Local Booms and Innovation*

Federica Coelli[a] and Paul Pelzl[b]

[a]University of Zurich (UZH) & CEPR

[b]NHH Norwegian School of Economics

February 28, 2024

Abstract

Using oil and gas shocks as exogenous source of local business cycles in U.S. commuting zones, we provide novel evidence that innovation responds very locally to economic booms. Total patenting is procyclical, which we show to be driven by positive agglomeration economies. In contrast, patenting in oil and gas technology – the sector most closely linked to the boom – is countercyclical, consistent with larger opportunity costs of innovation when an industry is booming. These findings shed new light on the spatial dimension of innovation, inform the recent policy debate on local industrial policy, and help explain the mixed evidence on the cyclicality of innovation.

JEL codes: L71, O12, O30, Q30, Q35

Keywords: Innovation, patents, natural resources, regional development & inequality

*We thank Hunt Allcott and Daniel Keniston for generously sharing data, and Knut Blind, Eugenie Dugoua, Marion Dumas, Gunnar Eskeland, Torfinn Harding, David Hemous, Hans Hvide, Ralf Martin, Maria-Eugenia Sanin, Florian Seliger, Vincent Sterk, Jan Stuhler, Bram Timmermans, seminar participants at NHH Norwegian School of Economics and University of Stavanger, and conference participants at the Ninth ZEW/MaCCI Conference on the Economics of Innovation and Patenting, the MaCCI Annual Conference 2023, and the EAERE Conference 2023 for valuable comments.
Federica Coelli: University of Zurich, Schönberggasse 1, 8001 Zurich. Email: federica.coelli@econ.uzh.ch.
Paul Pelzl: NHH Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway. Email: paul.pelzl@nhh.no.

1 Introduction

In the United States, innovation is highly concentrated in some large metropolitan areas. For example, residents of Santa Clara County in the Silicon Valley produced an average of 3,000 patents per year over 1970-2010, resulting in a staggering 2.2 patents per thousand residents. In many rural areas of the country, innovative activity is largely lacking, which is often paralleled by large joblessness (Glaeser and Hausman, 2020). Okmulgee County in Oklahoma, for instance, had 35,000 inhabitants, one patent annually, and a ratio of employment to working age population (age 15-64) equal to a mere 58% during 1970-2010. These disparities have sparked a vivid policy debate in recent years. For example, Gruber and Johnson (2019) and Atkinson et al. (2019) advocate a stronger geographic dispersion of research funding to promote local innovation clusters outside of superstar cities, to help "jump-start the American growth engine" (Gruber and Johnson, 2019, p.11). However, other studies have challenged this view, 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 an economic impetus at the local geographic level. Does a rise in local economic activity promote the creation of new ideas? Are the effects of economic booms undone during downturns, or are they sustained? Does the impact vary across urban areas and more remote places, or across other local characteristics such as human capital? And, what are the mechanisms at play? These questions have been largely overlooked in existing literature, but they are crucial for a better understanding of local innovation dynamics and regional development.

The contribution of this paper is twofold. First, we provide novel and plausibly causal evidence that innovation responds very locally – and positively – to economic booms whose genesis is unrelated to innovation. We derive this result from a long sample of US patenting activity at the commuting zone level covering over four decades (1969-2012). Notably, local innovation is procyclical also in commuting zones with little patenting, and in fact we observe the strongest effects in commuting zones that are urban but not metropolitan. These findings are highly policy relevant, as they suggest that in regions far from the current hotspots of American innovation, innovative activity does respond relatively strongly to an increase in economic activity – and thus potentially also to place-based policies. Our results also indicate that patent *quality* is not lower during local economic upswings, and that the rise in innovation during booms is not offset by a decline in patenting during busts. This suggests that economic booms can have a lasting impact on local innovation, and thereby potentially on local economic development. The effects are stronger for commuting zones with more human capital and/or a larger patenting intensity at the beginning of our sample period. In terms of innovation by type, we find that patenting typically rises more strongly in technologies the commuting zone is historically familiar with, indicating path dependence in innovation (see also Aghion et al., 2016; Manso et al., 2023). We also test for several mechanisms that may explain the procyclicality of local innovation, and identify positive agglomeration economies as the main channel. Other potential mechanisms such as relaxed financial constraints are not found to play an important role.

Our second contribution is to highlight heterogeneous effects across industries, which help explain a much-debated puzzle in the literature on the cyclicality of innovation. Specifically, theoretical work inspired by Schumpeter (1939) predicts that firms undertake productivity-improving activities during recessions, because of temporarily lower opportunity costs (Davis and Haltiwanger, 1990; Hall, 1991; Aghion and Saint-Paul, 1998). However, empirical studies typically find that innovation is procyclical (Geroski and Walters, 1995; Comin and Gertler, 2006; Ouyang, 2011). While several papers have addressed this puzzle in different ways, our identification strategy allows to empirically test – and confirm – the 'opportunity cost theory' in a novel, more direct, and more plausibly causal fashion. In particular, we find that sectors that are more exposed to higher aggregate demand during exogenously determined local booms, and thereby face a larger increase in their opportunity cost of innovation, either keep patenting constant or reduce it. In contrast, sectors whose product demand is less affected by the local boom significantly raise patenting. This shows that, contrary to earlier conjectures in the literature, the procyclicality of innovation cannot be explained by a low empirical relevance of varying opportunity costs of innovation over the business cycle.

We measure local economic booms through exogenous oil and gas shocks. These shocks are defined as the interaction of a commuting zone's initial oil and gas endowment with time-series variation in national oil and gas employment. This shift-share approach is similar to Allcott and Keniston (2018), who kindly shared their proprietary data on county-level oil and gas endowment with us. Our long sample period covers the US oil boom of the late 1970s and early 1980s, a long ensuing bust until the late 1990s, and the recent fracking boom in the 2000s, each of which created or destroyed hundred thousands of oil and gas jobs. Initial oil and gas endowment is a function of geology – especially because it includes "undiscovered reserves" – and does not correlate with other commuting zone characteristics. This makes our "shares" much more exogenous relative to most other shift-share designs, but nonetheless we account for recent methodological advances in this literature (Goldsmith-Pinkham et al., 2020).

While our shift-share interaction is unrelated to time-varying local confounders, it is a significant driver of various outcome measures of local economic activity, and thereby serves as a very good proxy for local economic booms. Specifically, in the commuting zone with an initial oil and gas endowment of five million dollars per square mile, a doubling of national oil and gas employment leads to a rise in population by 1.9%, employment by 3.7%, earnings per worker by 2.2%, GDP by 5.4%, and local government revenue by 6.3%. In terms of innovation, the total number of granted patents rises by 8.8%, which is equivalent to roughly one more patent in the commuting zone with median patent activity. By using oil and gas shocks to proxy for local business cycles, our approach is similar to Feyrer et al. (2017), who exploit local fracking booms to study the geographic dispersion of local economic shocks.

The result of procyclical overall innovation masks substantial heterogeneity. First, we show that local patenting in oil and gas – the industry most directly affected by the local boom – is *counter*cyclical. This is consistent with theory, since the higher prices or lower production costs that have characterized national peaks of oil and gas employment imply larger profits for local producers, thereby raising the opportunity cost of innovation.[1] In contrast, we find that patenting in non-oil and gas technology – which represents 98% of total patenting during our sample period – is *pro*cyclical overall. However, the effect differs markedly across different sectors of manufacturing, which jointly accrues for 97% of total patenting. Our sector distinction here is guided by the extent to which a local boom affects the sector's product demand, and thereby its change in the opportunity cost of innovation. Firms in highly traded sectors mainly sell at fixed prices outside the local commuting zone and thus hardly benefit from higher wages

[1] Moreover, since booms are typically temporary, firms may not update their expectations regarding the future state of the oil and gas sector, and thereby opt against raising innovation.

and demand during the local boom, so their opportunity cost of innovation is largely unaffected. Meanwhile, the boom may raise their innovative capacity, for instance due to positive agglomeration economies (Moretti, 2021).[2] In line with this reasoning, we find a statistically significant increase in patenting by highly-traded goods producers during local booms. The picture looks different for producers of relatively lowly-traded goods. These *are* able to raise prices upon higher aggregate demand since they face low import competition, allowing them to raise profits by raising production. Such producers thus face higher opportunity costs of innovation during local economic upturns; but on the other hand, the factors that make highly traded goods producers innovate more during booms may equally apply to lowly-traded sectors. Consistent with these mixed predictions, we find no statistically significant change in patenting by lowly-traded goods producers during local oil and gas booms.

To better understand why local innovation is procyclical overall, and to challenge the narrative of the previous paragraph, we investigate various mechanisms. This analysis focuses on patenting in non-oil and gas technologies. We start by testing for a financial channel, following a literature showing that relaxed credit constraints can help explain procyclical innovation (Ouyang, 2011; Aghion et al., 2012; Nanda and Nicholas, 2014). Using the method of Rajan and Zingales (1998) to distinguish more versus less financially constrained industries, we do not find heterogeneous innovation responses. However, we do find that within a technology class, listed firms – which are typically less financially constrained – raise patenting by less than other firms during booms, although the innovation response is positive and similar in magnitude for both firm types. These results indicate that financial constraints may play a role, but cannot fully explain our results. Second, we investigate whether the rise in non-oil and gas innovation can be attributed to input-output linkages, given that the expansion of the oil and gas sector during booms may represent a positive market size effect for local input suppliers. Classifying industries as either upstream or downstream to oil and gas and evaluating heterogeneous effects, we find that input-output linkages are unlikely to drive our results.

Our findings do reveal two mechanisms that contribute to explaining our results. First, we

[2] 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 effects are unlikely to occur in the market for inventors, given our evidence of reduced innovation in oil and gas technology during booms.

find that local oil and gas booms raise the number of college graduates and creative class workers in a commuting zone, which is consistent with positive agglomeration economies that raise inventor productivity. Second, we show that innovation in technologies more closely related to oil and gas experiences a stronger response during boom periods compared to less related technologies. This indicates a redirection of innovation from oil and gas to related technologies during boom times.[3]

Contribution to the literature

This paper contributes to several literatures. First, we add to a literature on local innovation. This body of work has studied topics such as the impact of socio-economic conditions on local patenting (Crescenzi and Rodríguez-Pose, 2013; Hasan et al., 2020), local innovation spillovers (Jaffe et al., 1993; Audretsch and Feldman, 1996; Matray, 2021), the importance of spatial proximity (Roche, 2020; Xiao et al., 2021), the effects of place-based policies (Glaeser and Hausman, 2020; Tian and Xu, 2022), or the benefits of local innovation in terms of regional economic development (Moretti and Wilson, 2014; 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 solve these issues and thereby contribute to the literature by showing that innovation responds very locally to economic booms, by highlighting heterogeneous effects across different types of regions and sectors, and by identifying mechanisms.

