zone times century times 2-digit SIC industry fixed effects, which implies 20 industry dummies per commuting zone for the pre-fracking period (1969-2000) and 20 industry dummies per commuting zone for the fracking period (2001-2012); 2-digit SIC industry times three-year period fixed effects; and state times three-year period fixed effects. Furthermore, we control for differential commuting zone-specific patenting trends across the pre-fracking period (1969-2000) and the fracking period (2001-2012). In columns 3-4 we use the same fixed effects structure, but with technology class rather than industry fixed effects. In columns 6-7, we use the same fixed effects as in Table 2. The same applies to column 5, except for adding firm times three-year period fixed effects. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table 5: Mechanisms II: Agglomeration (Part I: patenting by inventor type)

Dependent Variable: # Non-oil&gas patents by... $\rightarrow$	All inventors	Incumbent inventors	In-moving inventors	New inventors
	(1)	(2)	(3)	(4)
ln(National oil&gas employment) $\times$ endowment	0.051*	0.071***	0.012	0.023
	(0.030)	(0.025)	(0.049)	(0.032)
Observations	27,404	25,541	23,337	27,317
Sample period	76-12	76-12	76-12	76-12

Notes: In this table we investigate whether the rise in non-oil and gas patenting during local booms is driven by a rise in patents by (i) incumbent inventors (see column 2), (ii) inventors that move into the booming commuting zone (column 3), or (iii) new inventors (column 4). We classify a patent in year t as an incumbent inventor's patent if the inventor's physical address indicated on his or her most recent pre-t patent is in the same commuting zone as the address on the year t patent. Similarly, a patent in year t is classified as an in-mover's patent if the address on the inventor's most recent pre-t patent belongs to a different commuting zone. Patents by new inventors, in year t, are those by inventors that appear in our entire data set for the first time in year t. Our sample period is 1976-2012 due to data availability. We perform all regressions at the annual level, for reasons described in Section 4.3.3. In all columns we estimate Poisson pseudomaximum likelihood regressions. All regressions include commuting zone times century (1969-2000 versus 2001-2012) fixed effects and state times three-year period fixed effects, and control for commuting zone-specific patenting trends over the pre-fracking period (1969-2000) and the fracking period (2001-2012). Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table 6: Mechanisms II: Agglomeration (continued)

Dependent Variable →	Adult Population	Adult Population, College	Adult Population, Non-College	Adult Population	Creative Class Workers	#Non-OG Patents / Work-age Population
Data Frequency →	1990, 2000, 2006-10, 2011-15			Annual, '69-'12	'90, '00, '07-'11	3y periods, '69-'12
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Nat. oil&gas empl.) × endow.	0.013***	0.016***	0.013***	0.013***	0.018**	0.010***
	(0.004)	(0.006)	(0.004)	(0.003)	(0.008)	(0.003)
Observations	3,028	3,028	3,028	33,308	2,270	11,355
Sample period	90-15	90-15	90-15	69-12	90-11	69-12

Notes: In this table we study other dependent variables, which matter in the context of agglomeration: adult (age 25+) population (columns 1 and 4, which differ in terms of data frequency), college- versus non-college-educated adult population (columns 2-3), the number of "creative class workers" (column 5), and patents per capita (column 6). In columns 1-5, the dependent variable is the natural log of the indicated variable. In column 6, the dependent variable is the total number of non-oil and gas patents divided by the sum of all annual population counts (in terms of 1,000 residents) in the relevant three-year period. College-educated refers to a completed college degree. Data on educational attainment are from the 1990 census; the 2000 census; and the American Community Survey (ACS), from which we use the five-year averages over 2006-2010 (earlier data are unavailable) and 2011-2015. On the right-hand side, we evaluate national oil and gas employment in 1990, in 2000, as average over 2006-2010, and as average over 2011-2015. Data on *Creative Class Workers* are obtained from the Economic Research Service, which refined the original definition of Florida (2002) by excluding "many occupations with low creativity requirements and those involved primarily in services to the residential community". All regressions are estimated using OLS. In column 6 we include the same fixed effects as in Equation (1). In all columns except for column 6, we (i) use total endowment for the entire sample period rather than update endowment in 2001; (ii) include commuting zone fixed effects rather than commuting zone times century fixed effects; and (iii) only include one (linear) time trend per commuting zone. We make these modifications to Equation (1) since we only have one post-2000 data point per commuting zone for creative class workers, prohibiting us to apply the fixed effects structure of Equation (1). All regressions include state times three-year period fixed effects. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table 7: Geographic spillovers

Dependent Variable →	Adult Population	#Non-O&G Patents	Adult Population	#Non-O&G Patents	Adult Population	#Non-O&G Patents
Proximity Concept →	Same state		Distance doughnuts		Neighbors and 2^{nd} order neighbors	
Data Frequency →	annual	3-year periods	annual	3-year periods	annual	3-year periods
	(1)	(2)	(3)	(4)	(5)	(6)
ln(National oil&gas employment) × endowment	0.013***	0.082**	0.010***	0.069**	0.010**	0.053*
	(0.003)	(0.035)	(0.003)	(0.034)	(0.004)	(0.031)
ln(N. O&G e.) × end. of other c-zones in same state	0.007	-0.057				
	(0.006)	(0.076)				
ln(N. O&G e.) × end. of c-zones within 100 miles			0.008***	0.046**		
			(0.003)	(0.023)		
ln(N. O&G e.) × end. of c-zones 100-200 miles away			0.007	0.008		
			(0.006)	(0.064)		
ln(N. O&G e.) × end. of c-zones 200-300 miles away			-0.004	0.012		
			(0.011)	(0.097)		
ln(N. O&G e.) × end. of c-zones 300-400 miles away			-0.013*	-0.118		
			(0.007)	(0.102)		
ln(N. O&G e.) × neighbors' endowment					0.008	0.210***
					(0.005)	(0.032)
ln(N. O&G e.) × 2^{nd} order neighbors' endowment					0.003	-0.075***
					(0.005)	(0.026)
Observations	33,396	11,142	32,868	10,962	33,308	11,112
Sample period	69-12	69-12	69-12	69-12	69-12	69-12

Notes: In this table we test for geographic spillovers. *Endowment of other commuting zones in same state* equals total reserves of all commuting zones in the state except for commuting zone c, divided by the total area of the state excluding c. For the endowment variables in columns 3-6, we also compute total reserves and then divide by the corresponding total area, based on the indicated proximity concept. All endowment variables reported in this table are scaled by one standard deviation of conventional endowment at the commuting zone level (which equals around 5 million dollars per square miles). When computing the distance-specific endowment variables in columns 3-4, we consider the distance from a commuting zone's centroid to the centroid of c. *2^{nd} order neighbors* of commuting zone c are defined as commuting zones that are contiguous to commuting zones which share a border with c, but are not contiguous to c themselves. In columns 1, 3 and 5 we estimate OLS regressions (compare column 4 of Table 6), while in columns 2, 4 and 6 we employ Poisson pseudomaximum likelihood regressions (compare column 2 of Table 2). In columns 3-6 we include commuting zone times century (1969-2000 versus 2001-2012) fixed effects and state times three-year period fixed effects, and control for commuting zone-specific patenting trends over the pre-fracking period (1969-2000) and the fracking period (2001-2012). In columns 1-2 we use the same fixed effects structure except for replacing state times period fixed effects by census-region times period fixed effects. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table 8: Heterogeneity across different sectors

Dependent Variable →	# Non-oil&gas patents					
Included Manufacturing Industries →	All	Lowly Traded	Highly Traded	Least Traded Tercile	Intermediate Tercile	Most Traded Tercile
	(1)	(2)	(3)	(4)	(5)	(6)
ln(National oil&gas empl.) × endowment	0.090***	0.044	0.092***	0.018	0.068*	0.093***
	(0.034)	(0.037)	(0.028)	(0.030)	(0.035)	(0.029)
Observations	11,112	11,112	11,112	11,112	11,112	11,112
Sample period	69-12	69-12	69-12	69-12	69-12	69-12

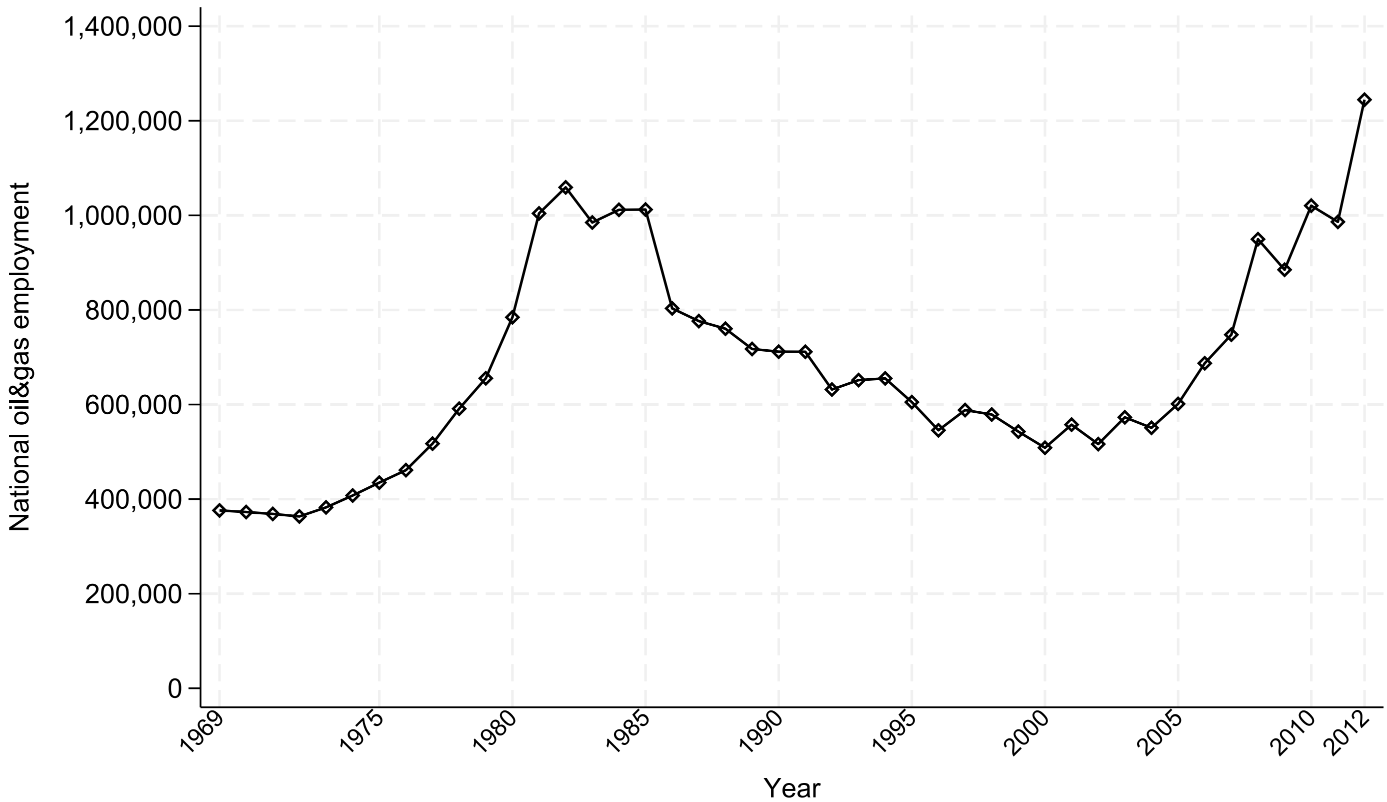
Notes: In this table we estimate Equation (1) on different subsamples, distinguishing patents in manufacturing sectors producing highly- versus lowly traded goods. Manufacturing industries' tradedness is measured by their distance elasticity to trade as calculated by Holmes and Stevens (2014). In column 3 we define highly traded industries as those with a below-median distance elasticity, across all 4-digit SIC manufacturing sectors. Conversely, lowly traded industries (see column 2) are those with the median or an above-median distance elasticity. In columns 4-6 we distinguish patents by sectors in the first, second, and third tercile of tradedness, respectively. In column 1, the dependent variable is the (non-oil and gas) patent count across all industries of the overall manufacturing sector. In all columns we estimate Poisson pseudomaximum likelihood regressions. All regressions include commuting zone times century (1969-2000 versus 2001-2012) fixed effects and state times three-year period fixed effects, and control for commuting zone-specific patenting trends over the pre-fracking period (1969-2000) and the fracking period (2001-2012). Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table 9: Main robustness checks

Nature of Robustness Check →	Base-line	Control for: X_c × O&G Empl.	Control for coal booms	Convent'l endowm. in all years	Total endowm. in all years	Shift variable = Oil price
Panel I: Dependent Variable →	# Non-oil&gas patents					
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Nat. O&G empl.) × endow.	0.089***	0.082**	0.088**	0.112**	0.089**	
	(0.034)	(0.037)	(0.035)	(0.055)	(0.045)	
ln(Oil price) × endowment						0.030***
						(0.006)
Observations	11,112	11,112	11,112	11,134	11,134	11,112
Panel II: Dependent Variable →	# Oil&gas patents					
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Nat. O&G empl.) × endow.	-0.085***	-0.081**	-0.084***	-0.079**	-0.047*	
	(0.033)	(0.037)	(0.032)	(0.033)	(0.026)	
ln(Oil price) × endowment						-0.032**
						(0.016)
Observations	7,767	7,767	7,767	7,767	7,767	7,767