Second, we add to a small literature which has tried to reconcile the diverging results on the cyclicality of R&D spending and innovation across theory and practice. Aghion et al. (2012) demonstrate the role of relaxed financial constraints during booms; Bernstein et al. (2021) show that inventors become less productive during recessions due to negative household wealth shocks; and Manso et al. (2023) show that during recessions, firms take more risk by patenting in technologies they are less familiar with, which they attribute to lower opportunity costs. Assuming that inventors take their R&D projects to market very quickly, Barlevy (2007) the-

[3] This result raises the question of whether the stronger effects for highly traded goods sectors merely reflects that these sectors tend to patent in technologies that are more closely related to oil and gas. We empirically dispel this concern in Section 4.6.

orizes that R&D – and its ensuing commercialization – are less profitable during recessions, and benefits competitors during the next boom. Fabrizio and Tsolmon (2014) show consistent empirical evidence that R&D and patenting are more procyclical in industries with faster obsolescence or a larger threat of imitation. Overall, the above studies thus typically explain procyclical innovation or R&D by introducing an additional factor that may outweigh higher opportunity costs, and test for the factor's relevance by exploiting its heterogeneity across firms or industries. We contribute to resolving the conundrum by instead exploiting heterogeneity across the opportunity cost of innovation across different industries within a given boom and locality, thereby testing for the opportunity cost theory more directly. Our boom measure is also less prone to endogeneity compared to a majority of the above studies which proxies industry cycles via aggregate industry output, thereby prohibiting causal inference (Manso et al., 2023). While Barlevy (2007) asks, "is the simple opportunity cost model inappropriate when it comes to R&D?", our results on patented innovations provide plausibly causal evidence against this hypothesis.

Third, we add to the ongoing discussion about the relationship between natural resources and economic development, which has been studied extensively both theoretically (Corden and Neary, 1982; van Wijnbergen, 1984; Mehlum et al., 2006) and empirically (e.g. Auty, 1993; Sachs and Warner, 2001; Sala-i Martin and Subramanian, 2013; Aragón and Rud, 2013; Caselli and Michaels, 2013; Allcott and Keniston, 2018; De Haas and Poelhekke, 2019; Pelzl and Poelhekke, 2021). This body of work 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 non-resource patenting, which speaks against a "resource curse" in innovation.

The remainder of the paper is structured as follows. Section 2 outlines our empirical strategy, Section 3 discusses data sources, Section 4 presents our results, and Section 5 concludes.

2 Identification

We set out to estimate plausibly causal effects of local business cycles on innovation at the local level over 1969-2012. Our empirical strategy is to exploit aggregate oil and gas shocks as exogenous source of local business cycles. The idea is that these aggregate 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. This approach is implemented via a “shift-share instrument” in the spirit of Bartik (1991). Following Allcott and Keniston (2018), we define exogenous local oil and gas booms as the interaction between cross-sectional variation in initial (1960) oil and gas reserves at the local level (“share”) and time series variation in oil and gas employment at the national level (“shift”). National oil and gas employment varies greatly across our sample period, reflecting the booms of the late 1970s and the 2000s and a bust period in-between (see Figure 1).[4] The endowment measure is largely a function of geology, especially because it includes “undiscovered” reserves, which are unrelated to exploration efforts (see Section 3 for details). In line, we find that initial oil and gas reserves do not correlate with other commuting zone characteristics (see Table A3 in the Appendix). This greatly eases identification concerns, given that the recent shift-share literature has found that exogeneity of the shares (conditional on controls) is sufficient to establish causality. We further discuss identification assumptions and how we deal with them after presenting our empirical specification next.

Since our main dependent variables are local patent counts, which potentially includes zeros, we use a Poisson specification. We estimate the following model,[5] where the geographic unit of observation is a commuting zone:

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

$Y_{c,\tau}$ equals the total number of granted patents in commuting zone c and period τ. Depending on the specification, “total” refers to either the universe of patents or a specific subset. τ

4 A potential alternative shift variable would be global oil and gas prices. However, this is not as good a proxy of US oil and gas booms because declining natural gas prices towards the end of our sample period (see Figure 2) do not indicate a gas bust, but rather reflect increased shale gas supply during the fracking boom. That said, Figure 1 shows that national oil and gas employment and the *oil* price follow very similar trajectories during our sample period, and our results are robust to using the oil price as a shift variable (see Table 8, column 6).

5 For estimate purposes, we use the ppmlhdfe command from (Correia et al., 2020), which makes it possible to estimate Poisson regression models with high-dimensional fixed effects.

stands for three-year periods, given that innovation takes time and the full impact of booms on patenting would be observed with some lag.[6] *National Oil&Gas Employment*$_\tau$ is computed as the average over the three-year period. *Initial Oil&Gas Reserves*$_c$ equals oil and gas endowment as of 1960 divided by commuting zone area, to account for its size. Following Allcott and Keniston (2018), from 2001 onwards, the measure includes endowment that is recoverable using hydraulic fracturing ("fracking") techniques.[7] For ease of interpretation of $\hat{\beta}_1$, we scale initial oil and gas reserves by the standard deviation of (pre-fracking) endowment, which equals \$4.5 million per square mile (see Table A1). $\gamma_{s,\tau}$ are state times (three-year) period fixed effects[8], and $\delta_{c,T}$ are commuting zone times period T fixed effects, where T={1969-2000 ; 2001-2012}. The inclusion of $\delta_{c,T}$ implies intuitively that we demean the equation separately, once for 1969-2000 and once for 2001-2012. This allows us to estimate one regression over the whole sample period rather than splitting the sample, and yet isolates the resource endowment that is relevant for a particular time frame (1969-2000 versus 2001-2012).[9] $\delta_{c,T} * \tau$ stands for commuting zone-specific linear trends, where for each commuting zone we estimate one linear slope for the pre-fracking period and one for the fracking period.

β_1 indicates the effect of local oil and gas booms on a commuting zone's patent count *relative* to commuting zones that experience a smaller or no oil and gas boom. β_1 is an unbiased estimator of this effect if, conditional on our rich fixed effect structure, the interaction term does not correlate with unobserved economic trends. While our long sample period covers multiple national oil and gas booms such as during the 1970s and 2000s, and national busts such as during the 1980s and 1990s, national oil and gas employment correlates with national GDP, and potentially also other variables. This per se is not a threat to identification: only if such

[6] Taking three-year averages also addresses the issue of serial correlation in the errors, since the averaging over periods ignores time-series information (see Bertrand et al., 2004). Note that since we update reserves starting from 2001, the last pre-fracking period consists of two years, 1999 and 2000.

[7] Our results are robust to using either of the two endowment measures (early endowment versus total endowment including fracking reserves, both as of 1960) for the entire sample period (see Table 8, columns 4-5).

[8] 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 its population in 1969.

[9] Demeaning across the whole sample period would mean that in every 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 larger fracking reserves starting from 2001, this would imply that for the pre-fracking period, we subtract a value that is influenced by a later change in endowment that is completely irrelevant at the time.

national-level variables affect patenting in commuting zones with more oil and gas endowment differently than patenting in other commuting zones in the same state, then β_1 is biased. While this appears unlikely, one might be concerned that oil and gas endowment is correlated with other local characteristics such as average income per capita, and changes in national GDP might in turn affect richer commuting zones differently than others. Therefore, as a first step, we evaluate the correlation of initial oil and gas endowment with multiple commuting zone characteristics. The results (see Table A3) show that endowment is uncorrelated with all included variables. This is supportive evidence that the shares in our shift-share design are exogenous. Nonetheless, we address the described identification concern in yet another way. Following Goldsmith-Pinkham et al. (2020), the idea is that to ensure a causal interpretation of the estimated coefficient $\hat{\beta}_1$, it is sufficient that the shares are exogenous conditional on controls. Therefore, we add interactions of national oil and gas employment with all commuting zone characteristics used in Table A3 to Equation (1), although our evidence indicates that they are not (statistically significantly) correlated with oil and gas endowment. The results are very robust to this alternative specification (see Table 8, column 2).

Table A2 shows that our measure of local oil and gas booms has a statistically significant and positive impact on local population, employment, earnings per worker, GDP, and local government revenue. This can be interpreted as strong evidence on a conceptual first stage of our reduced-form approach. Note that we do not apply a two-stage least squares IV estimator because the boom effect likely operates via multiple of the above variables, rather than only via population, for instance. As a consequence, the exclusion restriction would likely be violated in any type of IV setup.

3 Data

In this section we describe our sample and the data on oil and gas endowment and patenting. Details and other data sources are described as they become relevant throughout the text, and/or in the Online Data Appendix (see Section OA3).

Our sample covers the lower 48 states. We start our analysis in 1969 since data on most variables are not available for earlier years, and we end in 2012 due to patent data availability.

We aggregate all data from the county to the commuting zone level, mainly because commuting zones better represent local labor markets.[10] Several counties split over our sample period, and the Bureau of Economic Analysis (BEA) merges selected counties in its regional data reporting. We address these data issues by defining commuting zones that are consistent over 1969-2012. There are 759 such commuting zones in our sample. Table A1 reports descriptive statistics.

3.1 Oil and gas endowment

We obtain county-level data on economically recoverable oil and gas reserves as of 1960 from Allcott and Keniston (2018), which we aggregate to the commuting zone level. The data are not publicly available but have been generously shared by the authors for this project. "Economically recoverable" depends on the available extraction technology. In the early 2000s, the combination of horizontal drilling with hydrofracturing ("fracking"), as pioneered by Mitchell Energy in Texas' Barnett Shale (Hinton, 2012), made large amounts of existing reserves economically recoverable. Following Allcott and Keniston (2018), we account for this by defining two distinct endowment measures: pre-fracking endowment equals reserves as of 1960 that are economically recoverable using conventional extraction methods, and total endowment equals 1960 reserves that are economically recoverable using conventional or fracking techniques. In our empirical analysis, endowment equals pre-fracking reserves until the year 2000 and total reserves from 2001 onwards (see Section 2).

There are no county-level estimates of oil and gas endowment reaching back to the 1960s. Therefore, reserves are simply computed as the sum of remaining reserves in year T>1960 and total production between 1960 and T, where T=1995 for pre-fracking endowment and T=2011 for total endowment. Dividing by a commuting zone's area (including both land and water area) to account for its size, the endowment measure looks as follows[11]:

[10] Feyrer et al. (2017) show that the wider regional impact of fracking booms on jobs and income is three times as large as the immediate county effect, which suggests that a county is too small a unit to capture the full effect of oil and gas booms on local innovation.

[11] Oil and gas production data come from a novel dataset that relies mostly on information from private data provider DrillingInfo; proven reserves data are from the non-publicly available 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. See Allcott and Keniston (2018) for more details.

$$Oil\&Gas\ Reserves_{c,1960} = \frac{Proven\ Reserves_{c,T} + Undisc.\ Res._{c,T} + \sum_{t=1960}^{T} Production_{c,T}}{Area\ in\ Square\ Miles_c}$$

As shown by the formula, Allcott and Keniston (2018) include undiscovered reserves in their endowment measure, which represents a novelty compared to previous literature. Undiscovered reserves are "postulated from geologic knowledge and theory to exist outside of known fields" (Schmoker and Klett, 1999, p.1). Their inclusion makes the reserves measure more closely related to geology and thus more exogenous to economic outcomes.

In order to add up oil and gas units and ease the interpretation of our results, we follow Allcott and Keniston (2018) and transform physical measures to dollar values using average (real 2010) prices over 1960-2011: \$34.92 per barrel of oil and \$3.2 per million British thermal units (MMBtu) of natural gas. This step also allows us to quantify the relative importance of oil versus gas in national oil and gas endowment: oil accounts for 42% of total endowment (including fracking reserves), while gas accounts for 58%.

595 of the 759 commuting zones in our sample have nonzero pre-fracking endowment ($R_c^{early} > 0$), and 613 commuting zones have nonzero total endowment ($R_c^{total} > 0$). These statistics reflect the wide geographic spread of oil and gas endowment across the country. The average pre-fracking endowment across all commuting zones equals \$1.5 million per square mile, and the standard deviation equals \$4.6 million; for R_c^{total}, the average and standard deviation equal \$2.9 and \$7 million, respectively. At the national level, undiscovered reserves make up 35% of R^{total}. 153 commuting zones have undiscovered oil reserves but do not produce oil over 1960-2011, and 255 commuting zones have undiscovered gas reserves yet no gas production. These statistics illustrate that the endowment measure does not simply include producing commuting zones, which mitigates identification concerns. As we motivate in Section 2, our measure of local oil and gas booms is the interaction of initial commuting zone-level oil and gas endowment (we use R_c^{early} until 2000 and R_c^{total} thereafter) with national oil and gas employment over time, which we obtain from the Bureau of Economic Analysis (BEA).

3.2 Patents

We measure innovation using patent data. Patents are typically highly correlated with R&D and other indirect measures of innovation (Griliches, 1990). Our main data source is the European Patent Office's (EPO) Worldwide Patent Statistical Database (PATSTAT).[12] PATSTAT contains the population of all patents filed globally since the mid-1960s. For each application, PATSTAT collects a wide range of information, including bibliographic information, technology fields, family links, citations, etc. We merge these data with information on disambiguated inventors' addresses from PatentsView (1976-2012) and HistPat (1969-1975).

We leverage these rich data to construct a novel panel data set of patenting by technology field and geographic unit over more than four decades. We measure the innovative activity in a technology class j in commuting zone c and year t as the number of granted patents filed in year t and class j by inventors located in commuting zone c. We proceed as follows. First, we identify the research team behind a patent and the inventors' location at the time of patent filing. This allows us to assemble the population of all inventors residing in the United States, compute their innovation output in the period of analysis, and flexibly map it to the geographic unit of interest.[13] Second, we observe the filing and granting dates for each patent office where a patent is submitted. These data enable us to date patents based on their earliest filing date ("priority date"), which most closely approximates when the innovation project was conducted. Third, we observe technology fields, which are categorized by the patent office based on the patent's technical characteristics following the Cooperative Patent Classification (CPC) scheme.[14]

In our analysis, a patent corresponds to a unique invention. This means that if the same invention is patented in multiple countries, it is counted only once. When a patent comprises several technology codes, we count the patent fractionally, with a weight proportional to the frequency of each technology code. Similarly, if inventors are located in more than one com-

[12] We use the 2018 version.

[13] The inventor address reported on a patent document is usually the professional address of the inventor and therefore is a better indication of where the innovation takes place than the address of the applicant, who may not be located where the R&D takes place (OECD, 2009).

[14] In each version of PATSTAT, the most recent and detailed CIPC technical codes are assigned to all previously filed patents. This enables us to consistently track patent technologies over time.

muting zone, we count the patent fractionally.[15] 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 US, and non-missing CPC technology codes. While the literature has discussed extensively the advantages and disadvantages of using patent data,[16], it is the only source of information that enables an analysis of innovation by technology field and at the local level over our long sample period.

Following the Derwent World Patents Index (DWPI; Clarivate Analytics, 2020), we classify the technology classes C10G, C10L, C10M, and E21B as oil and gas classes. This classification is frequently used in the academic literature (see e.g. Duch-Brown and Costa-Campi, 2015) and yields total oil and gas patent counts that are comparable to those reported in non-academic outlets.[17] We define a patent as oil and gas patent if it contains at least one of the oil and gas technology codes. Oil and gas patents make up around 2% of all patents on average over our sample period.

Patent quality is known to be highly heterogeneous. We address this in two ways. First, our sample consists of granted patents at the USPTO, which are considered highly valuable inventions in the empirical innovation literature. Second, we use citations as a proxy for patent quality because highly valuable inventions are more extensively cited than low value patents (Harhoff et al., 1999).

The annual number of patents varies greatly across space, with 5% of commuting zones being responsible for almost three quarters of patenting throughout our sample period. The number of patents in oil and gas rich commuting zones (which we define as those with above-median reserves across all 759 commuting zones) over 1969-2012 averages to 76 per year, while in oil and gas poor commuting zones the average equals 71. The medians are much smaller, being equal to 4.1 and 4.5, respectively. Patents *per capita* (evaluated as patents per 100,000 persons) average to around 9.1 (median = 5.7) and 9.5 (median = 6.1) across oil and gas rich and oil

[15] For example, suppose that the patent office assigns technology codes 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 county A and one residing in county B; then we have 0.5 $\times$0.75 patents in technology A01B and 0.5 $\times$0.75 in technology A01C in county A, and 0.5 $\times$0.25 patents in technology A01B and 0.5 $\times$0.25 patents in technology A01C in county B.

[16] See for example Griliches (1990), OECD (2009), and Nagaoka et al. (2010).

[17] For example, Reuters reports 2,188 oil and gas patents filed in the US 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.

and gas poor commuting zones, respectively.

4 Results

We structure the presentation of our results as follows. First, we briefly discuss the impact of oil and gas booms on measures of local economic activity, in order to empirically justify the chosen boom measure (see Table A2). In Section 4.1, we then present the effects on total patent activity at the local level, as well as patenting in oil and gas versus non-oil and gas technologies (see Table 1). After discussing mechanisms behind our results on oil and gas patenting, we devote the remainder of Section 4 to studying non-oil and gas patenting in more depth. Sections 4.2 and 4.3 document heterogeneous effects across different types of commuting zones and across booms and busts, respectively (Table 2). In Section 4.4, we explore a wide range of mechanisms that may explain the rise in non-oil and gas patenting during boom times (Tables 3-5). In Section 4.5, we show evidence of path dependence in innovation (Table 5). The conclusions we derive from Section 4.4 help guide our analysis of patenting by highly-traded versus lowly-traded goods producers (Section 4.6, Table 6), whose opportunity cost of innovation is differently affected by the local boom. Finally, Section 4.7 discusses boom effects on innovation quality and green versus non-green innovation (Table 7), and Section 4.8 presents a wide range of robustness checks (Table 8).

Oil and gas booms correlate with measures of local economic activity

Table A2 shows that local oil and gas booms significantly raise the level of economic activity in a commuting zone. In particular, in the commuting zone with an initial oil and gas endowment equal to one standard deviation (approximately five million dollars per square mile), a doubling of 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%, and GDP by 5.4%. Data for these variables are obtained from the Regional Economic Accounts database provided by the Bureau of Economic Analysis (BEA). Moreover, column 5 shows a 6.3% rise in own-source local government revenue (thus excluding state and federal transfers, and including property taxes, for instance) during booms. We obtain these five-yearly data from the Census of Governments, aggregate the data across all county governments in the commuting zone, and adjust Equation (1) to account for the different data frequency. These results show that our shift-share interaction term is a very good proxy for the local business cycle.

4.1 Main Results

Table 1 reports the estimated effect of oil and gas booms on local innovation. Column 1 shows that in a commuting zone with an oil and gas endowment of five million dollars per square mile, a doubling of national oil and gas employment leads to an increase in the total patent count by around 8.3%. Column 3 shows a significant increase of similar magnitude for non-oil and gas patents. Table OA1 in the Online Appendix shows the effect on non-oil and gas patenting by two-digit technology class. In column 5 of Table 1 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.[18] Table 8 shows that these key results are robust to a wide range of modifications to the baseline specification, such as using the oil price as shift variable or accounting for recent work in the shift-share literature.

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 also analyze whether the above results differ across the pre-fracking and the fracking period. To this end, in columns 2, 4 and 6 of Table 1, we include an interaction of the boom variable with a dummy that equals one starting from the three-year period 2001-2003. The coefficient signs suggest that if anything, the documented effects become stronger after 2000, but the interaction terms are not statistically significant.

Reasons for countercyclical oil and gas innovation

The result of countercyclical oil and gas innovation may be explained by at least two factors. First and foremost, the oil and gas sector's opportunity cost of innovation is higher during boom periods. Figures 1 and 2 show that national oil and gas employment closely co-moves

[18] Note that the number of observations is smaller in column 5. 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.

with the oil price, which implies higher profits for oil producers during boom times. For natural gas producers, the same holds true before 2000; thereafter, the rise in national employment is mainly shaped by a reduction in production costs due to the fracking revolution. This implies that throughout our sample period, times of high national oil and gas employment are times of high profits for both oil and gas producers, leading firms to prioritize extraction and sales (as evidenced by a rise in local oil and gas activity during boom times, see Online Appendix Table OA2) rather than innovation. In this sense, our results provide strong empirical evidence of the opportunity cost theory: oil and gas, facing the most apparent rise in the opportunity cost of innovation, reduces patenting during booms, while other sectors that are less central to the boom increase patenting.

A second reason why oil and gas innovation is not procyclical may be that the incentives to innovate do not depend on the current state of the oil and gas sector – such as current oil prices – but on the expected future state of the industry. If firms consider booms to be (very) temporary, they may not adjust their expectations, in which case innovation is unlikely to increase.

4.2 Heterogeneity across commuting zone characteristics

An important question is whether our results are driven by a particular type of commuting zone, for instance large metropolitan commuting zones or those with a high patent activity. We study such sources of heterogeneity in Table 2 by either restricting our sample or adding interaction terms with our boom variable, and focus on innovation in non-oil and gas technologies as outcome variable. In column 1, we repeat the baseline results from column 3 of Table 1. In column 2, we restrict the sample to commuting zones with an average of less than 20 patents (of any kind) per three-year period over 1969-2012, which is true for 60% of commuting zones. The results are very similar for this restricted sample, which indicates 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. For this purpose we use the Rural-Urban Continuum Codes published by the US government's *Economic Research Service* (ERS).[19] The results show that compared to metropolitan commuting zones, non-oil and gas patenting in non-metropolitan commuting zones rises by about twice as much during local economic booms. In column 4, we distinguish non-metropolitan commuting zones into more versus less urban ones.[20] We observe that the effects are in fact largest overall for commuting zones that are non-metropolitan but relatively urban. In column 5, starting from the specification of column 4 we add interaction terms with several other variables: initial 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, as of 2018.[21] In order to interpret the non-interacted boom coefficient as the effect in the average commuting zone across the five sources of heterogeneity we test for, we demean the urban-rural dummies and the three additional variables before performing the regression. The results show that commuting zones with a higher initial patent activity experience a significantly larger rise in patenting during boom times. This suggests that booms reinforce existing innovation capacity rather than create it from scratch. The coefficients on the human capital and college density interactions are not statistically significant.[22] In column 6, we test for heterogeneous effects within the sample of urban non-metropolitan commuting

[19] 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". 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, based on the 1974 edition of the data (earlier data are unavailable). We define commuting zones with an average value of 4 or larger as non-metropolitan, which is true for 90% of commuting zones.

[20] We define a commuting zone as urban non-metropolitan if 4≤rural-urban code<8 (see also footnote 19). These 496 commuting zones produce around 80 patents per three-year period on average. Rural non-metropolitan commuting zones are those with a code≥8. These 190 commuting zones produce an average of around 6 patents.

[21] County-level education data, which we aggregate to the commuting zone level, are available from the ERS. Note that these data do not contain information on the number of residents who actually obtained a college degree. 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. Given that we have information on a college's number of employees, we are able to account for heterogeneity in college size across space, 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).