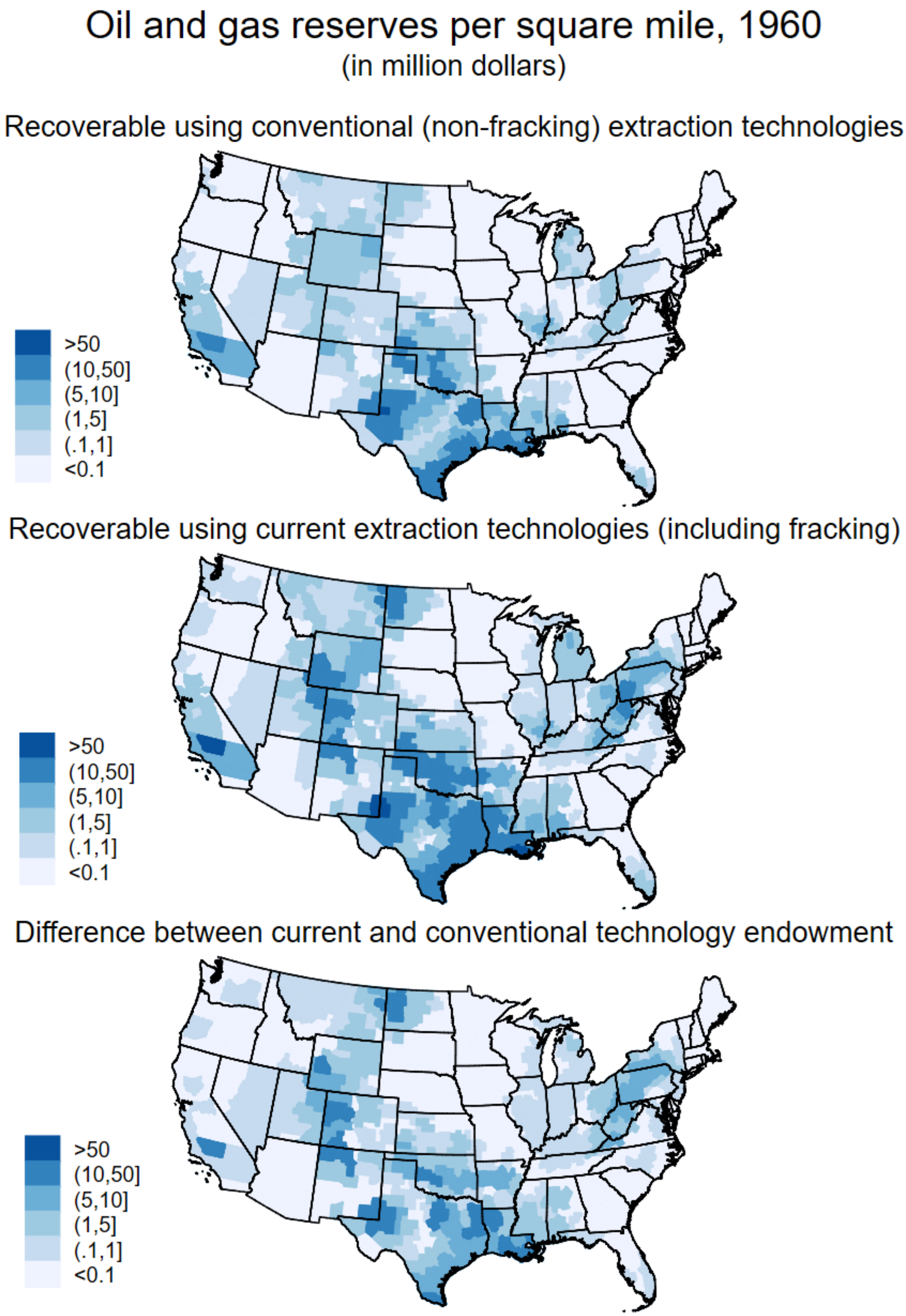
Notes: In this table we carry out several robustness checks on the results of column 2 of Table 2 (see Panel I) and the results of column 3 of Table 2 (see Panel II). In both panels, column 1 repeats the relevant baseline results. In column 2 we add interactions of national oil and gas employment with commuting zone-level personal income per capita (measured in 1969), residents' average education level (measured in 1970), and the dummy variables Urban non-metropolitan commuting zone and Rural non-metropolitan commuting zone (measured in 1974), respectively (compare Table A2). In column 3 we add an interaction of 1960 coal reserves at the commuting zone level and coal employment at the national level. In column 4 we use conventional endowment for the entire sample period, while in column 5 we use total endowment for the entire sample period. In column 6 we use the oil price rather than national oil and gas employment as shift variable in our key interaction term. In all columns we estimate Poisson pseudomaximum likelihood regressions. All regressions include commuting zone times century (1969-2000 versus 2001-2012) fixed effects (except for columns 4-5, which include commuting zone fixed effects) and state times three-year period fixed effects, and control for commuting zone-specific patenting trends over the pre-fracking period (1969-2000) and the fracking period (2001-2012). Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Figure 1: National oil and gas employment, 1969-2012



Notes: The data are sourced from the Bureau of Economic Analysis (BEA). See Online Appendix Section OA3 for details.

Figure 2: Oil and gas endowment at the commuting zone level



Notes: This figure provides information on commuting zone-level oil and gas endowment. The endowment variable in Panel I corresponds to Equation (2) with T=1995; the variable in Panel II corresponds to Equation (2) with T=2011; and the variable in Panel III is the difference in endowment across Panel II and Panel I.

Appendix

Table A1: Summary Statistics

	Mean	Median	Min	Max	sdev	N
Panel I: Variables at the commuting zone level over time						
# Patents (per three-year period)	216.0	12.2	0	29,128	998.8	11,385
# Oil and gas patents	5.50	0	0	3,106	54.3	11,385
# Non-oil and gas patents	210.5	12	0	29,090	983.1	11,385
...in Metropolitan CZs	1,643	671.7	4	29,090	2,704	1,095
...in Urban non-metropolitan CZs	78.2	16.3	0	5,964	260.2	7,440
...in Rural non-metropolitan CZs	5.40	2	0	89.1	9.09	2,850
...in Upstream industries	35.2	2.73	0	2,383	135.9	11,385
...in Non-linked industries	159.1	7.86	0	25,943	813.3	11,385
...by Compustat firms	107.2	2	0	19,867	559.4	8,349
...by non-Compustat firms	98.3	9.33	0	9,478	402.6	8,349
...by incumbent inventors	38.0	1	0	8,428	216.3	33,396
...by in-moving inventors	2.94	0	0	625.8	15.2	33,396
...by new inventors	25.4	2	0	2,812	109.8	33,396
...divided by working age pop. (in ’000 residents)	0.14	0.087	0	4.32	0.21	11,385
...by highly-traded industries (highest tercile)	135.0	5.32	0	24,786	739.5	11,385
...by medium-traded industries (medium tercile)	44.2	3.81	0	2,963	165.0	11,385
...by lowly-traded industries (lowest tercile)	24.2	1.98	0	1,313	89.3	11,385
# Forward citations per non-oil and gas patent	6.14	4.42	0	160.7	6.26	9,677
Patent generality	0.36	0.37	0	0.90	0.16	9,404
Population (in ’000)	334.3	87.2	0.43	18,046	957.7	11,385
Total employment (in ’000)	179.2	41.4	0.23	9,728	529.1	11,385
Earnings per worker (in ’000 real 2010 dollars)	31.8	31.3	15.8	80.5	5.92	11,385
Personal income per capita (in ’000 real 2010 dollars)	27.2	26.4	8.93	115.1	7.18	11,385
GDP per capita (in ’000 real 2010 dollars)	41,793	37,823	16,188	333,063	20,084	3,036
Local government revenue (in ’000 real 2010 dollars)	204,723	38,695	189.1	25,929,687	791,209	6,809
Adult population (25+ ; in ’000)	245.60	61.45	0.28	12,001	696.59	3,036
...with completed college degree (in ’000)	63.69	9.32	0.03	3,517	215.30	3,036
...without completed college degree (in ’000)	181.91	51.13	0.20	8,484	488.17	3,036
# Creative class workers (in ’000)	41.26	6.21	0.00	2,174	135.41	2,276
Panel II: Commuting zone level variables						
O&G reserves / Area, 1960 (excl. shale res.; million 2010 dollars)	1.48	0.03	0	56.96	4.58	759
O&G reserves / Area, 1960 (incl. shale res.; million 2010 dollars)	2.86	0.18	0	73.76	7.05	759
# Non-O&G patents 1960-69 / Population 1969 in ’000	4.43	1.99	0	63.58	7.06	759
Percentage of adult pop. with $\geq$1 year of college, 1970	16.72	16.08	6.19	47.60	5.39	759
Percentage of population employed in local colleges etc.	3.10	0.95	0	59.40	5.70	759

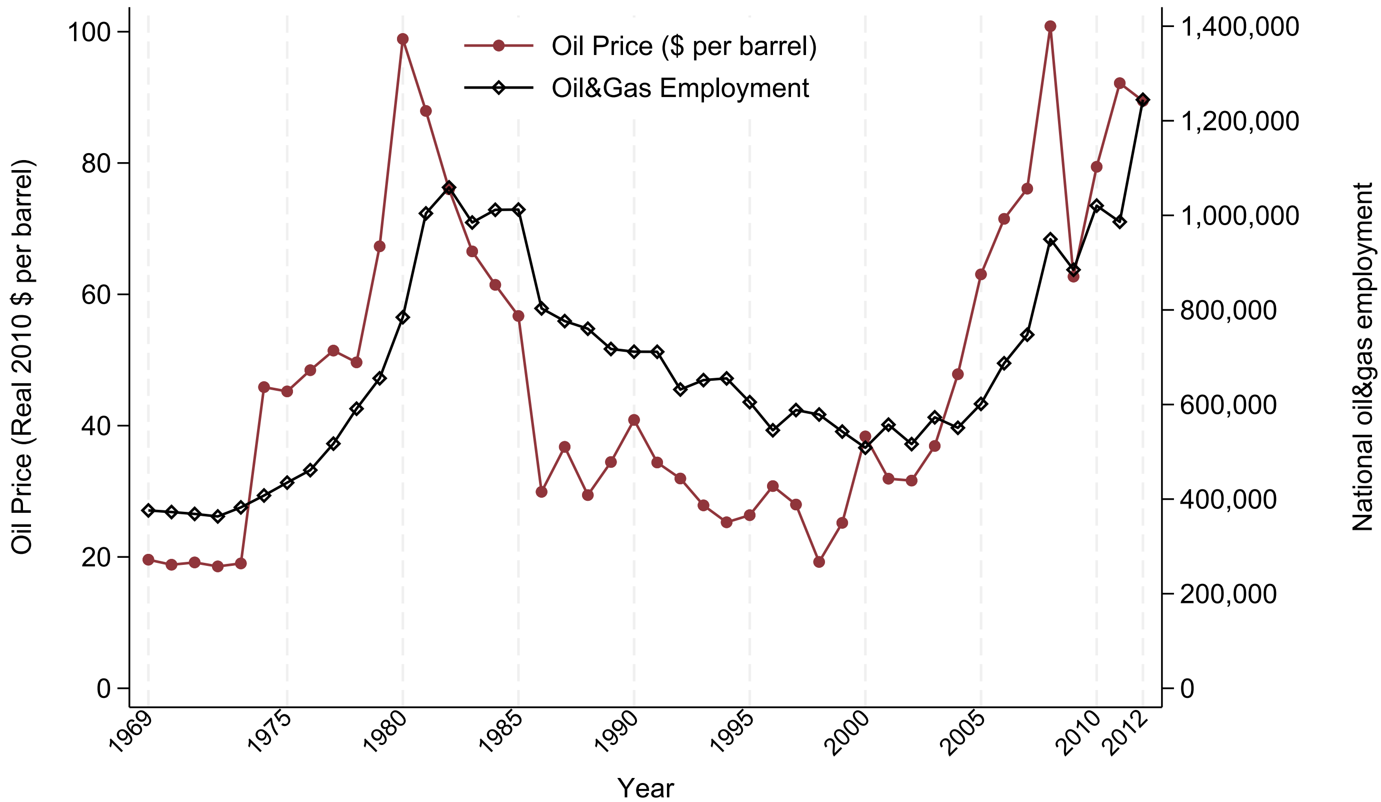
Notes: This table provides summary statistics on the variables used in our analysis. Values larger than 1,000 are rounded to the nearest integer. *Upstream industries* refers to industries that are upstream to the oil and gas sector, but not downstream. *Non-linked* means neither upstream nor downstream. The distinction into patents by Compustat versus non-Compustat firms is based on the period 1975-2006 due to data availability in the crosswalk from patents to Compustat firms. In Panel I, all statistics are as of three-year periods, except for (i) patents by incumbent, in-moving and new inventors (one-year periods, see Section 4.3.3) and (ii) the last five rows (annual data, used in the years in which the data are available, without aggregating to multi-year periods). Non-oil and gas patents per working age (15-64) population is defined as the total number of non-oil and gas patents divided by the sum of all annual working age population counts (in terms of 1,000 residents) in the relevant three-year period. The tradedness of industries is computed based on Holmes and Stevens (2014); see Section 4.5. *Local government revenue* equals the sum of own-source county government revenue across all counties in the commuting zone. Data are available five-yearly from the Census of Governments in the years 1972, 1977, 1982, 1987, 1992, 1997, 2002, 2007, and 2012. Data on adult population by educational attainment are based on data from the 1990 and the 2000 population census, and on the five-year averages over 2006-2010 and 2011-2015, respectively, from the American Community Survey (ACS). Data on the number of creative class workers are based on the 1990 and 2000 census and the 2007-2011 ACS average. In *local colleges etc.*, *colleges* refers to “colleges, universities, and professional schools”, and *local* refers to institutions located in the commuting zone. *# Forward citations per non-oil and gas patent* equals the total number of forward citations of all non-oil&gas patents over the following five years, divided by the total number of non-oil&gas patents. The concept of patent generality is explained in Online Appendix Section OA1.1.

Table A2: Correlation of oil and gas endowment with other local characteristics

Dependent Variable →	Conventional endowment		Total endowment	
	(1)	(2)	(3)	(4)
Urban non-metropolitan	-0.080		-0.047	
	(0.188)		(0.252)	
Rural non-metropolitan	-0.127		-0.073	
	(0.186)		(0.258)	
Personal income per capita 1969	0.104*	0.114**	0.115	0.127
	(0.060)	(0.054)	(0.088)	(0.078)
Human capital 1970	-0.047	-0.040	-0.020	-0.015
	(0.034)	(0.029)	(0.070)	(0.063)
Population 1969		-0.002		-0.016
		(0.019)		(0.023)
Observations	757	757	757	757

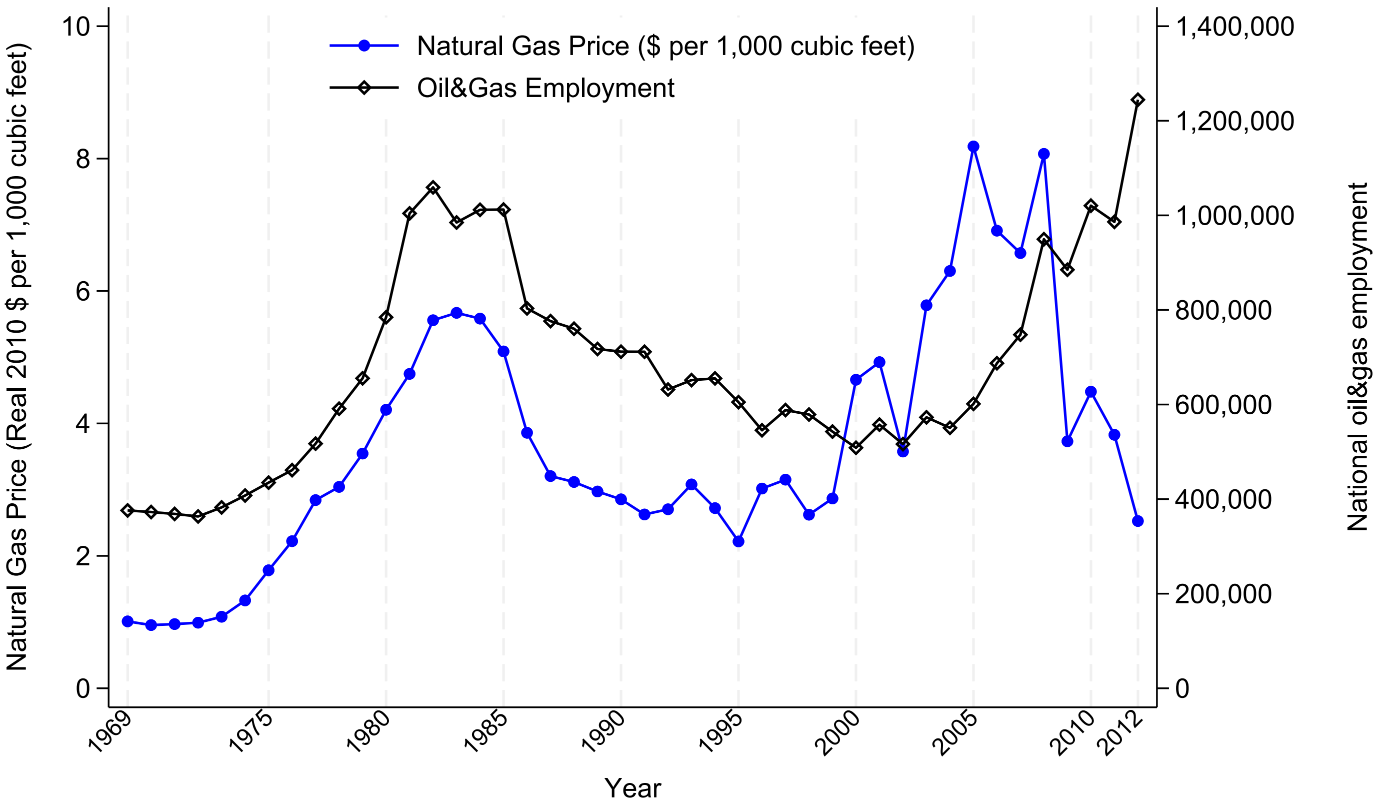
Notes: In this table we test the correlation between initial oil and gas endowment and other commuting zone characteristics. Endowment equals reserves divided by commuting zone size, as in our baseline specification. Human capital equals the share of the commuting zone's population aged 25+ with at least one year of college education, as of 1970. The dependent variables are scaled by the standard deviation of conventional endowment. The explanatory variables are scaled by their respective standard deviations. All specifications include state fixed effects. Standard errors in parentheses are clustered at the state level. Online Appendix Section OA3 provides data sources on the right-hand side variables used in this table. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Figure A3: Oil price and National oil and gas employment, 1969-2012



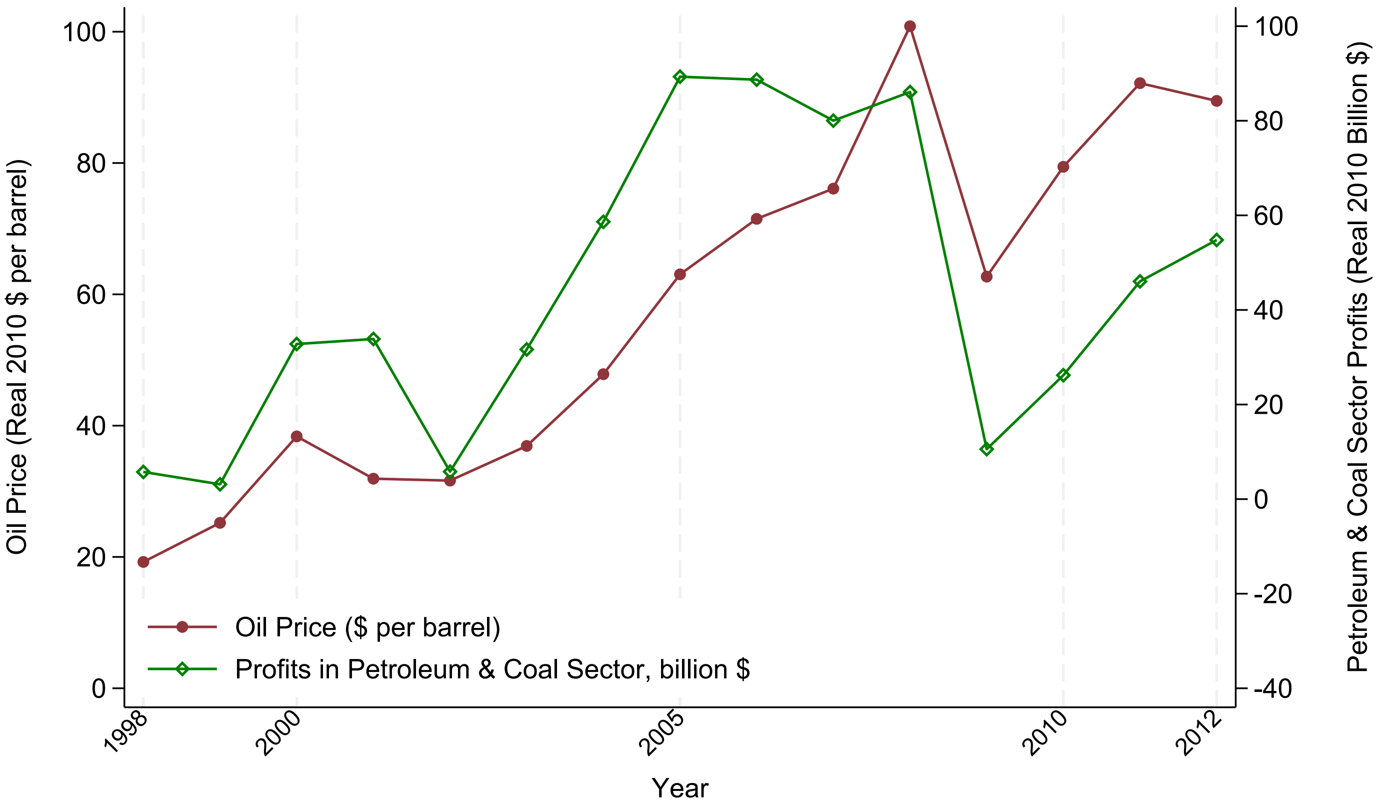
Notes: The oil price corresponds to the spot crude oil price of West Texas Intermediate (WTI). Oil prices are in 2010 real dollars.

Figure A4: Natural gas price and National oil and gas employment, 1969-2012



Notes: The natural gas price corresponds to the U.S. Natural Gas Wellhead Price. Prices are in 2010 real dollars.

Figure A5: Oil price and Profits in the petroleum & coal sector, 1998-2012



Notes: The oil price corresponds to the spot crude oil price of West Texas Intermediate (WTI). Profits in the oil and gas sector are not reported separately by the Bureau of Economic Analysis (BEA); therefore, we report profits in the most disaggregated sector for which data are available, which is "petroleum and coal products". The figure starts with the year 1998 since earlier data are unavailable. Both time series are expressed in 2010 real dollars.

References

Acemoglu, D., C. García-Jimeno, and J. A. Robinson (2015). State capacity and economic development: a network approach. *American Economic Review 105*(8), 2364–2409.

Aghion, P., P. Askenazy, N. Berman, G. Cette, and L. Eymard (2012). Credit constraints and the cyclicality of R&D investment: evidence from France. *Journal of the European Economic Association 10*(5), 1001–1024.

Aghion, P., A. Dechezleprêtre, D. Hemous, R. Martin, and J. Van Reenen (2016). Carbon taxes, path dependency, and directed technical change: evidence from the auto industry. *Journal of Political Economy 124*(1), 1–51.

Aghion, P. and G. Saint-Paul (1998). Virtues of bad times: interaction between productivity growth and economic fluctuations. *Macroeconomic Dynamics 2*(3), 322–344.

Akcigit, U., S. Baslandze, and S. Stantcheva (2016). Taxation and the international mobility of inventors. *American Economic Review 106*(10), 2930–2981.

Akcigit, U., J. Grigsby, and T. Nicholas (2017). The rise of American ingenuity: innovation and inventors of the golden age. NBER Working Paper 23047.

Akcigit, U., J. Grigsby, T. Nicholas, and S. Stantcheva (2022). Taxation and innovation in the twentieth century. *The Quarterly Journal of Economics 137*(1), 329–385.

Allcott, H. and D. Keniston (2018). Dutch disease or agglomeration? The local economic effects of natural resource booms in modern America. *The Review of Economic Studies 85*(2), 695–731.

Amore, M. D., C. Schneider, and A. Žaldokas (2013). Credit supply and corporate innovation. *Journal of Financial Economics 109*(3), 835–855.

Andrews, M. J. (2023). How do institutions of higher education affect local invention? Evidence from the establishment of US colleges. *American Economic Journal: Economic Policy 15*(2), 1–41.

Aragón, F. M. and J. P. Rud (2013). Natural resources and local communities: evidence from a Peruvian gold mine. *American Economic Journal: Economic Policy 5*(2), 1–25.

Atkinson, R., M. Muro, and J. Whiton (2019). The case for growth centers: how to spread tech innovation across America. *Brookings Institution: Washington, DC, USA*.

Audretsch, D. B. and M. P. Feldman (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review 86*(3), 630–640.

Autor, D., D. Dorn, G. H. Hanson, G. Pisano, and P. Shu (2020). Foreign competition and domestic innovation: evidence from US patents. *American Economic Review: Insights 2*(3), 357–374.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review 103*(5), 1553–1597.

Barlevy, G. (2007). On the cyclicality of research and development. *American Economic Review 97*(4), 1131–1164.

Bartik, T. J. (1991). Who benefits from state and local economic development policies? Kalamazoo, MI: Upjohn Institute for Employment Research.

Bernstein, S., T. McQuade, and R. R. Townsend (2021). Do household wealth shocks affect productivity? Evidence from innovative workers during the Great Recession. *The Journal of Finance 76*(1), 57–111.

Blanchard, O. J. and L. F. Katz (1992). Regional evolutions. *Brookings Papers on Economic Activity 1992*(1), 1–75.

Canton, E. and H. Uhlig (1999). Growth and the cycle: creative destruction versus entrenchment. *Journal of Economics 69*, 239–266.

Carlino, G. and W. R. Kerr (2015). Agglomeration and innovation. In *Handbook of Regional and Urban Economics*, Volume 5, pp. 349–404. Elsevier.

Carlino, G. A., S. Chatterjee, and R. M. Hunt (2007). Urban density and the rate of invention. *Journal of Urban Economics 61*(3), 389–419.

Caselli, F. and G. Michaels (2013). Do oil windfalls improve living standards? Evidence from Brazil. *American Economic Journal: Applied Economics 5*(1), 208–238.