[22] While recent literature shows that local patenting is positively affected by local university presence (Andrews, 2023) and innovation (Hausman, 2022), our results suggest that local universities do not mediate the impact of local economic booms on patenting.

zones, given that these experience the largest overall effects. The coefficient on initial patenting remains unchanged in magnitude, but turns insignificant due to a larger standard error in this restricted sample. Interestingly, the coefficient on human capital is now positive, larger in size, and statistically significant, indicating that a well-educated population is key for innovation to respond to economic booms in urban non-metropolitan areas. Overall, the findings in columns 5 and 6 resonate with the proposal of Gruber and Johnson (2019) to increase research funding in local innovation clusters with sufficient human capital – although the 102 places the authors recommend "for jump-starting America" are mostly (but not exclusively) located in metropolitan commuting zones.

4.3 Booms versus busts

While a rise in local innovation during boom times is good news for a locality, the question arises whether these effects are reversed by a decline in innovation during bust periods. We test for this question by restricting our sample to a time frame that is characterized by a continuous decline in national oil and gas employment from one three-year period to the next: over 1981-2000, the variable halved from one million employees to a mere 510,000 (see also Figure 1). If the coefficient on our key interaction term remains positive and significant using this subsample, then this is evidence that oil and gas busts lead to a decline in local non-oil and gas patenting. However, the results (see Table 2, column 7) do not show this pattern: the coefficient is not statistically significant, and in fact negative. To scrutinize the conclusion that *booms* drive our results, in column 8 we restrict our sample to a time frame in which national oil and gas employment continually rose. This occurred over 2001-2012 during the fracking boom, when US oil and gas employment bounced back from 510,000 to 1.25 million. Although the sample period is shorter than in column 7 which negatively affects statistical power, we find a positive and statistically significant effect. Taken together, these effects suggest that the rise in innovation during local boom periods tends to be sustained, rather than only short-lived.

4.4 Mechanisms behind the procyclicality of non-oil&gas innovation

In this section, we examine underlying mechanisms that may contribute to the rise in non-oil and gas innovation during economic upswings. We evaluate the importance of relaxed financial constraints, input-output linkages, agglomeration effects, public finance, household wealth effects, strategic patent timing, and a redirection of local innovation from oil and gas to non-oil and gas technologies.

4.4.1 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). 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 exploit the widely-applied notion of Rajan and Zingales (1998) that the degree of financial constraints typically differs by industry. To do so, we first map patent technology classes to four-digit industries based on the Standard Industrial Classification (SIC) system. Mapping technologies to industries has been challenging. Existing techniques based on text similarities, such as Lybbert and Zolas (2014), have a limitation in distinguishing between producing and using industries. To address this limitation, we employ a two-step approach. First, we use the fact that for each patent in our dataset, we know whether it was filed by a firm included in the Compustat database, and if so, in which four-digit SIC industry the firm operates. By examining these patents filed by Compustat firms, which account for 34% of all patents in our sample, we can 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. The advantage of this method is that we can subsequently map all patents to industries, irrespective of whether the patent is filed by a Compustat firm or not. This allows us to map the complete set of patents in our analysis to the corresponding SIC industry. We then compute a measure of dependence on external finance for each of the 20 two-digit SIC manufacturing industries following the procedure of Rajan and Zingales (1998)[23], and correlate it with the industry's patent activity over the local business cycle. We do so by estimating Equation (1) at the commuting zone – industry level, where we replace commuting zone fixed effects with commuting zone times industry fixed effects, and further include industry times three year period fixed effects. The results (see Table 3, column 2) suggest that firms in more financially constrained industries 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 extent to which we can plausibly map technologies to industries, 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 those 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 oil and gas booms, we estimate a specification at the commuting zone – technology class level, with commuting zone times technology class fixed effects and technology class times three year period fixed effects. The results (see Table 3, column 4) indicate that patenting by Compustat firms rises by significantly less in boom times, but the overall effect on these firms is still positive and similar in size to the baseline coefficient (see column 3). These findings suggest that relaxed financial constraints during booms may play a role in explaining our results, but are unlikely to fully explain them.[24]

4.4.2 Input-output linkages

Next, we study whether the increase in local innovation in non-oil and gas technologies is driven by input-output linkages. We do so mainly by testing whether patents in industries that

[23] 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, and then define the industry's realization as the median observation across all firms in the industry.

[24] Note that the results in columns 1 and 3 of Table 3 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 columns 1 and 3 include sector-period and technology class-period fixed effects, respectively, and their results confirm the baseline result of column 3 of Table 1, such concerns are invalidated.

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), and the expansion of the oil and gas sector during booms represents a market size effect for input suppliers if local oil and gas producers increase their demand for intermediate inputs.

Using the Input-Output tables from the Bureau of Economic Analysis for the year 1987, we first calculate the direct and indirect share of an industry's output sold to the oil and gas sector[25]. Following Allcott and Keniston (2018), an industry is defined as "upstream" if the sum of direct and indirect oil and gas output share is larger than 0.1%, and "downstream" if the direct input share exceeds 0.1%.[26] Industries that do not fall into either category are classified as "non-linked". Subsequently, we aggregate non-oil and gas patents into patents by upstream, downstream, and non-linked industries and use these are left-hand side variables in Equation 1.

The results presented in columns 5-7 of Table 3 suggest that the input-output linkages channel is unlikely to account for the rise in local non-oil and gas innovation. While we find a positive and significant elasticity of innovation in upstream industries to oil and gas shocks, the elasticity is even larger in non-linked industries. Furthermore, the effect on innovation in downstream industries is small and imprecisely estimated.[27]

4.4.3 Agglomeration effects

The above results indicate that weakened financial constraints and input-output linkages cannot (fully) explain the rise in non-oil and gas innovation during local booms. Moreover, in Section 4.4.5 we rule out public finance, household wealth effects, and strategic patent timing as important drivers of our results. This evidence, and a literature highlighting the importance of agglomeration for innovation (Carlino and Kerr, 2015; Moretti, 2021), suggest the presence of agglomeration economies as a possible mechanism behind our results. Perhaps inventors move to the booming commuting zone and patent there, and/or their move makes incumbent

[25] 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.

[26] Including indirect shares in the downstream measure computation would be overly affected by the electricity intensity of a sector.

[27] We cannot reasonably evaluate the effects on industries that are downstream and not upstream, because all but two downstream industries are also upstream.

inventors more productive; or, more generally, high-skilled workers in creative occupations immigrate and promote the local creation of new ideas. While it is impossible to identify inventors' movements unless they file a patent, we can test the relevance of the high-skilled workers hypothesis by analyzing immigration by type of worker. We start by studying the impact of local oil and gas booms on the number of residents with and without a college degree. Data on county-level educational attainment are only available from the years 1990 and 2000 via the Population Census, and thereafter as moving five-year averages via the annual American Community Survey (ACS), starting with 2006-2010.[28] On the left-hand side, we use these three data points and the 2011-15 ACS average, and aggregate to the commuting zone level; on the right-hand side, we evaluate our key interaction term in 1990, 2000, as average over 2006-10, and as 2011-15 average. The results (see Table 4) show that oil and gas booms raise not only total adult population (column 1), but also both college- and non-college educated population (columns 2-3). The coefficient is in fact larger for college-educated population, but the difference to the non-college coefficient is not statistically significant.

Given the low data frequency, one might be concerned that the rise in college graduates during boom times does not reflect migration, but educational choices of the native population. However, column 4 of Table 4 shows that total population does respond (positively) to oil and gas booms measured at annual frequency; and Cascio and Narayan (2022) show that oil and gas booms typically have a *negative* impact on residents' schooling, both of which speak against this alternative hypothesis.

One limitation of studying population by educational attainment is that college graduates may not necessarily work in creative occupations that are close to the local innovation process. We therefore perform an additional test in which we analyze the number of 'creative class workers' at the local level over time. While this concept was first introduced and defined by Florida (2002), we use an improved classification by the ERS, which excludes from the original measure "many occupations with low creativity requirements and those involved primarily in services to the residential community". The list of included occupations features for instance Computer and Mathematical, Architecture and Engineering, or Life, Physical, and Social Science

[28] 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 (compare e.g. Weber, 2014).

occupations, but also Management occupations. The ERS provide county-level data for 1990, 2000, and the average over the 2007-11 ACS rounds, which we aggregate to the commuting zone level. Averaging the creative class share across all available data points reveals that 17% of workers belong to the creative class (median=16%). The results in column 5 of Table 4 show that oil and gas booms lead to a statistically significant rise in the local number of creative class workers. This suggests that during booms, there is a larger amount of people working in occupations that are relatively close to local innovation processes, which implies agglomeration economies that are fruitful to local innovation and likely help explain our results. Also inventors in particular are well-represented in the creative class definition we use, based on evidence by Akcigit and Goldschlag (2023) that 50% of inventors are in the technical occupations listed further above, and 26% are in managerial occupations.

4.4.4 Innovation redirected from the oil and gas sector to related technologies

As innovation in oil and gas decreases during boom times, inventors from the oil and gas industry may transition to firms working on non-oil and gas technologies. Although oil and gas patenting only accounts for a small share of total patenting, this could partly explain the rise in non-oil and gas innovation during boom times. If such a mechanism plays a role, non-oil and gas innovation should increase relatively more in technologies that are closely related to the oil and gas sector. We can test for this by building a measure of technological relatedness to oil and gas based on technology co-classification information contained in patent documents, inspired by previous literature (e.g. Kogler et al., 2013). Intuitively, the more often a certain non-oil and gas technology class is listed together with an oil and gas technology, the closer it is to oil and gas in the technology space.[29] The result is a value between zero and one for each two-digit technology class indicating its relatedness to oil and gas technology. We then use this measure in a specification that features, as dependent variable, the number of non-oil and gas patents in

[29] We define the relatedness of technology class k to oil and gas technologies as follows:

$$R_k^{OG} = \frac{\sum_{t=1969}^{t=2013} \omega_{ik} p_{ik}^{OG}}{P^{OG}}$$

where ω_{ik} denotes the relative importance of technology class k in patent i, $p_{ik}^{OG} = 1$ if patent i lists technology class k and is an oil and gas patent, and P^{OG} denotes the total number of oil and gas patents filed and granted globally in the sample period. k denotes a 2-digit IPC code.

a given two-digit technology class in a given commuting zone in a given three-year period, $y_{kc,\tau}$. In particular, in column 1 of Table 5, we estimate our baseline specification at the two-digit IPC class by commuting zone level, augmented by an interaction of our standard boom variable with the relatedness measure. Given our focus on non-oil and gas innovation, we remove oil and gas patents from an IPC class' total patent count whenever applicable. The specification contains two-digit IPC class times commuting zone times century fixed effects, state times three-year period fixed effects, two-digit IPC class times three-year period fixed effects, and controls for differential commuting zone-specific patenting trends across the pre-shale gas period (1969-2000) and the shale gas period (2001-2012). In column 2, we also add an interaction of our boom measure and a variable indicating the general importance of the IPC class, measured as total patents in the IPC class over 1969-2012 divided by all patents in 1969-2012. This addition controls for the possibility that technologies that are close to oil and gas are generally more (or less) common, and, during booms, firms patent disproportionally more (or less) in common technology classes. The results show a positive and statistically significant (at the 10% level) coefficient on the relatedness interaction, which is consistent with innovation being redirected from oil and gas to non-oil and gas technologies.[30] That said, the non-interacted coefficient is also positive and significant, indicating that non-oil and gas patenting also rises in technology classes that are unrelated to oil and gas.

4.4.5 Public finance, household wealth effects, and strategic patent timing

In this subsection we discuss a number of other potential channels which turn out to be unlikely drivers of our results.

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 A2 shows that total own-source revenue of county governments rises during oil and gas booms.[31] This raises the question on the existence of a public finance

[30] In line with this evidence, we find that out of all nine two-digit technology classes that experience a statistically significant rise in patenting during booms (see Online Appendix Table OA1), only one is in the least related tercile (which constitutes only 1% of the relatedness mass; see Figure OA1).

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

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 increases (Newell and Raimi, 2015, 2018). Nonetheless, we test the public finance hypothesis by exploiting the variation in counties' oil and gas revenue participation across states. The results (see Section OA1.3) show that commuting zones located in states with a larger county government "take" on oil and gas revenue do not raise patenting by more than others, which speaks against a public finance channel.

Local oil and gas booms may increase incumbent inventors' wealth, for instance via raising house prices. In turn, inventor wealth has been documented to affect productivity: Bernstein et al. (2021) show that inventors who experience a negative housing wealth shock during the Great Recession produce fewer and less important patents thereafter, and attribute this to resource constraints and/or psychological or financial distress. Since our results are driven by positive boom effects rather than negative bust effects (see Section 4.3), the question arises whether a *reduction* in distress during booms can raise inventors' productivity. Bernstein et al. (2021) find evidence against this: repeating their analysis during the housing boom period between 2002 and 2007, they find no impact of house price increases on inventors' patent output. Moreover, in fact we do not find strong evidence that local oil and gas booms raise local house prices in the first place (see Section OA1.3). These findings suggest that wealth effects are unlikely to explain the rise in non-oil and gas innovation during boom times.

Alternatively, 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 4.6), which speaks against this channel.

4.5 Path dependence

The rise in non-oil and gas patenting may reflect innovation in novel technological areas for a commuting zone or areas in which the commuting zone already has prior experience. To examine this, we calculate commuting zone c's relative historical experience in each two-digit technology class as the ratio of patents in the two-digit IPC code to total patents in the period before our analysis (1960-1969). We then proceed by estimating Equation 1 at the two-digit IPC class by commuting zone level, including an interaction with the boom variable and the technology's share in total patents in the pre-period. The findings are presented in column 3 of Table 5. In column 4, we additionally control for differential patenting in more or less common technology classes during boom times. The findings indicate that the rise in non-oil and gas patenting is not exclusively driven by technology classes the commuting zone is historically familiar with. However, these familiar technology classes do experience a larger rise in patenting in percentage terms. We interpret this as evidence for path dependence of innovation, consistent with Aghion et al. (2016) and Manso et al. (2023).

4.6 Different sectors, different opportunity costs of innovation

In Section 4.1, we showed that innovation in oil and gas – a sector that clearly experiences a rise in the opportunity cost of innovation during local oil and gas booms – is countercyclical. In contrast, firms in other sectors, which (at least on average) experience a smaller rise in the opportunity cost of innovation during local booms, raise innovation. In this section we provide an additional test of the opportunity cost theory, by studying heterogeneous effects on non-oil and gas innovation across highly- and lowly-traded goods producers. We do so because of these sectors' varying exposure to higher local demand during the boom, and thus their different change in the opportunity cost of innovation. Highly traded goods producers mainly sell outside of the local market, which implies a low or zero change to their product demand, and thus at best a small increase in their opportunity cost of innovation. In contrast, lowly traded goods producers can raise prices and sales upon higher local demand, which implies higher opportunity costs of innovation during the boom. We classify a four-digit SIC sector in manufacturing as relatively highly traded if it has a below median distance elasticity. The latter equals the percentage change in trade volume as distance increases by one percent, as calculated by Holmes and Stevens (2014) using industries' average shipment distance reported in the *US Commodity Flow Survey*. 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 industries into highly- and lowly-traded, we map industries to technology classes as described above, and allocate patents into the two categories of tradedness. The results (see Table 6, column 1) 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 due to agglomeration economies (see the results of Table 4, and consistent evidence by Moretti, 2021). In contrast, and consistent with theory, highly traded goods producers significantly raise patenting in boom times, at similar magnitude as our baseline effect (column 2).

As a robustness check, we divide industries into tradedness terciles. If the results in columns 1-2 can indeed be explained by varying exposure to local aggregate demand, then we should see the weakest impact for the least traded tercile, stronger effects for the intermediate tercile, and the strongest impact on the most traded tercile. This is exactly what we observe in columns 3-5 of Table 6.

While this result is reassuring, one might be concerned that the strongest effects on patenting by highly-traded goods producers merely reflects that these firms innovate in technological areas that are more related to oil and gas, given that these see a larger rise in patenting during booms (see Table 5). We test for this possibility by analyzing the correlation between the number of patents in highly traded sectors and the number of patents in oil and gas-related technologies, at the commuting zone level over time. We define 'highly traded patents' as those filed by industries in the highest tradedness tercile, and 'highly oil and gas-related patents' consistently as those filed in technology classes in the highest relatedness tercile. Given that both of these variables rise in oil and gas-endowed commuting zones during booms, estimating their relationship based on the full sample could easily indicate a spurious positive correlation. To test whether oil and gas-related patents are typically filed by firms in more traded sectors, it is therefore more constructive to use the sample of commuting zones without oil and gas reserves. Regressing highly traded sectors' patents on highly related patents and including the same fixed effects as in Equation (1) yields a negative and insignificant correlation (results are available upon request), which speaks against the described concern.

4.7 Innovation quality and green versus non-green innovation

4.7.1 Innovation quality

During boom periods, resource-rich regions that experience increased wealth may 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, as found by Makridis and McGuire (2023). 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 $q_{ic\tau}$ denote the number of forward citations five years after a patent i was filed. The cumulative sum in commuting zone c in period τ is:

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

where $P_{c\tau}$ is 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 $\bar{Q}_{c\tau} = Q_{c\tau}/P_{c\tau}$.

Our second proxy is a measure of generality. This measure 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 (5-year) citations from patents of technology class k to patent i, and $cit_i = \sum_k cit_{ik}$ denotes the total number of (5-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 various fields, whereas low generality occurs when citations are concentrated in a few specific fields. A high generality index suggests that the patent had a significant influence on future innovations across a wide range of 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.[32]

We proceed by using the average quality and generality of patents in each commuting zone and period as the dependent variable and estimate our baseline model again. Note that in this case, the sample includes all commuting-zone-period observations with at least one patent. This is 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 columns 2 and 3 of Table 7 and suggest that there is no significant decrease in the average quality or generality of patents. Hence, the increase in non-oil and gas innovation does not seem to be driven by the deterioration of innovation quality.

4.7.2 Green innovation

In this section, we examine the effect of oil and gas booms on green innovation. The underlying motivation is recent theoretical work suggesting that increased oil and gas activity during the fracking boom crowded out green innovation (Acemoglu et al., 2019). Green patents are defined as those including one or more technology classes capturing "climate change mitigation technologies" (C/IPC in the Y02 class), following previous literature (see e.g. Angelucci et al., 2018). We estimate equation (1) using the number of green patents and the share of green within total non-oil and gas patents as dependent variables. The findings, presented in columns 4 and 5 of Table 7, indicate a positive response of green innovation to oil and gas booms. However, there is a decline in the share of green patents. This aligns with the findings of Acemoglu et al. (2019), who observed a similar decline in green patents following the fracking revolution in the US. Notably, our results indicate that this decline begins earlier than the fracking revolution.

[32] Details on how we compute this measure at the local level are provided in Online Appendix Section OA2.

4.8 Robustness Checks

In Table 8 we carry out several robustness checks. Panel I reports results for non-oil and gas innovation, while Panel II reports equivalent robustness checks for oil and gas innovation. Column 1 repeats the baseline results of Table 1 as reference. In column 2 we add interactions of national oil and gas employment with commuting zone-level population density, personal income per capita, and population, respectively, all measured in 1969. This addresses the concern that initial oil and gas reserves per square mile may be correlated with these variables, thus making the shares in our shift-share design endogenous (see Goldsmith-Pinkham et al., 2020). In column 3 we add an interaction of 1960 coal reserves at the commuting zone level with national coal employment to control for the impact of local coal booms. In column 4 we use pre-fracking reserves as of 1960 for the entire sample period, while in column 5 we use endowment in 1960 including fracking reserves. In column 6 we show that the baseline results are robust to using the oil price rather than national oil and gas employment in our key interaction term. The magnitude of the coefficient is smaller, which arguably reflects that a doubling of the oil price is a smaller shock than a doubling of national oil and gas employment, as indicated by Figure 1. Overall, the results reported in Table 8 show that our baseline results are robust to a wide range of modifications to our baseline specification.

5 Conclusion

In this paper we asked several basic questions about the responsiveness of innovation to changes in the level of economic activity at the local geographic level. Exploiting nationwide oil and gas shocks as exogenous source of local business cycles in oil and gas-endowed commuting zones, we find novel evidence that patent activity responds very locally, and positively so, to economic booms. The results are strongest in non-metropolitan yet relatively urban areas. Importantly, we do not find that the effects are reversed by a decline in patenting during bust periods, suggesting that the effects tend to be permanent rather than only short-lived. In terms of mechanisms, we find that relaxed financial constraints and input-output linkages are unlikely to explain our findings. In contrast, we show evidence consistent with positive agglomeration economies that raise the productivity of local innovative firms.

Our findings inform the recent academic and policy debate on the spatial distribution of innovation across the US. Gruber and Johnson (2019) advocate a stronger geographic dispersion of research finding to local innovation clusters outside of large metropolitan areas, in order to revitalize US economic growth across geographies. Regarding the local consequences of such policies, Glaeser and Hausman (2020) argue that "[p]erhaps cases exist where a modest "big push" would propel a region into perpetual growth, but we do not know how to identify these cases in the United States today." (p.264). In that regard, our findings of procyclical local patenting could be seen as evidence that a local stimulus has the potential to promote local innovation permanently, and thereby potentially also local economic development.

As a second contribution, this paper highlights heterogeneous effects on local innovation across economic sectors, which help explain the puzzle that innovation is typically procyclical despite higher opportunity costs during economic upswings. Specifically, we show that sectors whose opportunity cost of innovation rises the most in boom times do not raise patenting during oil and gas booms: for producers of locally-sold products, we observe no significant change in patenting, and patenting in oil and gas – the industry at the center of the boom – actually declines. This indicates that contrary to earlier conjectures in the literature, a low empirical relevance of varying opportunity costs of innovation across the business cycle cannot explain the procyclicality of innovation.