Clarivate Analytics (2020). Derwent World Patents Index. https://clarivate.com/derwent/wp-content/uploads/sites/3/dlm_uploads/2019/08/dwpi-classification-guide-2020.pdf.

Clarke, D. (2017). Estimating difference-in-differences in the presence of spillovers. MPRA Paper No. 81604. https://mpra.ub.uni-muenchen.de/81604/.

Combes, P.-P. and L. Gobillon (2015). The empirics of agglomeration economies. In *Handbook of Regional and Urban Economics*, Volume 5, pp. 247–348. Elsevier.

Comin, D. and M. Gertler (2006). Medium-term business cycles. *American Economic Review 96*(3), 523–551.

Corden, W. M. and J. P. Neary (1982). Booming sector and de-industrialisation in a small open economy. *The Economic Journal 92*(368), 825–848.

Correia, S., P. Guimarães, and T. Zylkin (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal 20*(1), 95–115.

Crescenzi, R. and A. Rodríguez-Pose (2013). R&D, socio-economic conditions, and regional innovation in the U.S. *Growth and Change 44*(2), 287–320.

Davis, S. J. and J. Haltiwanger (1990). Gross job creation and destruction: microeconomic evidence and macroeconomic implications. *NBER Macroeconomics Annual 5*, 123–168.

De Haas, R. and S. Poelhekke (2019). Mining matters: natural resource extraction and firm-level constraints. *Journal of International Economics 117*, 109–124.

Dechezleprêtre, A., D. Hémous, M. Olsen, and C. Zanella (2025). Induced automation innovation: evidence from firm-level patent data. *Journal of Political Economy, forthcoming*.

Duch-Brown, N. and M. T. Costa-Campi (2015). The diffusion of patented oil and gas technology with environmental uses: a forward patent citation analysis. *Energy Policy 83*, 267–276.

Dugoua, E. (2023). Induced innovation and international environmental agreements: evidence from the ozone regime. *Review of Economics and Statistics*, 1–45.

Fabrizio, K. R. and U. Tsolmon (2014). An empirical examination of the procyclicality of R&D investment and innovation. *Review of Economics and Statistics 96*(4), 662–675.

Feyrer, J., E. T. Mansur, and B. Sacerdote (2017). Geographic dispersion of economic shocks: evidence from the fracking revolution. *American Economic Review 107*(4), 1313–1334.

Florida, R. (2002). *The rise of the creative class*. New York: Basic Books.

Geroski, P. A. and C. F. Walters (1995). Innovative activity over the business cycle. *The Economic Journal 105*(431), 916–928.

Gilje, E. P. (2019). Does local access to finance matter? Evidence from US oil and natural gas shale booms. *Management Science 65*(1), 1–18.

Gilje, E. P., E. Loutskina, and P. E. Strahan (2016). Exporting liquidity: branch banking and financial integration. *The Journal of Finance 71*(3), 1159–1184.

Giroud, X. and H. M. Mueller (2017). Firm leverage, consumer demand, and employment losses during the Great Recession. *The Quarterly Journal of Economics 132*(1), 271–316.

Glaeser, E. L. and N. Hausman (2020). The spatial mismatch between innovation and joblessness. *Innovation Policy and the Economy 20*(1), 233–299.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: what, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Gorodnichenko, Y. and M. Schnitzer (2013). Financial constraints and innovation: why poor countries don't catch up. *Journal of the European Economic Association 11*(5), 1115–1152.

Griliches, Z. (1990). Patent statistics as economic indicators: a survey. *Journal of Economic Literature 28*(4), 1661–1707.

Gruber, J. and S. Johnson (2019). *Jump-starting America: how breakthrough science can revive economic growth and the American dream*. London: Hachette UK.

Hall, R. E. (1991). Recessions as reorganizations. *NBER Macroeconomics Annual 6*, 17–47.

Hasan, I., C.-K. S. Hoi, Q. Wu, and H. Zhang (2020). Is social capital associated with corporate innovation? Evidence from publicly listed firms in the US. *Journal of Corporate Finance 62*, 101623.

Hausman, N. (2022). University innovation and local economic growth. *Review of Economics and Statistics 104*(4), 718–735.

Holmes, T. J. and J. J. Stevens (2014). An alternative theory of the plant size distribution, with geography and intra-and international trade. *Journal of Political Economy 122*(2), 369–421.

Hunt, J. and M. Gauthier-Loiselle (2010). How much does immigration boost innovation? *American Economic Journal: Macroeconomics 2*(2), 31–56.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics 108*(3), 577–598.

James, A. G. and B. Smith (2020). Geographic dispersion of economic shocks: evidence from the fracking revolution: Comment. *American Economic Review 110*(6), 1905–1913.

Kovalenko, A. (2023). Natural resource booms, human capital, and earnings: evidence from linked education and employment records. *American Economic Journal: Applied Economics 15*(2), 184–217.

Ladino, J. F., S. Saavedra, and D. Wiesner (2021). One step ahead of the law: the net effect of anticipation and implementation of Colombia's illegal crops substitution program. *Journal of Public Economics 202*, 104498.

Lybbert, T. J. and N. J. Zolas (2014). Getting patents and economic data to speak to each other: an 'Algorithmic Links with Probabilities' approach for joint analyses of patenting and economic activity. *Research Policy 43*(3), 530–542.

Manso, G., B. Balsmeier, and L. Fleming (2023). Heterogeneous innovation over the business cycle. *The Review of Economics and Statistics 105*(5), 1224–1236.

Matray, A. (2021). The local innovation spillovers of listed firms. *Journal of Financial Economics 141*(2), 395–412.

Mian, A. and A. Sufi (2014). What explains the 2007–2009 drop in employment? *Econometrica 82*(6), 2197–2223.

Miguel, E. and M. Kremer (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica 72*(1), 159–217.

Moretti, E. (2021). The effect of high-tech clusters on the productivity of top inventors. *American Economic Review 111*(10), 3328–3375.

Moretti, E. and D. J. Wilson (2014). State incentives for innovation, star scientists and jobs: evidence from biotech. *Journal of Urban Economics 79*, 20–38.

Moretti, E. and D. J. Wilson (2017). The effect of state taxes on the geographical location of top earners: evidence from star scientists. *American Economic Review 107*(7), 1858–1903.

Mosquera, R. (2022). The long-term effect of resource booms on human capital. *Labour Economics 74*, 102090.

Nanda, R. and T. Nicholas (2014). Did bank distress stifle innovation during the Great Depression? *Journal of Financial Economics 114*(2), 273–292.

OECD (2009). *Patent Statistics Manual*. OECD Publishing, Paris.

Ouyang, M. (2011). On the cyclicality of R&D. *Review of Economics and Statistics 93*(2), 542–553.

Pelzl, P. and S. Poelhekke (2021). Good mine, bad mine: natural resource heterogeneity and Dutch disease in Indonesia. *Journal of International Economics 131*, 103457.

Popp, D. (2002). Induced innovation and energy prices. *American Economic Review 92*(1), 160–180.

Rajan, R. G. and L. Zingales (1998). Financial dependence and growth. *American Economic Review 88*(3), 559–586.

Roche, M. P. (2020). Taking innovation to the streets: microgeography, physical structure, and innovation. *Review of Economics and Statistics 102*(5), 912–928.

Sachs, J. D. and A. M. Warner (2001). The curse of natural resources. *European Economic Review 45*(4-6), 827–838.

Saunders, A. and S. Steffen (2011). The costs of being private: evidence from the loan market. *The Review of Financial Studies 24*(12), 4091–4122.

Schmoker, J. W. and T. R. Klett (1999). *US Geological Survey Assessment Model for Undiscovered Conventional Oil, Gas, and NGL Resources–the Seventh Approximation*, Volume 2165. US Department of the Interior, US Geological Survey.

Schmookler, J. (1966). *Invention and Economic Growth*. Harvard University Press.

Schumpeter, J. A. (1939). *Business cycles: a theoretical, historical and statistical analysis of the capitalist process*. New York: Mcgraw-Hill.

Stein, J. C. (1997). Internal capital markets and the competition for corporate resources. *The Journal of Finance 52*(1), 111–133.

Tian, X. and J. Xu (2022). Do place-based policies promote local innovation and entrepreneurship? *Review of Finance 26*(3), 595–635.

Valero, A. and J. Van Reenen (2019). The economic impact of universities: evidence from across the globe. *Economics of Education Review 68*, 53–67.

van Reenen, J. (2022). Innovation and human capital policy. In *Innovation and Public Policy*, pp. 61–84. University of Chicago Press.

van Wijnbergen, S. (1984). The 'Dutch disease': a disease after all? *The Economic Journal 94*(373), 41–55.

Weber, J. G. (2014). A decade of natural gas development: the makings of a resource curse? *Resource and Energy Economics 37*, 168–183.

Weinstein, A. L., M. D. Partridge, and A. Tsvetkova (2018). Follow the money: aggregate, sectoral and spatial effects of an energy boom on local earnings. *Resources Policy 55*, 196–209.

Xiao, H., A. Wu, and J. Kim (2021). Commuting and innovation: are closer inventors more productive? *Journal of Urban Economics 121*, 103300.

Online Appendix

"Local Booms and Innovation"

Federica Coelli and Paul Pelzl

December 22, 2025

Table of Contents

OA1 Additional Results

OA1.1 Innovation quality (Table OA1)

During boom periods, resource-rich commuting zones might prioritize incremental innovations to improve existing products or processes rather than focusing on new innovations. If this holds true, patent quality may decline in boom periods. To examine this hypothesis, we measure the average quality of patents using two indicators and examine whether average quality tends to deteriorate during booms.

Our first proxy for quality is based on the number of forward citations. We use citations because high-value inventions are more extensively cited than low-value patents (Harhoff et al., 1999). We calculate the average quality of patents in each period as follows. Let q_{ict} denote the number of forward citations five years after a patent i was filed, where t denotes the patent filing year. The cumulative sum in commuting zone c in period τ is:

$$Q_{ct} = \sum_{t \in \tau} \sum_{i \in c} q_{ict}$$

Let $P_{c\tau}$ be the total number of patents filed in period τ in commuting zone c. The average quality of patents filed each period is then calculated as $\bar{Q}_{c\tau} = Q_{c\tau}/P_{c\tau}$.

We also study patent generality as a second measure of patent quality, which captures the importance of patents for later developments and the number of fields they influence (Hall et al., 2001). We define the generality of patent i filed in year t:

$$g_{it} = 1 - \sum_{k \in K} \left(\frac{cit_{ik}}{cit_i} \right)^2$$

where cit_{ik} is the number of (five-year) citations from patents of technology class k to patent i, and $cit_i = \sum_k cit_{ik}$ denotes the total number of (five-year) citations to patent i. This measure resembles a Herfindahl–Hirschman index. A patent has high generality when it is cited by subsequent patents in many fields, whereas low generality occurs when citations are concentrated in a few specific fields. We scale the generality measure with the average generality of patents filed in the same year and technology classes to account for the fact that patent generality may increase over time. We then compute the average generality of patents in commuting zone c and period τ (see Section OA2 for details).

We proceed by estimating Equation (1) using the average quality and generality of patents as dependent variables. We include all commuting-zone-period observations with at least one patent into the sample. We do so because we want to distinguish true zeros, i.e. situations when citations or generality are zero, from situations where the quality and generality measures are undefined because there are no patents. The results are reported in Table OA1 and suggest that there is no significant decrease in the average quality (column 1) or generality (column 2) of patents during boom periods. Hence, the increase in non-oil and gas innovation does not seem to be paralleled by a deterioration of innovation quality.

OA1.2 Changing the period length from 3 years to other durations (Table OA2)

Given the somewhat arbitrary nature of defining a period in our analysis as a three-year window, in this section we test the robustness of our results to alternative period lengths and to using a lag structure. The results are presented in Table OA2. In column 1, we define a period simply as one year. In column 2, we use the same definition but lag our key explanatory variable, *ln(National oil&gas employment)* $\times$ *endowment*, by two years, similar to Dechezleprêtre et al. (2025) who also use a patent count as the dependent variable. Since in this specification, endowment is de facto only updated in 2003 (rather than in 2001), separately demeaning the estimating equation across the time frame for which we use conventional endowment and the time frame for which we use total endowment (as in our baseline) requires us to redefine the "centuries" in our fixed effects structure to 1969-2002 and 2003-2012. Note that because this important step makes the specification more complex and less intuitive relative to our baseline, we only use it in this robustness check. In column 3, we include three lags of our key interaction term. To make this specification tractable, we use total endowment for the entire sample period, instead of updating reserves to equal total endowment only in 2001. Therefore, we also replace commuting zone times century fixed effects with commuting zone fixed effects; thus, we demean the data once, rather than separately over the pre-fracking and the fracking era. In columns 4 and 6 of Table OA2, we estimate Equation (1) with a period defined as two and four years, respectively. Column 5 repeats the results of our baseline specification, where a period is three years long. In column 7, we define a period as roughly an entire decade, specifically using the periods 1969-1979, 1980-1990, 1991-2000, and 2001-2012. Considering the fracking boom, for instance, this specification imposes no restriction on how long it takes for the boom to affect local patenting. However, the specification has two downsides: (i) it averages national oil and gas employment over long and potentially diverse timeframes; and (ii), it implies a comparatively limited number of observations. Since we only have one period after 2000 in this specification, we cannot update endowment and demean the data separately over the two centuries; therefore, we use total endowment for the entire sample period and include only one dummy and one linear time trend per commuting zone.