Tables and Figures

Table 1: Local booms and patent activity

Dependent Variable →	# Patents		# Non-oil&gas patents		# Oil&gas patents	
	(1)	(2)	(3)	(4)	(5)	(6)
ln(National oil&gas employment) × endowment	0.083***	0.081***	0.088***	0.086***	-0.084**	-0.083***
	(0.028)	(0.027)	(0.034)	(0.033)	(0.033)	(0.032)
ln(Nat. O&G empl.) × endowment × Shale period		0.084		0.076		-0.091
		(0.077)		(0.082)		(0.174)
Observations	11,097	11,097	11,097	11,097	7,748	7,748
Sample period	69-12	69-12	69-12	69-12	69-12	69-12

Notes: In this table we analyze the impact of local economic booms, proxied by localized oil and gas shocks (see Table A2), on patent activity at the commuting zone level. We only consider applications which result in granted patents later on. The sample period is 1969-2012. In columns 1-2 we consider patents in all technology classes. In columns 3-4 we only include oil and gas patents (classes C10G, C10L, C10M and E21B), while in columns 5-6 we exclude those four classes. We aggregate the number of patents over a period of three years, except for the period 1999-2000 which constitutes a two-year period. *endowment* equals initial oil&gas reserves and is scaled by the standard deviation of this variable across all commuting zones. Endowment is updated to include fracking reserves from 2001 onwards. In all columns we estimate Equation (1) using Poisson pseudomaximum likelihood regressions. All regressions include commuting zone times century fixed effects and state times three year period fixed effects, and control for differential commuting zone-specific patenting trends across the pre-shale gas period (1969-2000) and the shale gas 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 2: Heterogeneous effects across commuting zone types and booms/busts

Dependent Variable →	# Non-oil&gas patents							
Sample →	All	<20 patents		All		Urban non-metro c-zones	Long bust spell	Long boom spell
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(National oil&gas employment) × endowment	0.089***	0.075***	0.072***	0.072***	0.106**	0.108***	-0.229	0.174*
	(0.034)	(0.024)	(0.028)	(0.028)	(0.046)	(0.016)	(0.147)	(0.105)
ln(N. O&G E.) × endowment × Non-metropolitan			0.070***					
			(0.027)					
ln(N. O&G E.) × endowment × Urban non-metro				0.070***	0.123***			
				(0.027)	(0.045)			
ln(N. O&G E.) × endowment × Rural non-metro				0.032	0.054			
				(0.134)	(0.162)			
ln(N. O&G E.) × endowment × Ini. pat-intensity					0.056**	0.054		
					(0.023)	(0.053)		
ln(N. O&G E.) × endowment × Human capital					0.031	0.137**		
					(0.035)	(0.068)		
ln(N. O&G E.) × endowment × College density					-0.040	-0.040		
					(0.029)	(0.032)		
Observations	11,097	6,445	11,097	11,097	11,097	7,363	5,173	2,880
Sample period	69-12	69-12	69-12	69-12	69-12	69-12	81-00	01-12

Notes: In this table we analyze heterogeneous effects on non-oil and gas patenting across different commuting zone types and across boom versus bust periods. In each column, we estimate Equation (1) (columns 1,2,7,8) or depart from it and add one or more interactions of the boom variable with a commuting zone-specific variable (columns 3-6). In column 2, we restrict the sample to commuting zones with an average of less than 20 patents per three-year period over 1969-2012. *Non-metropolitan* is originally defined at the county level and takes one of ten values defined as Rural-Urban Continuum Codes by the U.S. Department of Agriculture's Economic Research Service. The values rise with ruralness and classify the county as non-metropolitan if the county's value is four or larger. In column 2, *Non-metropolitan* is a dummy that equals one if the average value as of the year 1974 across all counties in the commuting zone is larger or equal four. In column 4, *Urban micro* equals one for a commuting zone average between (including) 4 and (excluding) 8, and *Rural micro* equals one for a commuting zone average larger or equal to 8. *Initial patent intensity* equals 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 or more with at least one year of college education, as of 1970, scaled by its standard deviation. *College density* equals the fraction of residents employed in "colleges, universities, and professional schools" that are located in the commuting zone, as of 2018. In columns 5 and 6, before performing the regressions, we demean the commuting zone characteristics that feature in the triple interaction terms. This allows us to interpret the non-interacted boom coefficient as the effect in the average commuting zone across the sources of heterogeneity we test for. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table 3: Mechanisms I: finance, input-output linkages

Dependent Var.: # Non-oil&gas patents by... →	All industries				Up-stream industries	Down-stream industries	Non-linked industries
Unit of observation →	Commuting zone – industry		Commuting zone – technology class – firm type		Commuting zone		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln(National oil&gas empl.) × endowment	0.065**	0.064**	0.061**	0.065**	0.066***	-0.009	0.090***
	(0.028)	(0.026)	(0.029)	(0.026)	(0.024)	(0.024)	(0.033)
ln(N. O&G E.) × endowm. × Ext. finance dep.		0.000					
		(0.003)					
ln(N. O&G E.) × endowm. × Compustat firms				-0.015***			
				(0.007)			
Observations	221,684	221,684	305,906	305,906	11,097	11,082	11,097

Notes: In this table we test different potential mechanisms behind the baseline results in column 5 of Table 1. In column 2, we test whether two-digit SIC industries that are more financially constrained (Rajan and Zingales, 1998) are differently affected by a local oil and gas boom in terms of their patenting behavior. In column 4, we test whether within a given IPC2 technology class, firms that are included in Compustat (and are thus publicly listed and less financially constrained, on average) are affected differently. The sample period in columns 3-4 starts with the three-year period 1975-1977 because the relevant data are only available from 1975 onwards. In column 5 we study the number of patents in industries that are upstream to the oil and gas sector, and not downstream. In column 6 we study patents by industries that are downstream to oil and gas (and potentially upstream). In column 7 we study patents in industries that are neither upstream nor downstream. In all columns we estimate Equation (1) or adjusted specifications using Poisson pseudomaximum likelihood regressions. The regressions in columns 1-2 contain commuting zone times century times two-digit SIC industry fixed effects, which implies 20 dummies per commuting zone for the pre-fracking period (1969-2000), and 20 for the fracking period (2001-2012); two-digit industry times period fixed effects; state times three year period fixed effects; and commuting zone-specific patenting trends, one for the pre-fracking period and another for the fracking period. In columns 3-4 we use the same fixed effects structure, but with technology class rather than industry 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 II: Agglomeration

Dependent Variable → ln()...	Adult Population	Adult Population, College	Adult Population, Non-College	Adult Population	Creative Class Workers
Data Frequency →	1990, 2000, 2006-10, 2011-15			Annual, '69-'12	'90, '00, '07-'11
	(1)	(2)	(3)	(4)	(5)
ln(National oil&gas empl.) × endowment	0.013***	0.016***	0.012***	0.013***	0.018**
	(0.004)	(0.006)	(0.004)	(0.003)	(0.008)
Observations	3,024	3,024	3,024	33,264	2,267
Sample period	90-15	90-15	90-15	69-12	90-11

Notes: In this table we mainly study the impact of local oil and gas booms on college- versus non-college educated population (columns 2-3), and on the number of 'creative class workers' (column 5). To ease the interpretation of the results in columns 2 and 3, in column 1 we use *total* adult population as dependent variable. College-educated refers to a completed college degree, and "adult" refers to age 25+. 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 across 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 across 2006-2010, and as average across 2011-2015. In column 4, we study total population on an annual basis, over 1969-2012. *Creative Class Workers* in column 5 are obtained from the *Economic Research Service* (ERS), 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". 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 III: Local redirection of innovation, and path dependence

Dependent Variable →	# Non-oil&gas patents			
Unit of observation →	Commuting zone – technology class			
	(1)	(2)	(3)	(4)
ln(National oil&gas empl.) × endowment	0.060***	0.053***	0.048***	0.049**
	(0.016)	(0.019)	(0.017)	(0.021)
ln(Nat. oil&gas empl.) × endowment × IPC2's relatedness to O&G	0.088*	0.085*		
	(0.049)	(0.051)		
ln(Nat. O&G empl.) × end. × IPC2's share in total c-zone patents, 1960-69			0.181***	0.184***
			(0.039)	(0.052)
ln(Nat. O&G empl.) × end. × IPC2's share in total US patents 1969-2012		0.075		-0.011
		(0.059)		(0.085)
Observations	181,661	181,661	180,822	180,822
Sample period	69-12	69-12	69-12	69-12

Notes: In this table we continue to test different potential mechanisms behind the baseline results in column 5 of Table 1. The unit of observation equals two-digit IPC technology classes in a given commuting zone, thus there are 23 observations per commuting zone and time period. Whenever applicable, we remove oil and gas patents from the total patent count in a two-digit IPC class, thereby focusing on non-oil and gas patenting. In columns 1 and 2 we test whether the rise in non-oil and gas patenting during booms is dis-proportionally driven by technology classes that are related to oil and gas. *IPC2's relatedness to O&G* equals the share of oil and gas patents for which the specific non-oil and gas IPC2 class is also among the listed technology classes of the patent. These ratios are computed based on all oil and gas patents globally over 1969-2012. In columns 3 and 4 we test whether the rise in non-oil and gas patenting is dis-proportionally driven by two-digit IPC classes that the county is historically focusing on. In columns 2 and 4 we control for the general relevance of a two-digit IPC class, over 1969-2012. In all columns we estimate Poisson pseudomaximum likelihood regressions. All regressions include two-digit IPC class times commuting zone times century fixed effects, state times three year period fixed effects, two-digit IPC class times three-year period fixed effects, and control for differential commuting zone-specific patenting trends across the pre-shale gas period (1969-2000) and the shale gas period (2001-2012). Note that the number of observations in columns 3-4 is smaller than in columns 1-2 because there are commuting zones that do patent at least once during 1969-2012 but do not patent during 1960-1969. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table 6: Heterogeneity across different sectors

Dependent Variable →	# Non-oil&gas patents				
Included Industries →	Lowly Traded	Highly Traded	Least Traded Tercile	Inter-mediate Tercile	Most Traded Tercile
	(1)	(2)	(3)	(4)	(5)
ln(National oil&gas empl.) × endowment	0.044	0.092***	0.017	0.068*	0.093***
	(0.037)	(0.027)	(0.031)	(0.035)	(0.029)
Observations	11,097	11,097	11,097	11,097	11,097
Sample period	69-12	69-12	69-12	69-12	69-12

Notes: In this table we estimate Equation (1) on different sub-samples in order to study heterogeneous effects on non-oil and gas innovation across sectors producing highly- versus lowly-traded goods. Industries' tradedness is measured by their distance elasticity to trade as calculated by Holmes and Stevens (2014); in columns 1-2, we measure highly traded industries as those with a below-median distance elasticity, across all four-digit SIC manufacturing sectors. In columns 3-5, we aggregate patents into those filed by firms in sectors in the first, second, and third tercile, respectively, in terms of tradedness. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table 7: Patent quality and green innovation

Dependent Variable →	# Non-oil&gas patents	Average # FWcitations, non-O&G patents	Average Generality, non-O&G patents	# Green Patents	# Green Patents / # Total Patents
	(1)	(2)	(3)	(4)	(5)
ln(National oil&gas empl.) × endowment	0.087***	0.010	-0.026	0.065***	-0.315**
	(0.034)	(0.027)	(0.024)	(0.018)	(0.148)
Observations	10,340	9,591	9,306	9,423	9,163
Sample period	69-09	69-09	69-09	69-12	69-12