The results show that our baseline findings (both on non-oil and gas patenting and oil and gas patenting) are robust to using the described alternative specifications. The results in column 3 show that the strongest patenting response to a boom occurs 2-3 years after the boom. This is consistent with the time lag found in the literature analyzing the effect of other shocks on patenting (Popp, 2002; Dugoua, 2023; Dechezleprêtre et al., 2025).

OA1.3 Testing for heterogeneous effects in the fracking period (Table OA3)

Given the significant changes to the oil and gas industry due to the fracking revolution and its strong impact on local economies (Feyrer et al., 2017), we analyze whether our baseline results in Table 2 differ across the pre-fracking and the fracking period. To this end, in columns 2, 4 and 6 of Table OA3, we include an interaction of the boom variable with a dummy that equals one starting from the three-year period 2001-2003 (columns 1, 3 and 5 repeat the results of columns 1, 2 and 3 of Table 2, respectively). The coefficient signs suggest that if anything, the effects become stronger after 2000, but the interaction terms are not statistically significant.

OA1.4 Oil and gas booms and local oil and gas activity (Table OA4)

In this section we show that during local oil and gas booms – as defined by our shift-share interaction – local oil and gas activity increases. This result complements our discussion in the beginning of Section 4.5. In terms of data on local oil and gas activity, the BEA's Regional Economic Accounts provide county-level information on employment in the "mining" sector. Besides the oil and gas sector, "mining" also comprises metals mining, coal mining, and mining and quarrying of nonmetallic minerals (except fuels). However, oil and gas represents more than two thirds of total mining employment at the national level during our sample period.

We aggregate county-level "mining" employment to the commuting zone level and use it as outcome variable to study local oil and gas activity during oil and gas booms. As shift variable in our shift-share interaction term, we use the oil price rather than national oil and gas employment to avoid simultaneity. We control for coal booms (using the coal price as shift variable), given that coal mining is the second-largest component of overall mining employment. The results show that local employment in oil and gas and other mining rises during local oil and gas booms (see Table OA4). Note that we do not use local oil and gas production as outcome variable since US production has typically taken more than three years to substantially respond to price changes (Konrad, 2012), while employment data capture production- and sales-enhancing activities that respond with a much smaller time lag.

OA1.5 Testing for additional mechanisms (Table OA5)

In this section, we discuss and empirically test for additional mechanisms which, as motivated below, a priori appear less likely to drive our main results. Consistent with this prior, we do not find compelling empirical support for any of these channels (see Table OA5).

Public finance

County governments receive a share of locally derived oil and gas revenue, which varies across states (ranging from 0.1% in Ohio to 2.3% in Alaska) and is primarily accrued via property taxes on oil and gas reserves, production, or related equipment (Newell and Raimi, 2018). Consistently, Table 1 shows that total own-source revenue of a commuting zone's county governments rises during oil and gas booms.[1] This raises the question of whether there is a public finance channel, through which local county governments in booming commuting zones use oil and gas windfalls to support local innovation. However, this channel is very unlikely to play a role: local governments spend oil and gas revenue mostly on public services such as primary education, as well as infrastructure projects that often become necessary as local oil and gas activity

[1] Unreported regressions show a positive effect of similar magnitude on own-source revenue specifically derived from local property taxes.

increases (Newell and Raimi, 2015, 2018). Nonetheless, we test the public finance hypothesis by exploiting the mentioned variation in counties' oil and gas revenue participation across states. We do so by regressing the number of non-oil and gas patents on our boom variable and an interaction of our boom variable with the share of oil and gas revenue accrued by counties in the producing commuting zone. This share (which does not vary within a state) is documented by Newell and Raimi (2018) for 16 US states, which jointly account for more than 97% of US oil and gas production. In our sample, we therefore include all commuting zones in these 16 states (except for Alaska, which is excluded from our baseline regressions), and further add commuting zones with zero total oil and gas endowment.[2] To isolate the impact of revenue sharing from the effect of other variables that vary across space, we further include all interaction terms from the specification in column 5 of Table 3, thereby accounting for factors such as urban/rural status or average human capital. The results, which are reported in column 2 of Table OA5, show that in commuting zones where the local county governments receive a larger share of local oil and gas revenue, local oil and gas booms have a significantly *smaller* impact on non-oil and gas patenting. This clearly speaks against the hypothesis that local county governments in booming commuting zones use oil and gas windfalls to support local innovation.

Strategic timing of patent filing

Next, our results might be explained by firms' decision to delay the implementation of innovation projects or the filing of patent applications to periods of high demand (Shleifer, 1986; Barlevy, 2007; Fabrizio and Tsolmon, 2014). Given our focus on *local* economic booms, such effects should be stronger for producers of locally-sold products, as they face a larger rise in demand during boom times. However, we find that the procyclicality of innovation is less pronounced for lowly traded goods producers compared to producers of highly traded goods (see Section 4.5), which speaks against this channel.

[2] For commuting zones that have zero total reserves (thus have neither conventional nor shale reserves), whenever data on oil and gas revenue participation is missing (we only have this information for 16 states), revenue sharing is defined to be zero. We do so because in the absence of local oil and gas endowment, there is no revenue sharing independent of the actual revenue sharing rule in the commuting zone's state. In practice, this choice has no implications on our results since the triple interaction *ln(National O&G employment)* $\times$ *endowment* $\times$ *O&G revenue share* takes the value zero no matter which value *O&G revenue share* takes, given that *endowment*=0.

Inventor-level wealth effects

Finally, booms might increase the wealth of local inventors (for example via house price increases), which might have a positive impact on their personal productivity. While we are unable to test this channel directly, Bernstein et al. (2021) find no relationship between inventor-level house price increases and innovative output during the boom preceding the Great Recession.[3] Given this evidence, which is also based on US data, a positive wealth shock mechanism appears unlikely to drive our own results.

OA1.6 Localizing oil and gas patenting (Table OA6)

This section complements our short description of the location of oil and gas patenting at the end of Section 3.2. Among all 63,000 oil and gas patents during 1969-2012, 76% were produced in metropolitan commuting zones (in comparison, 75% of *non*-oil and gas patents were produced in metropolitan commuting zones). The Houston, TX commuting zone – where most large American oil companies are headquartered – ranked first, producing 28% of oil and gas patents during 1969–2012. In terms of geographic distribution, all US states in our sample produced 20 or more oil and gas patents during 1969-2012, and 10 states produced more than 1,000 oil and gas patents each. This indicates that oil and gas innovation exhibits some degree of geographic spread rather than being very narrowly concentrated in big oil states like Texas, New Mexico, or North Dakota (which jointly accounted for 60% of oil and gas patenting over 1969-2012). Table OA6 reveals that, in the cross-section of commuting zones, average annual oil and gas patenting per capita over 1969-2012 increases with both oil and gas endowment and production.[4] Specifically, a one standard deviation increase in conventional oil and gas reserves per square mile is associated with 0.28 more oil and gas patents per year per 100,000 residents (see column 1). This positive association continues to hold when using total endowment rather than conventional endowment, thereby also accounting for fracking reserves (see column 2). Moreover, a one standard deviation rise in cumulative (1969-2012) oil and gas *production* per

[3] In contrast, they find significant negative effects during the Great Recession. These pieces of evidence lead the authors to conclude that above a certain level, additional personal resources do not appear to raise productivity, while losses do reduce productivity (for instance via psychological distress due to loss aversion).

[4] Average annual oil and gas patenting per capita over 1969-2012, at the commuting zone level, is computed as the ratio of total patents in 1969-2012 and the sum of all annual population counts over 1969-2012 divided by 100,000.

square mile is associated with 0.24 more oil and gas patents per year per 100,000 residents (see column 3). Unreported regressions show that these positive associations are robust to (i) excluding Texas, New Mexico and North Dakota; or (ii) restricting the sample to rural commuting zones and thereby excluding oil and gas hubs (in terms of both production and innovation) such as the broader Houston, TX area or the commuting zone including Odessa city, TX. In sum, the presented descriptive statistics and results indicate that oil and gas patenting extends beyond major oil and gas cities and frequently occurs within areas of large hydrocarbon production. This makes it meaningful to study the impact of local oil and gas booms on local oil and gas innovation as part of our main analysis.

OA1.7 Migration patterns during booms (Table OA7)

In this section we show the migration results discussed in Section 4.4, after providing details on the underlying data and discussing the used specifications. The migration data from the Internal Revenue Service (IRS) are based on year-to-year address changes reported on individual income tax returns submitted to the IRS, which keeps a record of all income tax filings. Individuals not obligated to file United States federal income tax returns are excluded from this dataset, leading to an underrepresentation of the poor and elderly. Additionally, the dataset omits the small proportion of tax returns submitted after late September of the filing year. Coverage is high nonetheless: Molloy et al. (2011) estimate that the data cover around 87% of all U.S. households. Data are available from 1991 until 2022, but post-2011 data are not comparable to prior data due to methodological changes.

The dataset which the IRS makes publicly available (see Section OA3 on how to access the data) reports the number of individuals moving from county i to county j between year $t-1$ and t, as long as at least 10 households actually moved from i to j between $t-1$ and t.[5] Given this censoring practice by the data provider, it is not possible to study in-migration from neighboring counties, for example. However, the data does separately report migration into j originating from counties in the same state, in-migration from counties in other states, and

[5] A move from i to j between $t-1$ and t implies that the household resided in county i in year $t-1$ and resided in county j in year t.

immigration from abroad.[6] We use the data on these variables to study how our boom measure (this time defined at the county level)[7] impacts county-level in-migration from the different stated origins, at an annual level. As outcome variable, we use the number of migrants moving into county j in year t coming from type of origin r, divided by j's initial population (thus population in 1990, which we express in 10,000 residents) following the migration literature (see e.g. Wilson, 2022).[8] In terms of the used specification, we estimate Equation (1) at the county-year level using OLS rather than Poisson, over the sample period 1991-2011. The results are reported in Table OA7. Column 1 shows that in a county with an oil and gas endowment of around five million dollars per square mile, an increase in national oil and gas employment by 100 log points leads to 4.43 more in-migrants per 10,000 residents. Column 2 shows that the coefficient on in-migrants coming from the same state equals 3.54, while column 3 shows that the coefficient on in-migrants coming from different states or abroad (which we study separately in columns 4 and 5) equals 0.89.[9] Only the coefficient on same-state in-migration is statistically significant (at the 5% level), but this is attributable to the larger coefficient size rather than a substantially lower standard error.

Given that we express our dependent variable as the ratio of in-migrants to initial population, our analysis in columns 2-3 effectively reflects a decomposition of the total in-migration effect (see column 1) into a same-state effect (column 2) and an outside-same-state effect (column 3): $3.54 + 0.89 = 4.43$. We can therefore conclude that among 100 migrants flowing into a booming county, $(3.54/4.43) \times 100 = 80$ migrants move in from counties in the same state, while the remaining $(0.89/4.43) \times 100 = 20$ migrants move in from elsewhere. Columns 4 and 5 show

6 Prior to 1995-1996, the data do not report these variables directly, but they can be easily computed using the reported data.

7 We study in-migration at the county level because aggregating the data to the commuting zone level would not make sense. To see this, suppose counties i and j make up one commuting zone and both counties experience an inflow of 100 people originating from other counties in the same state. In this case, we would not be able to identify how many of the 100 same-state migrants into i came from j as opposed to other counties in the same state but in different commuting zones, making it impossible to observe same-state in-migration at the commuting zone level.

8 On average, counties experienced an annual inflow of 574 people per 10,000 residents over 1991-2011, while the median is 504.

9 For 3% of county-years, the number of in-migrants coming from the same state versus other states or abroad is not reported. This can occur in county-years with comparatively little in-migration (in absolute numbers), for reasons related to the discussed data reporting threshold. Although we do have data on total in-migration for these county-years, we drop these county-years in column 1 in order to have the same sample across columns 1-3.

that in the average county, all of these 20 migrants move in from other US states as opposed to other countries. This suggests that migration from abroad plays virtually no role during local oil and gas booms in the US.

OA1.8 Extending the distance radius in Equation (4) (Table OA8)

In Table OA8, we include into Equation (4) the interaction of our standard boom variable with another distance doughnut, which captures the endowment of commuting zones located within 400-500 miles from the "home zone" c. The coefficients are close to zero, insignificant, and much smaller in absolute value than the coefficients for the 300-400 miles interaction. This suggests that non-negligible geographic spillovers only occur until 400 miles away from a booming commuting zone, on average. Based on the iterative, data-driven model selection approach suggested by Clarke (2017), this confirms Equation (4) as the optimal specification to estimate absolute effects net of geographic spillovers.

Table OA1: Patent quality

andDependent Variable →	Average # forward citations, non-O&G patents	Average Generality, non-O&G patents
	(1)	(2)
ln(National oil&gas employment) × endowment	-0.025	-0.024
	(0.035)	(0.029)
Observations	9,599	9,307
Sample period	69-09	69-09

Notes: In this table we estimate Equation (1) to study patent quality measures. In column 1, the dependent variable is the total number of forward citations of all non-oil and gas patents over the following five years, divided by the total number of non-oil and gas patents. This ratio proxies for the average quality of patents produced in a given three-year period. In column 2, the dependent variable is the average (and normalized) generality score across all non-oil&gas patents, which indicates whether patents are cited in many or few different technology fields. We only include commuting zone – three-year periods with at least one patent. The sample period omits the final period (2010-2012) as we cannot evaluate forward citations of patents in this period. In all columns we estimate Poisson pseudomaximum likelihood regressions. All regressions include commuting zone times century (1969-2000 versus 2001-2012) fixed effects and state times three-year period fixed effects, and control for commuting zone-specific trends over the pre-fracking period (1969-2000) and the fracking period (2001-2012). Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table OA2: Using time periods of different length and using lag structures

Period Length →	One year			Two years	Three years	Four years	Decade
RHS interaction term is measured in... →	t	t–2	t, t-1, t-2, t-3	Period average			
Panel I: Dependent Variable →	# Non-oil&gas patents						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln(National O&G empl.) × endowment	0.080**	0.081**	0.047	0.084**	0.089***	0.088***	0.065**
	(0.032)	(0.037)	(0.031)	(0.033)	(0.034)	(0.032)	(0.031)
ln(National O&G empl. in t-1) × endowment			0.012				
			(0.019)				
ln(National O&G empl. in t-2) × endowment			0.054***				
			(0.011)				
ln(National O&G empl. in t-3) × endowment			-0.014				
			(0.019)				
Observations	32,696	31,238	30,498	16,329	11,112	8,113	2,992
Sample period	69-12	71-12	72-12	69-12	69-12	69-12	69-12
Panel II: Dependent Variable →	# Oil&gas patents						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln(National O&G empl.) × endowment	-0.078**	-0.082***	-0.014	-0.081**	-0.085***	-0.077**	-0.072**
	(0.034)	(0.032)	(0.045)	(0.033)	(0.033)	(0.037)	(0.029)
ln(National O&G empl. in t-1) × endowment			0.012				
			(0.072)				
ln(National O&G empl. in t-2) × endowment			-0.013				
			(0.051)				
ln(National O&G empl. in t-3) × endowment			-0.038***				
			(0.014)				
Observations	21,857	20,585	19,615	11,309	7,767	4,956	2,020
Sample period	69-12	71-12	72-12	69-12	69-12	69-12	69-12

Notes: In this table we perform our baseline analysis using different definitions of the length of one period (columns 1 and 4-7) and using annual data with a lag structure (columns 2-3). The period length increases from left to right in the table: one year (columns 1-3), two years (column 4), three years (as in our baseline; see column 5), four years (column 6), and one decade (column 7), where we define decades as the time periods 1969-1979, 1980-1990, 1991-2000, and 2001-2012. In columns 1 and 4-6, we estimate Equation (1) using the indicated time period length. In column 2 we lag *ln(National O&G employment)* × *endowment* by two years, such that the sample period is reduced to 1971-2012. In this column we estimate Equation (1) but define the "centuries" 1969-2002 and 2003-2012 (instead of 1969-2000 and 2001-2012 as in our baseline) for our commuting zone times century fixed effects. We do so to account for the fact that endowment is de facto only updated in 2003 in this specification, due to the described two-year lag. In column 3 we include three lags of our key interaction term. To make this specification tractable, we use total endowment for the entire sample period (instead of updating reserves to equal total endowment only in 2001), and thus include commuting zone fixed effects instead of commuting zone times century fixed effects. In column 7, given that we only have one period after 2000, we estimate an adjusted version of Equation (1). This version includes commuting zone fixed effects instead of commuting zone times century fixed effects; includes one linear trend coefficient per commuting zone instead of two; and uses total endowment for the entire sample period rather than conventional endowment prior to 2001 and total endowment after 2000. In all columns we estimate Poisson pseudomaximum likelihood regressions. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table OA3: Testing for heterogeneous effects in the fracking era 2001-2012

Dependent Variable →	# Patents		# Non-oil&gas patents		# Oil&gas patents	
	(1)	(2)	(3)	(4)	(5)	(6)
ln(National oil&gas employment) × endowment	0.083***	0.080***	0.089***	0.086***	-0.085***	-0.083***
	(0.028)	(0.027)	(0.034)	(0.033)	(0.033)	(0.032)
ln(Nat. O&G empl.) × endow. × Fracking period		0.098		0.094		-0.113
		(0.082)		(0.091)		(0.189)
Observations	11,112	11,112	11,112	11,112	7,767	7,767
Sample period	69-12	69-12	69-12	69-12	69-12	69-12

Notes: In this table we test whether the effect of local oil and gas booms on local patenting activity (see Table 2, columns 1-3, which we repeat in columns 1, 3 and 5 of this table) differs across the pre-fracking period and the fracking period. We do so by interacting our baseline shift-share interaction term with a post-2000 dummy (see columns 2, 4 and 6). Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table OA4: Oil and gas booms and local oil and gas employment

Dependent Variable →	ln # Employees in oil&gas, mining and quarrying
	(1)
ln(Oil price) × oil and gas endowment 1960	0.021**
	(0.010)
ln(Coal price) × coal endowment 1960	0.095
	(0.062)
Observations	10,682
Sample period	69-12

Notes: In this table we analyze the impact of oil and gas booms on employment in oil and gas (and other mining) at the commuting zone level. We estimate Equation (1) but use the oil price rather than national oil and gas employment as shift variable, to avoid simultaneity. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table OA5: Testing for a public finance channel

Dependent Variable →	# Non-oil&gas patents	
Sample →	CZs in states with O&G revenue sharing info, and CZs without O&G endow.	
	(1)	(2)
ln(National oil&gas employment) × endowment	0.088**	0.138**
	(0.035)	(0.059)
ln(Nat. O&G empl.) × endowment × O&G revenue share		-0.152**
		(0.075)
ln(Nat. O&G empl.) × endowment × Urban non-metro		0.067**
		(0.032)
ln(Nat. O&G empl.) × endowment × Rural non-metro		0.159
		(0.212)
ln(Nat. O&G empl.) × endowment × Ini. pat-intensity		0.080**
		(0.035)
ln(Nat. O&G empl.) × endowment × Human capital		-0.046
		(0.035)
ln(Nat. O&G empl.) × endowment × College density		0.031
		(0.022)
Observations	6,152	6,152
Sample period	69-12	69-12

Notes: In this table we explore potential public finance effects. Fixed effects and standard error clustering are equivalent to Equation (1). For comparison, in column 1 we use the same sample as in column 2 and repeat the specification of Table 2, column 2. *O&G revenue share*, *Initial patenting intensity*, *Human capital*, and *College density* are scaled by their respective standard deviation. See the Notes of Table 3 for a description of the latter three variables. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table OA6: Correlating oil and gas supply with oil and gas patenting

Dependent Variable →	Average number of oil&gas patents per capita, 1969-2012		
Unit of observation →	Commuting zones (cross-section)		
	(1)	(2)	(3)
O&G endowment per square mile, 1960 (excl. fracking res.)	0.276***		
	(0.046)		
O&G endowment per square mile, 1960 (incl. fracking res.)		0.254***	
		(0.040)	
Total O&G production 1969-2012, per square mile			0.242***
			(0.037)
Observations	757	757	757

Notes: In this table we correlate measures of oil and gas supply with the intensity of oil and gas patenting over our sample period, both measured at the commuting zone level. The dependent variable is computed as the ratio of total patents in 1969-2012 and the sum of all annual population counts over 1969-2012 divided by 100,000. *Total O&G production 1969-2012, per square mile* is computed as the sum of annual production figures over 1969-2012, divided by commuting zone size in square miles. All explanatory variables are scaled by their respective standard deviations. All specifications include state fixed effects. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table OA7: Oil and gas booms and in-migration by type of origin

Dep. Var.: In-migration from ... / Initial Pop. →	All origins	Same state	Different states or abroad	Different states	Foreign countries
	(1)	(2)	(3)	(4)	(5)
ln(National oil&gas employment) × endowment	4.435*	3.545**	0.890	0.970	-0.080
	(2.515)	(1.708)	(1.806)	(1.710)	(0.156)
Observations	61,558	61,558	61,558	61,558	61,558
Sample period	91-11	91-11	91-11	91-11	91-11

Notes: In this table we study the impact of oil and gas booms on total in-migration and in-migration by type of origin. Given the nature of the migration data, the unit of observation is a county. The dependent variable equals the number of in-migrants from a certain origin type in a certain year, divided by initial population. Since our sample period is 1991-2011, initial population corresponds to population in 1990, which we express in 10,000 residents. In terms of the specification, we estimate an annual county-level version of Equation (1) using OLS rather than Poisson. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table OA8: Extending the distance in the geographic spillover analysis

Dependent Variable →	Adult Population		#Non-O&G Patents	
Proximity concept →	Distance doughnuts			
Data Frequency →	annual		3-year periods	
	(1)	(2)	(3)	(4)
ln(National oil&gas employment) × endowment	0.010***	0.010***	0.069**	0.069
	(0.003)	(0.003)	(0.034)	(0.045)
ln(N. O&G e.) × endow. of c-zones within 100 miles	0.008***	0.008***	0.046**	0.047**
	(0.003)	(0.003)	(0.023)	(0.023)
ln(N. O&G e.) × endow. of c-zones 100-200 miles away	0.007	0.007	0.008	0.009
	(0.006)	(0.006)	(0.064)	(0.061)
ln(N. O&G e.) × endow. of c-zones 200-300 miles away	-0.004	-0.004	0.012	0.010
	(0.011)	(0.011)	(0.097)	(0.122)
ln(N. O&G e.) × endow. of c-zones 300-400 miles away	-0.013*	-0.012*	-0.118	-0.118
	(0.007)	(0.007)	(0.102)	(0.100)
ln(N. O&G e.) × endow. of c-zones 400-500 miles away		-0.005		0.008
		(0.006)		(0.113)
Observations	32,868	32,868	10,962	10,962
Sample period	69-12	69-12	69-12	69-12

Notes: In this table we include an additional distance doughnut into the specification, relative to Equation (4) and columns 3-4 of Table 7. For comparison, column 1 repeats the results of column 3 of Table 7, while column 3 repeats the results of column 4 of Table 7. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

OA2 Details on patent data computations

OA2.1 Patent technology codes and technology classes

We use IPC codes to identify the technological characteristics of patents.

IPC classification system

The IPC codes form a hierarchical classification system; most patents have several of them.

The structure of the IPC classification is as follows:

- Section: Sections are the highest level of hierarchy of the Classification. Each section is designated by one of the capital letters A through H.

- Class: Each section is subdivided into classes which are the second hierarchical level of the Classification. Each class symbol consists of the section symbol followed by a two-digit number, e.g. H01.

- Subclass: Each class comprises one or more subclasses which are the third hierarchical level of the Classification. Each subclass symbol consists of the class symbol followed by a capital letter, e.g. H01S.

- Group: Each subclass is broken down into subdivisions referred to as "groups", which are either main groups (i.e. the fourth hierarchical level of the Classification) or subgroups (i.e. lower hierarchical levels dependent upon the main group level of the Classification). Each main group symbol consists of the subclass symbol followed by a one- to three-digit number, the oblique stroke and the number 00, e.g. H01S 3/00.

  - Subgroups form subdivisions under the main groups. Each subgroup symbol consists of the subclass symbol followed by the one- to three-digit number of its main group, the oblique stroke and a number of at least two digits other than 00, e.g. H01S 3/02.

In the following, we refer to sections, classes, subclasses, and main groups as 1-digit, 3-digit, 4-digit, and 6-digit codes respectively.

Fractional count

The patent office assigns one or (usually) several technology codes to each patent. When a patent comprises n different technology codes, we assign a weight to each technology code and we count the patent fractionally with a weight to each of the n technology codes.

We consider IPC 6-digit and IPC 4-digit codes, and we denote them as l and k respectively. Define the technology code vector for patent i filed in year t, p_{it}, as $L_{it} = \{l_{i1}, l_{i2}, ..., l_{in}\}$.

The weight of patent i's technology code k is:

$$\omega_{ik} = \frac{\sum_l IPC_{ikl}}{\sum_l IPC_{il}}, \tag{5}$$

with $\sum_k \omega_{ik} = 1$. For example, suppose the 6-digit IPC codes vector for patent i filed in year t, p_{it}, is $L_{it} = \{A01B1, A01B3, A01B5, A01C1\}$. The 4-digit IPC codes vector for patent p_{it} is $K_{it} = \{A01B, A01C\}$. Then, the weights of patent i's 4-digit IPC codes, K_{it}, are: $\omega_{iA01B} = 3/4$ and $\omega_{iA01C} = 1/4$.

OA2.2 Patent producing region

We use the inventor address of a patent to identify where a patent is invented. Using the inventor address gives a better approximation of where innovation is produced compared to using the applicant/assignee address.

Fractional count

Patents have one or (usually) more than one inventor. When a patent has inventors in different regions, we count patents fractionally. Let i denote patents and r regions. Ideally, we compute weights as

$$\omega_{ir} = \frac{I_{ir}}{I_i} \tag{6}$$

where I_{ir} is the number of patent i's inventors in region r, I_i is the total number of patent i's inventors, and $\sum_r \omega_{ir} = 1$.

However, in the earliest years of our sample, we only observe the location of the inventor, but not the number if inventors in each region r, therefore, we assign equal weight to each region r where the patens has inventors:

$$\omega_{ir} = \frac{1}{\#r} \tag{7}$$

where $\sum_r \omega_{ir} = 1$.

OA2.3 Patent generality

Denote:

- i: cited patent
- j: citing patent
- k: patent technology class

Let $cit_{ik} = \sum_j \omega_{jk} cit_{ij}$ be the number of (5-year) citations from patents of technology class k to patent i; and $cit_i = \sum_k cit_{ik}$ denote the total number of (5-year) citations to patent i. We count citations fractionally and ω_{jk} is the weight of patent j's technology class k defined in (5).