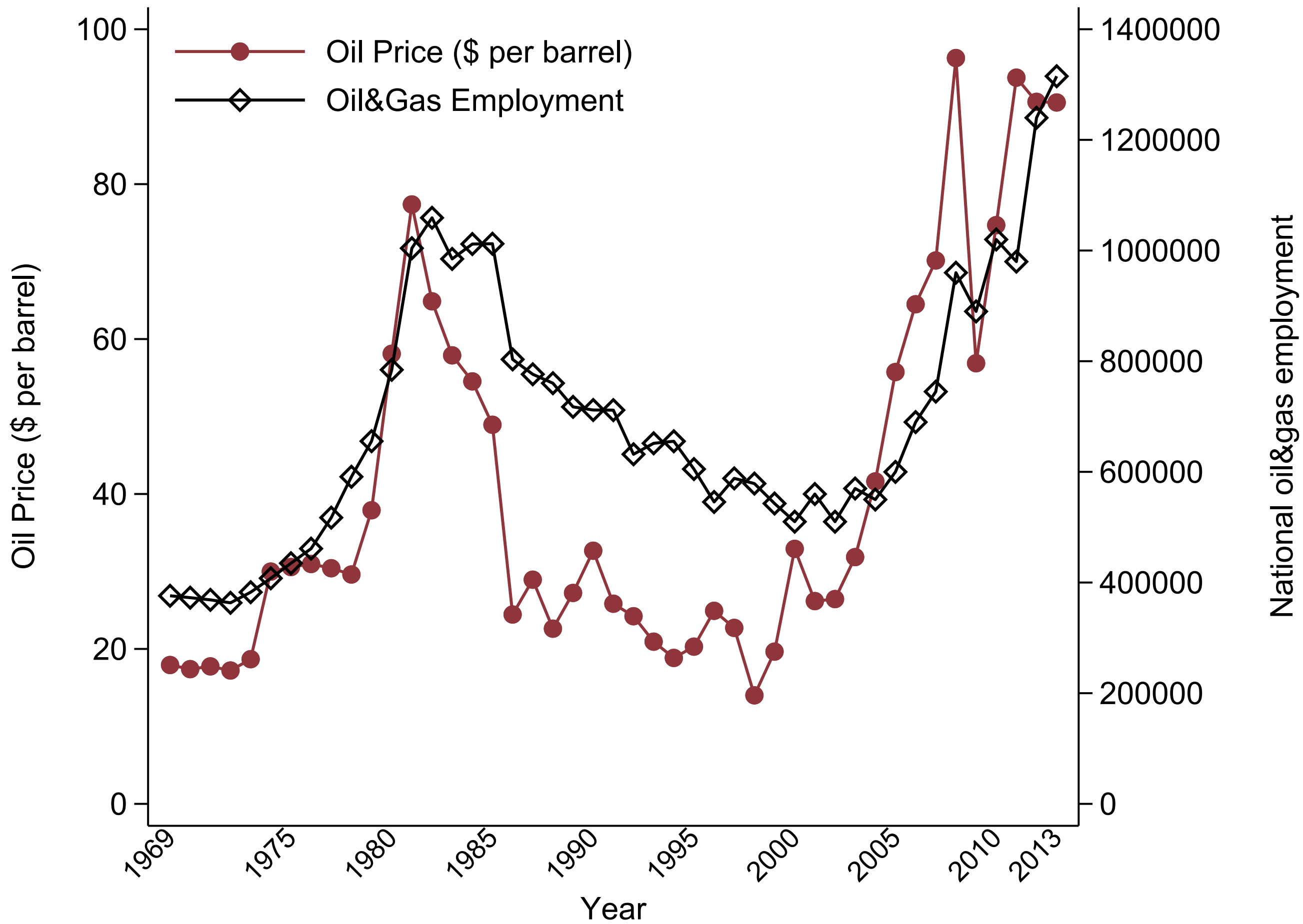
Notes: In this table we study patent quality measures (columns 2-3) and green patenting during oil and gas booms (columns 4-5). In column 2, the dependent variable equals the total number of forward citations of all non-oil&gas patents over the following five years, divided by the total number of non-oil&gas patents. This ratio proxies for the average quality of patents filed (and later granted) in a given three-year period. In column 3, the dependent variable equals the average (and normalized) generality score across all non-oil&gas patents. In columns 2 and 3 we only include commuting zone – three year periods with at least one patent. The sample period in columns 2 and 3 omits the final period (2010-2012) as we cannot evaluate forward citations of patents in this period. *Green Patents* are those classified into the CPC class Y02. In all columns we estimate Poisson pseudomaximum likelihood regressions. All regressions include commuting zone times century fixed effects and state times three year period fixed effects, and control for differential commuting zone-specific patenting trends across the pre-shale gas period and the shale gas period. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Table 8: Robustness Checks

Nature of Robustness Check →	Base-line	Control for: X_c × O&G Empl.	Control for coal booms	Pre-frack. 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.) × endowment	0.088***	0.086**	0.088**	0.111**	0.088**	
	(0.034)	(0.037)	(0.034)	(0.054)	(0.045)	
ln(Oil price) × endowment						0.031***
						(0.006)
Observations	11,097	11,097	11,097	11,119	11,119	11,097
Panel II: Dependent Variable →			# Oil&gas patents			
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Nat. O&G empl.) × endowment	-0.084**	-0.078**	-0.083***	-0.078**	-0.047*	
	(0.033)	(0.037)	(0.032)	(0.033)	(0.026)	
ln(Oil price) × endowment						-0.032**
						(0.016)
Observations	7,748	7,748	7,748	7,748	7,748	7,748

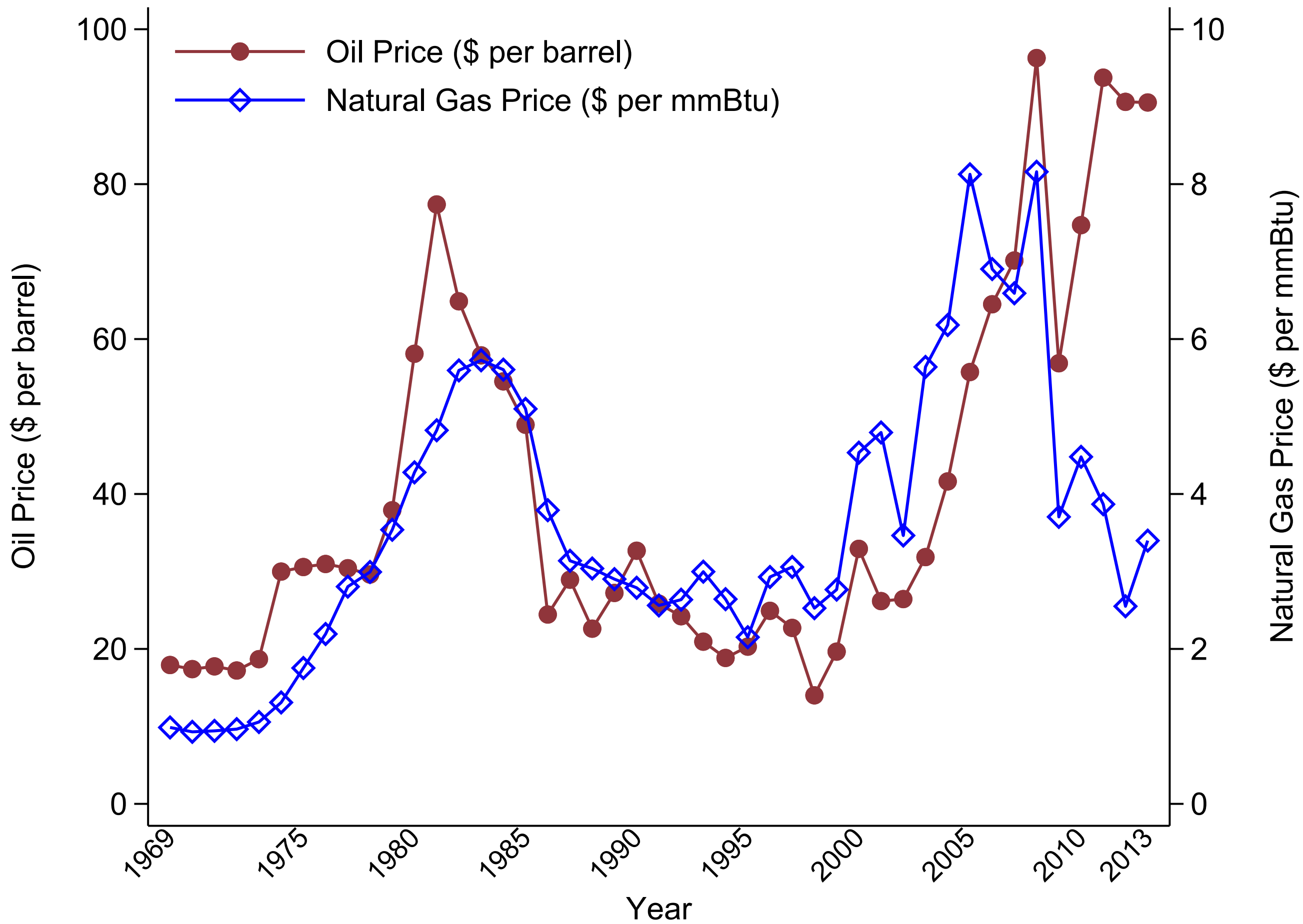
Notes: In this table we carry out several robustness checks on the results of column 3 (Panel I) and column 5 (Panel II) of Table 1, respectively. In both panels, column 1 repeats the relevant baseline results. In column 2 we add interactions of national oil and gas employment with county-level population density, personal income per capita, and population, respectively, all measured in 1969. In column 3 we add an interaction of 1960 coal reserves at the commuting zone level and national coal employment. In column 4 we use pre-fracking reserves for the entire sample period, and in column 5 we use total reserves for the entire sample period. In both columns, we demean the equation based on the full sample period rather than separately for 1969-2000 and 2001-2012 as in the baseline. In column 6 we use the oil price rather than national oil and gas employment in our key interaction term. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Figure 1: Oil Price and National Oil and Gas Employment 1969-2013



Notes: Oil price data are from the Energy Information Administration. Oil Prices are in 2010 real dollars.

Figure 2: Oil Price and Natural Gas Price 1969-2013



Notes: Oil price and natural gas price data are from the Energy Information Administration. Prices are in 2010 real dollars.