We define generality of patent i filed in year t as:

$$g_{it} = 1 - \sum_{k \in K} \left(\frac{cit_{ik}}{cit_i} \right)^2, \tag{8}$$

Note that the g_{it} is undefined if patents i is never cited.[10]

Normalization: Generality tends to be positively correlated with the number of citations a patent receives. To account for the fact that patent generality may increase over time, we scale the generality with the weighted average generality of patents filed in the same year and technology classes. We normalize the generality index as follows:

$$g_{it}^{norm} = \frac{g_{it}}{\bar{g}_t}, \tag{9}$$

[10] This can occur in two cases: if patent i is never cited, such that $cit_i = 0$, or if patent i is cited by patents with non-available technology codes because cit_{ik} is undefined for all $k \in K_j$.

where $\bar{g}_t = \frac{\sum_{k\in K_i}\omega_{ik}g_{kt}}{\sum_{k\in K_i}\omega_{ik}}$ is the weighted average generality of class k patents filed in year t and set of technology classes $k \in K_i$; K_i is the set of patent i's technology classes, and $g_{kt} = \frac{\sum_i \omega_{ik}g_{ik}}{\sum_i \omega_{ik}}$ is the average generality score of class k patents filed in year t. For example, suppose the generality of patent p_{it} is $g_{it} = 0.5$, and that, as in the above example, the 4-digit IPC codes vector for patent p_{it} is $K_{it} = \{A01B, A01C\}$, with weights $\omega_{iA01B} = 0.75$ and $\omega_{iA01C} = 0.25$. Suppose further that the weighted average generality of patents of technology class A01B and A01C filed in year t are $g_{A01Bt} = 0.6$ and $g_{A01Ct} = 0.3$. Then $\bar{g}_t = 0.75 \times 0.6 + 0.25 \times 0.3 = 0.525$ and the normalized generality score of patent i is $g_{it}^{norm} = \frac{0.5}{0.525} = 0.95$.

Region aggregation

We compute the average generality of patents in region r as the weighted average generality of patents filed in year t by inventors based in region r:

$$\bar{g}_{tr} = \frac{\sum_r \omega_{ir}g_{itr}}{\sum_r \omega_{ir}} \tag{10}$$

We compute this measure separately for O&G and non-O&G patents. Note that this regional aggregation excludes patents with missing generality.

OA2.4 Oil and gas patents by region

The number of O&G patents produced in US region r in year t is

$$p_{rt}^{OG} = \sum_i \omega_{ir} pat_{it}^{OG} \tag{11}$$

The number of non-O&G patents produced in US region r in year t is computed analogously.

OA2.5 Patents by technology and region

Let pat_{ikrt} denote a technology class k patent invented in region r in year t. We distinguish between O&G, pat_{ikrt}^{OG}, and non-O&G patents, pat_{ikrt}^{nOG}. The number of technology class k non-O&G patents produced in region r in year t is

$$p^{nOG}_{krt} = \sum_i \omega_{ik}\omega_{ir} pat^{nOG}_{ikrt} \tag{12}$$

We can compute p^{OG}_{krt} in the same way. $p^{nOG}_{krt} + p^{OG}_{krt} = p_{krt}$ is the total number of technology class k patents produced in region r in year t.

OA3 Online Data Appendix

This section complements Section 3, where we focused on describing our key variables: local oil and gas endowment and local patenting. We provide more detail, and/or describe other data sources and variable computations, dataset by dataset, below. In each of the below subtitles, we specify in brackets in which specific table(s) the variable or information is used – unless the variable or information is used in most tables.

Patent data (used in most tables)

Our primary source of patent data is the 2018 version of PATSTAT, from which we obtain all bibliographic and technological information. To geographically locate patents, we supplement PATSTAT with inventor address data from PatentsView and HistPat, as inventor locations at the time of filing provide a more accurate approximation of where innovation occurred than applicant addresses. We use PatentsView (version 31-12-2019) for patents filed from 1976 onward and HistPat for patents filed prior to 1976, as PatentsView coverage begins in 1976. Apart from geographic information, all data used in the analysis are sourced from PATSTAT.

Derwent World Patents Index (DWPI; used in most tables)

We identify oil and gas patents with the help of the DWPI classification system. Class H in the classification refers to petroleum, and identifies the relevant IPC codes. The data can be accessed here:

`https://clarivate.com/derwent/wp-content/uploads/sites/3/dlm_uploads/2019/08/DWPI-Classification-Guide-2020.pdf`.

The DWPI classifies four IPC classes as oil and gas technology classes. These classes jointly cover all aspects of the oil and gas industry, including obtaining crude oil and natural gas (exploration, drilling, well completion, production and treatment; C10G, E21B), unit operations (distillation, sorption and solvent extraction; C10G), petroleum processing (treating, cracking, reforming and gasoline preparation; C10G), processes pertaining to natural gas and LPG not included in C10G (C10L), and lubricants and lubrication (C10M). Below are the detailed descriptions of each code:

- C10G: Cracking hydrocarbon oils; production of liquid hydrocarbon mixtures, e.g. by destructive hydrogenation, oligomerization, polymerization; recovery of hydrocarbon oils from oil-shale, oil-sand, or gases; refining mixtures mainly consisting of hydrocarbons; reforming of naphtha; mineral waxes.
- C10L: Fuels not otherwise provided for; natural gas; synthetic natural gas obtained by processes not covered by subclasses C10G or C10K; liquefied petroleum gas; use of additives to fuels or fires; fire-lighters.
- C10M: Lubricating compositions; use of chemical substances either alone or as lubricating ingredients in a lubricating composition.
- E21B: Earth or rock drilling; obtaining oil, gas, water, soluble or meltable materials or a slurry of minerals from wells.

Oil and gas reserves (used in most tables)
Data on oil and gas reserves at the county level, as of 1960, have been generously shared by Hunt Allcott and Daniel Keniston (Allcott and Keniston, 2018). We aggregate these data to the commuting zone level. As described in Section 3, Allcott and Keniston (2018) compute reserves using various datasets, which originate from different sources. Oil and gas production data are mainly sourced from private data provider DrillingInfo; proven reserves data are from the Survey 23L data from the Energy Information Administration (EIA); and undiscovered reserves are estimated by the United States Geological Survey (USGS) based on expected oil, gas, and natural gas liquid yield using current technologies, including estimated future discoveries throughout the next 30 years.

National oil and gas employment (used in most tables)

Data are provided by the Bureau of Economic Analysis (BEA), via *SAEMP25: Total full-time and part-time employment by industry.* The data can be accessed at `https://apps.bea.gov/itable/?ReqID=70&step=1`. The BEA classifies industries according to the Standard Industrial Classification (SIC) from 1969-2000, and according to the North American Industry Classification System (NAICS) from 2001 onwards. In the SIC, we use employment in industry 13 = oil and gas extraction, which contains 131=crude petroleum and natural gas, 132=natural gas liquids, and 138=oil and gas field services. Note that SIC industry 138 does not map 1:1 to NAICS, since NAICS combines support activities for oil and gas extraction and support activities for mining in its subcategory 213. Following Allcott and Keniston (2018), we therefore define post-2000 oil and gas employment as the sum of employment in NAICS=213 and NAICS 211=oil and gas extraction. A comparison of SIC and NAICS data, which is possible for the years 1998-2000, reveals that this is a minor issue because employment in support activities for mining is comparatively very small.

Population (used in Tables 3, 1, A2, OA6)

County-level population data are obtained from the Bureau of Economic Analysis' (BEA) Regional Economic Accounts (previously referred to as Regional Economic Information System (REIS)). The data can be accessed at `https://apps.bea.gov/regional/downloadzip.cfm` or `https://apps.bea.gov/itable/?ReqID=70&step=1`. The population (number of persons) variable is available in the dataset *CAINC4*. Note that the data sources of population by educational attainment and by age are described further below.

Employment (used in Table 1)

County-level employment data are sourced from the BEA's Regional Economic Accounts (`https://apps.bea.gov/regional/downloadzip.cfm` or `https://apps.bea.gov/itable/?ReqID=70&step=1`). We use the variable "Total employment (Number of jobs)", which is available in the dataset *CAEMP25* or in the dataset *CAINC4*.

Personal income (used in Tables 9, 1, A2)

County-level personal income data are sourced from the BEA's Regional Economic Accounts (`https://apps.bea.gov/regional/downloadzip.cfm` or `https://apps.bea.gov/itable/?ReqID=70&step=1`), specifically from the dataset *CAINC4*. The BEA computes personal income as "the sum of wages and salaries, supplements to wages and salaries, proprietors' income with inventory valuation (IVA) and capital consumption adjustments (CCAdj), rental income of persons with capital consumption adjustment (CCAdj), personal dividend income, personal interest income, and personal current transfer receipts, less contributions for government social insurance plus the adjustment for residence". The data are reported as the sum across the entire county, so (after aggregating to the commuting zone level) we divide by population to compute personal income per capita. We deflate these nominal data using a consumer price index from the St. Louis Fed (see `https://fred.stlouisfed.org/series/CPIAUCNS`).

GDP (used in Table 1)

County-level GDP, in "thousands of chained 2012 dollars", is sourced from the BEA's Regional Economic Accounts (`https://apps.bea.gov/regional/downloadzip.cfm` or `https://apps.bea.gov/itable/?ReqID=70&step=1`), specifically from the dataset *CAGDP1*. GDP data are only available from 2001 onwards. We scale by population to compute GDP per capita.

(Local-level) oil and gas employment (used in Table OA4)

Oil and gas employment is not available as a separate data item in the Regional Economic Accounts. It is included in the broader "mining" item (SIC industry B, NAICS industry 21), which also includes metals mining, coal mining, and mining and quarrying of non-metallic minerals (except fuels). We use these data (available via the dataset *CAEMP25* at `https://apps.bea.gov/regional/downloadzip.cfm` or `https://apps.bea.gov/itable/?ReqID=70&step=1`) to measure local oil and gas employment over time. This is meaningful because national data reveals that the oil and gas sector represents more than two-thirds of total US mining employment during our sample period.[11]

[11] Note that oil and gas and other mining employment is not reported for some counties and years, in some cases to avoid disclosure of confidential information. When aggregating from the county to the commuting zone level, we treat those observations as equal to zero. The result that local employment in oil and gas and other mining rises during local oil and gas booms (see Table OA4) is robust to using the median employment value across all counties in the particular year.

Population by age (used in Tables 3, 6, 7, A2, OA8)

In column 4 of Table 6, our dependent variable is adult population at the commuting zone level, measured annually (unlike in column 1, where we compute adult population as the sum of college-educated adult population and non-college-educated adult population, both of which are not available annually; see the next item *Adult population by educational attainment*). To compute commuting zone-level adult population at an annual basis, we obtain annual county-level data on population by age, from the National Cancer Institute at `https://seer.cancer.gov/popdata/download.html#single`.

Adult population by educational attainment (used in Tables 3, 6, 9, A2, and OA5)

The nature of local-level data on educational attainment has evolved over time. Below, we describe two different educational attainment variables, where we use them in our analysis, and from where we obtained the data.

Until (and including) the 1980 round of the decennial population census, no question on actual college degree obtainment was included; instead, years of schooling was asked. Therefore, to measure local human capital at the beginning of our sample period (as used in Tables 3, 9, A2, and OA5), we compute the fraction of adults (age 25+) with at least one year of college education as of 1970. The data are obtained from the Economic Research Service (ERS) at `https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/`. More specifically, we take the sum of the two variables "Some college (1-3 years)" and "Four years of college or higher".

In our analysis in Table 6, we want to study local adult population by actual educational attainment (in terms of degrees rather than years of college education) over time. We use the more recent census data for this purpose, specifically data from 1990 and 2000. The data are accessible at the same web link as above, and are contained in the variable "Number of adults (age 25+) with bachelor degree or higher". Starting from 2010, the census no longer includes questions on education. Therefore, we resort to data from the American Community Survey

(ACS), which we downloaded from the Census website `data.census.gov`.[12] Note that the ACS has "insufficient coverage for a reliable county estimate in one year" (Weber, 2014). Therefore, we use the five-year rather than the one-year data, where the five-year data is an average over the indicated year and the previous four years. We use ACS2010_5Y, which represents averages across 2006-2010, and ACS2015_5Y, which represents averages across 2011-2015.[13]

Creative class workers (used in Table 6)

The concept of creative class workers was originally defined by Florida (2002). We use a refined classification by the Economic Research Service (ERS), which is accessible at `https://www.ers.usda.gov/data-products/creative-class-county-codes/`. The data lists "population employed in occupations that require "thinking creatively."" The ERS provide county-level data for 1990, 2000, and the average over the 2007-11 ACS rounds.

County government revenue (used in Table 1)

Panel data on county government revenue are collected via the five-yearly Census of Governments. The data can be downloaded at `https://www.census.gov/programs-surveys/gov-finances/data/historical-data.html`. We use the years 1972, 1977, 1982, 1987, 1992, 1997, 2002, 2007, and 2012. Total revenue is the sum of i) tax revenue (which includes property tax revenue, for instance), ii) intergovernmental transfers (IG), and iii) other, non-tax and non-IG revenue. Own-source revenue is the sum of i) and iii).[14]

Classifying commuting zones into metropolitan versus non-metropolitan etc.
(used in Tables 3, 9, A2, and OA5)

The Economic Research Service (ERS) publishes the Rural-Urban Continuum Codes, which provide information on how urban versus rural a certain US county is. They were originally developed in 1974 and have been updated each decennial since (1983, 1993, 2003, 2013). We use the 1974 classification since it reflects most closely the county's urban- versus rural-ness

[12] Specifically, see `https://data.census.gov/cedsci/table?q=american%20community%20survey%20education&tid=ACSST1Y2019.S1501`

[13] Comparing values across the census and five-yearly ACS averages is feasible and has been done in previous studies, see for example Weber (2014).

[14] See for example the 2012 questionnaire: `https://www2.census.gov/govs/forms/2012/f28_12.pdf`

at the beginning of our sample period. The data can be accessed at `https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/`.

Each county is assigned a value from 0 to 9, ranging from "Central county of metro areas of 1 million population or more" (Code=0) to "Completely rural or less than 2,500 urban population, not adjacent to a metro area" (see Figure OA1). Counties with value 4 to 9 are classified as "non-metropolitan". We bring this classification to the commuting zone level by taking the average value across all counties within a commuting zone, and define commuting zones with an average value of 4 or larger as non-metropolitan. We define a commuting zone as urban non-metropolitan if the rural-urban code is larger than (or equal to) 4 and smaller than 8. Rural non-metropolitan commuting zones are those with a code larger or equal to 8.

Figure OA1: Rural-Urban Continuum Codes

Description of the Rural-Urban Continuum Codes prior to 2003	
Code	Description
Metro counties:	
0	Central counties of metro areas of 1 million population or more.
1	Fringe counties of metro areas of 1 million population or more.
2	Counties in metro areas of 250,000 to 1 million population.
3	Counties in metro areas of fewer than 250,000 population.
Nonmetro counties:	
4	Urban population of 20,000 or more, adjacent to a metro area.
5	Urban population of 20,000 or more, not adjacent to a metro area.
6	Urban population of 2,500 to 19,999, adjacent to a metro area.
7	Urban population of 2,500 to 19,999, not adjacent to a metro area.
8	Completely rural or less than 2,500 urban population, adjacent to a metro area.
9	Completely rural or less than 2,500 urban population, not adjacent to a metro area.

Notes: This figure describes the classification of counties into the 10 different Rural-Urban Continuum Codes, as of the 1974 edition of the data. Figure source: `https://wayback.archive-it.org/5923/20110914000642/http://www.ers.usda.gov/Briefing/Rurality/RuralUrbCon/priordescription.htm`.

College density (used in Tables 3 and OA5)

College-level employment data are provided in the dataset *Homeland Infrastructure Foundation-Level Data* (HIFLD), which can be accessed at `https://hifld-geoplatform.opendata.arcgis.com/datasets/colleges-and-universities/explore`. These data are collected by the U.S. Department of Homeland Security. The dataset is composed of all Post Secondary Education facilities in the academic year 2018-19, as defined by the Integrated Post Secondary Education System (IPEDS), National Center for Education Statistics, US Department of Education (see `https://www.sciencebase.gov/catalog/item/4f4e4acee4b07f02db67fb39`). The dataset contains 6,559 institutions, which are classified into the following NAICS sectors (in parentheses we report the number of institutions): Business and Secretarial Schools (29), Colleges, Universities, and Professional Schools (2,579), Computer Training (20), Cosmetology and Barber Schools (1,178), Educational Support Services (71), Fine Arts Schools (34), Flight Training (13), Junior Colleges (1,562), Other Technical and Trade Schools (1,073). In our analysis, we only consider "Colleges, Universities, and Professional Schools" (NAICS code 611310), since this category appears most relevant for local innovation.[15] The dataset contains a variable indicating the county in which the college is located, which enables us to compute total employment in NAICS=611310 institutions at the commuting zone level.

Classifying industries as upstream to oil and gas (used in Table 4)

We use the 1987 Input-Output tables of the *Bureau of Economic Analysis* (BEA) to identify upstream plants. The dataset distinguishes 362 industries (BEA codes) within the manufacturing sector. The data can be accessed at
`https://www.bea.gov/industry/historical-benchmark-input-output-tables`, in the zip-file *1987 Benchmark I-O Table Six-Digit Transactions*, where we use TBL2-87 = "The use of commodities by industries".

Following Allcott and Keniston (2018), we classify four BEA industries as belonging to the oil and gas sector: 80000 (Crude petroleum and natural gas); 110601 (Petroleum and natural gas well drilling); 110602 (Petroleum, natural gas, and solid mineral exploration); and 120215

[15] Note that employment data are missing for several colleges. When aggregating from the county to the commuting zone level, we treat these observations as being equal to zero. The results (see Table 3) are robust to instead using the median number of college employees.

(Maintenance and repair of petroleum and natural gas wells). For each industry j in the input-ouput table, we compute its "upstreamness" to the oil and gas sector as the ratio of the sum of its direct and indirect sales to the oil and gas sector (as defined above) and its total sales ($Upstream_j$ thus takes a value between zero and one):

$$Upstream_j = \frac{Sales_{j,OG}}{\sum_j Sales_j} + \sum_{-J} \left[\frac{Sales_{j,-j}}{\sum_j Sales_j} \times \frac{Sales_{-j,OG}}{\sum_j Sales_{-j,j}} \right]$$

where $-J$ denotes the set of all industries apart from j.

We then walk from the BEA code to the 4-digit SIC87 code using a concordance table provided in the above-mentioned zip-file. Note that while a certain BEA industry sometimes maps to multiple SIC87 industries, each SIC87 industry maps to one unique BEA industry. We then define a 4-digit SIC87 industry as upstream to oil and gas if $\text{Upstream}_j > 0.01$, following Allcott and Keniston (2018).

Classifying industries into highly- versus lowly traded (used in Table 8)

For each of 457 4-digit SIC-1987 manufacturing industries, Holmes and Stevens (2014) estimate a (constant) distance elasticity, which equals the percentage reduction in trade volume as distance increases by one percent. For this purpose the authors use data from the 1997 *U.S. Commodity Flow Survey* (CFS), which documents the destination, product classification, weight and value of a broad sample of shipments. Holmes and Stevens (2014) estimate the distance elasticity via a standard log-log specification. The higher the trade costs of a specific industry, the shorter its optimal average shipment distance (equivalently, the higher its distance adjustment). Ready-Mix Concrete (4.2), Ice (3.0) and Asphalt (2.9) have the highest estimated distance elasticity. 29 industries have an estimated distance elasticity of zero, including Semiconductors, Analytical laboratory instruments and Aircraft, in which transportation costs are very low relative to product value. The data can be obtained at `http://users.econ.umn.edu/~holmes/data/plantsize/description_of_supplementary_files.html`.[16]

Based on the 456 sectors that are represented in our patent-by-industry data (see Section 4.3.2) and in the Holmes-Stevens data, we compute the median distance elasticity. This me-

[16] Note that SIC1992 in the data corresponds to SIC 1987; see `https://guides.loc.gov/industry-research/classification-sic`.

dian equals 0.58. Industries with a below-median value are classified as relatively highly traded, while all others are classified as relatively lowly traded. In another exercise, we classify industries into a most-traded tercile, an intermediate tercile, or a least-traded tercile. The cutoff for the most-traded tercile lies at a distance elasticity of 0.77. This value is close to 0.8, which corresponds to an average shipment distance of approximately 500 miles (and equals the cutoff chosen by Allcott and Keniston, 2018). Industries in the least-traded tercile thus have an average shipment distance of below 500 miles, while industries in the medium- and highly traded tercile have a larger average shipment distance.

Oil price (used in Table 9, Figures A3-A5, Table OA4)

We use the spot crude oil price of the blend West Texas Intermediate (WTI). Data are obtained from the Federal Reserve Bank of St. Louis and can be accessed here: `https://fred.stlouisfed.org/series/WTISPLC`. We average the monthly data to the annual level and then deflate the nominal series using a consumer price index (which we also first average to the annual level) from the St. Louis Fed (see `https://fred.stlouisfed.org/series/CPIAUCNS`).

Natural gas price (used in Figure A4)

Data on US Natural Gas Wellhead Prices are obtained from the EIA, at `https://www.eia.gov/dnav/ng/hist/n9190us3A.htm`.

Petroleum & Coal sector profits (used in Figure A5)

Data on corporate profits in the petroleum and coal sector are obtained from the BEA, at `https://www.bea.gov/data/income-saving/corporate-profits`.

National coal employment (used in Table 9)

Data from 1969-2000 are provided by the BEA via *SAEMP25: Total full-time and part-time employment by industry*. The data can be accessed at `https://apps.bea.gov/itable/?ReqID=70&step=1`. The corresponding SIC code is 12=Coal mining. Data from 2001 onwards is obtained via the BLS Data Viewer, via `https://beta.bls.gov/dataViewer/view/timeseries/IPUBN2121__W200000000`. The corresponding series name is *Annual employment (thousands of* *jobs) for NAICS 2121, coal mining, U.S. total.*[17,18]

Coal endowment (used in Tables 9 and OA4)

We obtain county-level data on coal endowment as of 1960 from Allcott and Keniston (2018).

Migration data from Internal Revenue Service (used in Table OA7)

The IRS migration data can be downloaded from `https://www.irs.gov/statistics/soi-tax-stats-migration-data`, under *County-to-county migration data.*

County splits and mergers (used in all tables)

Data on county splits and mergers are obtained from `https://www.census.gov/programs-surveys/geography/technical-documentation/county-changes.1980.html`.

County to commuting zone crosswalk (used in all tables)

We use data from the ERS, available at `https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/`. We depart from the data for the year 1980, since the commuting zone concept was first developed in the 1980s. We then complement these data with data on county splits and mergers (see above) to define commuting zones as of 1969.

Cartographic boundary shapefile (used in Figure 2)

We create the maps of commuting zone-level oil and gas endowment (Figure 2) based on a shapefile of US counties as of 1970. The shapefile can be downloaded from `https://www.nhgis.org/`. We choose the year 1970 because we define commuting zones as of 1969 for our

[17] Contrary to SAEMP25S, which does report employment for SIC=12, SAEMP25N does not separately report employment for NAICS=2121=Coal mining. It is for this reason that we use a slightly different data source for the years 2001-2012.

[18] This series is available from 1987 onwards. Comparing data points from 1987-2000 across SIC and NAICS using the two sources reveals that employment in NAICS=2121 is around 10% lower on average than employment in SIC=12, although both series have the name "Coal mining". However, this difference does not affect our results due to our fixed effects structure, which "separates" the periods 1969-2000 and 2001-2012 in our regressions.

analysis, and county shapefiles are not available annually but only decennially (in US census years).

References

Allcott, H. and D. Keniston (2018). Dutch disease or agglomeration? The local economic effects of natural resource booms in modern America. *The Review of Economic Studies 85*(2), 695–731.

Barlevy, G. (2007). On the cyclicality of research and development. *American Economic Review 97*(4), 1131–1164.

Bernstein, S., T. McQuade, and R. R. Townsend (2021). Do household wealth shocks affect productivity? Evidence from innovative workers during the Great Recession. *The Journal of Finance 76*(1), 57–111.

Clarke, D. (2017). Estimating difference-in-differences in the presence of spillovers. MPRA Paper No. 81604. https://mpra.ub.uni-muenchen.de/81604/.

Dechezleprêtre, A., D. Hémous, M. Olsen, and C. Zanella (2025). Induced automation innovation: evidence from firm-level patent data. *Journal of Political Economy, forthcoming*.

Dugoua, E. (2023). Induced innovation and international environmental agreements: evidence from the ozone regime. *Review of Economics and Statistics*, 1–45.

Fabrizio, K. R. and U. Tsolmon (2014). An empirical examination of the procyclicality of R&D investment and innovation. *Review of Economics and Statistics 96*(4), 662–675.

Feyrer, J., E. T. Mansur, and B. Sacerdote (2017). Geographic dispersion of economic shocks: evidence from the fracking revolution. *American Economic Review 107*(4), 1313–1334.

Florida, R. (2002). *The rise of the creative class*. New York: Basic Books.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). The NBER patent citation data file: lessons, insights and methodological tools. NBER Working Paper 8498.

Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel (1999). Citation frequency and the value of patented inventions. *The Review of Economics and Statistics 81*(3), 511–515.

Holmes, T. J. and J. J. Stevens (2014). An alternative theory of the plant size distribution, with geography and intra-and international trade. *Journal of Political Economy 122*(2), 369–421.

Konrad, T. (2012). The end of elastic oil. Forbes, January 26, 2012: https://www.forbes.com/sites/tomkonrad/2012/01/26/the-end-of-elastic-oil/?sh=4462954f36d6.

Molloy, R., C. L. Smith, and A. Wozniak (2011). Internal migration in the United States. *Journal of Economic Perspectives 25*(3), 173–196.

Newell, R. G. and D. Raimi (2015). Shale public finance: local government revenues and costs associated with oil and gas development. NBER Working Paper 21542.

Newell, R. G. and D. Raimi (2018). US state and local oil and gas revenue sources and uses. *Energy Policy 112*, 12–18.

Popp, D. (2002). Induced innovation and energy prices. *American Economic Review 92*(1), 160–180.

Shleifer, A. (1986). Implementation cycles. *Journal of Political Economy 94*(6), 1163–1190.

Weber, J. G. (2014). A decade of natural gas development: the makings of a resource curse? *Resource and Energy Economics 37*, 168–183.

Wilson, R. (2022). Moving to economic opportunity: the migration response to the fracking boom. *Journal of Human Resources 57*(3), 918–955.