Appendix

Table A1: Summary Statistics

	Mean	Median	Min	Max	sdev	N
Panel I: Commuting zone - year level variables						
# Patents	215.43	12.17	0	28,517	994.30	11,385
# Oil and gas patents	5.48	0	0	2,996	53.69	11,385
# Non-oil and gas patents	209.95	12	0	28,472	978.49	11,385
...in Metropolitan CZs	1639	671.68	4	28,472	2,690	1,095
...in Urban non-metropolitan CZs	78.07	16.33	0	5,962	259.31	7,440
...in Rural non-metropolitan CZs	5.39	2	0	89.08	9.06	2,850
...in Upstream industries	35.09	2.72	0	2,332	135.64	11,385
...in Downstream industries	9.29	0.60	0	524.07	35.91	11,385
...in Non-linked industries	158.59	7.86	0	25,401	809.23	11,385
...by Compustat firms	100.03	1.60	0	19,866	539.30	9,867
...by non-Compustat firms	121.07	9.50	0	27,175	597.46	9,867
...by highly traded industries (highest tercile)	134.64	5.31	0	24,278	735.73	11,385
...by medium-traded industries (medium tercile)	44.12	3.80	0	2,896	164.53	11,385
...by lowly-traded industries (lowest tercile)	24.20	1.98	0	1,313	89.16	11,385
# Forward citations per non-oil and gas patent	6.13	4.42	0	160.73	6.33	9,677
Patent generality	0.36	0.37	0	0.90	0.16	9,404
# Green patents	9.86	0	0	1,789	51.05	11,385
Green patent share, in %	4.72	1.73	0	100	9.76	10,397
Earnings per worker (in '000, real 2010 dollars)	31.84	31.31	15.81	80.53	5.92	11,385
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; mill.$)	1.48	0.03	0	56.96	4.58	759
O&G reserves / Area, 1960 (incl. shale; mill.$)	2.86	0.18	0	73.76	7.05	759
# Non-O&G pat. 1960-69 / Pop. 1969, in '000	4.43	1.99	0	63.58	7.06	759
Percentage of adult pop. with ≥1y college, 1970	16.72	16.08	6.19	47.60	5.39	759
Percentage of pop. 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. *Downstream industries* refers to industries that are downstream to oil and gas and may also be upstream, given that "pure" downstream industries are very rare. *Non-linked* means neither upstream nor downstream. The distinction into patents by Compustat versus non-Compustat firms is based on the period 1975-2012, since earlier data do not permit this distinction. 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. *Colleges etc.* refers to "colleges, universities, and professional schools", and *local colleges* 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 Section 4.7.

Table A2: Oil and gas booms and local economic activity

Dependent Variable → ln()...	Population	Employment	Earnings per Worker	GDP	Local Government Revenue
	(1)	(2)	(3)	(4)	(5)
ln(National oil&gas empl.) × endowment	0.019***	0.037***	0.022***	0.054**	0.063***
	(0.006)	(0.007)	(0.003)	(0.025)	(0.011)
Observations	11,340	11,340	11,340	3,024	6,791
Sample period	69-12	69-12	69-12	01-12	72-12

Notes: In this table we analyze the impact of oil and gas booms on various measures of local economic activity. For each variable we use the average realization over the three-year period. Data on county-level GDP is 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. Intuitively, we thus analyze for instance how the change in county-level revenue between 1982 and 1987 is affected by the change in national oil&gas employment between 1982 and 1987, interacted with initial oil and gas endowment at the county level. *endowment* equals initial oil&gas reserves and is scaled by the standard deviation of this variable across all commuting zones. Endowment is updated to include fracking reserves from 2001 onwards. In all columns we estimate Equation (1) using OLS. All regressions include commuting zone times century fixed effects and state times three year period fixed effects, and control for differential commuting zone-specific patenting trends across the pre-shale gas period (1969-2000) and the shale gas 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 A3: Correlation of oil and gas endowment with other local characteristics

Dependent Variable →	Pre-fracking endowment (R_c^{early})		Total endowment (R_c^{total})	
	(1)	(2)	(3)	(4)
Urban non-metropolitan commuting zone	-0.074		-0.043	
	(0.165)		(0.229)	
Rural non-metropolitan commuting zone	-0.111		-0.063	
	(0.169)		(0.249)	
Personal income per capita 1969	0.091	0.095	0.104	0.104
	(0.067)	(0.063)	(0.093)	(0.087)
Human capital 1970	-0.030	-0.029	0.001	-0.000
	(0.040)	(0.038)	(0.082)	(0.080)
Population 1969		0.021		0.020
		(0.030)		(0.040)
Observations	757	757	757	757
State FE	Yes	Yes	Yes	Yes

Notes: In this table we test the correlation between initial oil and gas endowment and other commuting zone characteristics. Endowment is scaled by commuting zone size, as in our baseline specification. Human capital equals the share of the commuting zone's population aged 25 or more with at least one year of college education, as of 1970. The dependent variables are scaled by the standard deviation of pre-fracking endowment. The explanatory variables are scaled by their respective standard deviations. Robust standard errors are in parentheses. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

References

Acemoglu, D., D. Hemous, L. Barrage, and P. Aghion (2019). Climate change, directed innovation, and energy transition: The long-run consequences of the shale gas revolution. Unpublished Working Paper.

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. and N. Goldschlag (2023). Where have all the "creative talents" gone? employment dynamics of us inventors' NBER Working Paper 31085.

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

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.

Angelucci, S., F. J. Hurtado-Albir, and A. Volpe (2018). Supporting global initiatives on climate change: the EPO's "Y02-Y04S" tagging scheme. *World Patent Information 54*, S85–S92.

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. 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.

Auty, R. (1993). *Sustaining economic growth in mineral economies: the resource curse thesis.* London: Routledge.

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.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics 119*(1), 249–275.

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

Cascio, E. U. and A. Narayan (2022). Who needs a fracking education? The educational response to low-skill-biased technological change. *Industrial and Labor Relations Review 75*(1), 56–89.

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.

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.

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.

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.

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

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

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.

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.

Hinton, D. D. (2012). The seventeen-year overnight wonder: George Mitchell and unlocking the Barnett Shale. *The Journal of American History 99*(1), 229–235.

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.

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.

Kogler, D. F., D. L. Rigby, and I. Tucker (2013). Mapping knowledge space and technological relatedness in US cities. *European Planning Studies 21*(9), 1374–1391.

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.

Makridis, C. A. and E. McGuire (2023). The quality of innovation "booms" during "busts". *Research Policy 52*(1), 104657.

Manso, G., B. Balsmeier, and L. Fleming (2023, 09). 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.

Mehlum, H., K. Moene, and R. Torvik (2006). Institutions and the resource curse. *The Economic Journal 116*(508), 1–20.

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.

Nagaoka, S., K. Motohashi, and A. Goto (2010). Patent statistics as an innovation indicator. In *Handbook of the Economics of Innovation*, Volume 2, pp. 1083–1127. Elsevier.

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

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.

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.

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.

Sala-i Martin, X. and A. Subramanian (2013). Addressing the natural resource curse: an illustration from Nigeria. *Journal of African Economies 22*(4), 570–615.

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.

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

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 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.

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

February 28, 2024

Table of Contents

OA1 Additional Results

OA1.1 Oil and gas booms and patenting by 2-digit technology class (Table OA1)

In Table OA1 we study the impact of oil and gas booms on patenting in all 23 two-digit technology classes. Otherwise, the specification is equivalent to Equation (1). While the results show positive effects in multiple one-digit technology classes, the effects are most pronounced in classes F (Engineering, broadly defined), G (Instruments, Physics), and H (Electricity). We also see that positive boom effects tend to be stronger in technology classes that are more related to oil and gas (compare Figure OA1).

OA1.2 Oil and gas booms and local oil and gas activity (Table OA2)

The BEA's Regional Economic Accounts data provides 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 the county-level data on "mining" employment to the commuting zone level and use them to study local oil and gas activity during oil and gas booms. As shift variable, we use the oil price rather than national oil and gas employment to avoid simultaneity. We control for initial coal endowment, 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 OA2).

Note that we do not use local oil and gas production as dependent 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.

OA1.3 Testing for public finance and wealth effects (Table OA3)

Public finance

In Table OA3 we explore potential mechanisms that appear to have a low likelihood of explaining our results (see also Section 4.4.5). In column 2 we test for a potential public finance channel, although local governments typically spend oil and gas revenue on other items such as education or infrastructure (Newell and Raimi, 2015, 2018). 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 varies by state (but not within a state) and is documented by Newell and Raimi (2018) for 16 US states, which jointly produce more than 97% of US oil and natural gas. 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 (pre-fracking and fracking) oil and gas endowment. As control variables, we include all interaction terms from the specification in column 5 of Table 2. The results 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 *smaller* impact on non-oil and gas patenting (the coefficient is statistically significant at the 10% level). This clearly speaks against the hypothesis that local county governments in booming commuting zones use oil and gas windfalls to support local innovation.

Wealth effects

In columns 3 and 4 of Table OA3, we test whether oil and gas booms lead to a rise in local house prices. We do so because rising housing prices might give rise to a household wealth-related mechanism driving our results – although Bernstein et al. (2021) find that positive wealth shocks do not affect inventor productivity (unlike negative shocks). We use a county-level house price index for our analysis, which reaches back as far as 1975 for some counties. The index is based on all transactions in a given year and is made available by the Federal Housing Finance Agency (FHFA). Our chosen unit of observation is a county, since the index nature of the data and varying index base years across counties make it impossible to meaningfully aggregate to the commuting zone level. In column 3 of Table OA3, we estimate Equation (1) with the dependent variable being the average index realization over a three-year period, as in our baseline analysis. In column 4, we use the original, annual data. The results suggest a positive effect of relatively small magnitude: in the commuting zone with oil and gas endowment equal to one standard deviation, a doubling of national oil and gas employment leads to an increase of local house prices by 0.5%. The coefficient is only (marginally) statistically significant if we use annual data (which increases statistical power). Taken together with the evidence of Bernstein et al. (2021), these results speak against inventor wealth effects as an important driver of our results.

Table OA1: Natural resource booms and non-oil&gas patents by two-digit technology class

Dep. Var. → # Non-oil&gas patents in IPC class...	A0 Agri-culture	A2 Food; Tob-acco	A4 Person. or dom. articles	A6 Health; Lifesav.; Amusem.	B0 Separ-ating; Mixing	B2 Shap-ing	B4 Print-ing	B6 Trans-port -ing	B8 Microstr. / Nano tech.	C0 Chem-istry	C2 Metall-urgy	C4 Combin-atiorial tech
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ln(Nat. oil&gas empl.) × endowment	-0.053	-0.160***	-0.058	0.034***	0.037	0.094**	-0.363***	0.004	0.745***	-0.038	0.429***	0.094
	(0.063)	(0.026)	(0.046)	(0.008)	(0.039)	(0.037)	(0.046)	(0.019)	(0.140)	(0.028)	(0.059)	(0.360)
Observations	10,255	7,525	9,939	9,780	9,172	9,721	6,975	10,545	2,752	9,306	6,449	1,149

Dep. Var. → # Non-oil&gas patents in IPC class...	D0 Textiles; Flexible materials	D2 Paper	E0 Buil-ding	E2 Earth or Rock Drilling	F0 Engines; Pumps	F1 Engin-eering in General	F2 Lighting; Heating	F4 Weapons; Blasting	G0 Instru-ments	G2 Nuclear physics/ engineer.	H0 Electr-icity
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
ln(National oil&gas empl.) × endowment	0.127	0.034	0.056**	-0.182*	-0.007	0.058***	-0.091	0.030***	0.067***	-0.005	0.150***
	(0.206)	(0.164)	(0.024)	(0.106)	(0.062)	(0.005)	(0.063)	(0.011)	(0.009)	(0.181)	(0.039)
Observations	6,649	4,004	9,688	2,530	8,975	8,991	9,195	7,916	9,892	3,479	9,698

Notes: In this table we study the impact of oil&gas booms on patent activity in the different two-digit IPC classes, excluding the oil and gas classes C10G, C10L, C10M and E21B. In all columns we estimate Equation (1) using Poisson pseudomaximum likelihood regressions. All regressions include commuting zone times century fixed effects and state times three year period fixed effects, and control for differential commuting zone-specific patenting trends across the pre-shale gas period (1969-2000) and the shale gas 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: Aggregate oil and gas booms and local oil and gas employment

Dependent Variable →	# 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	0.094
	(0.062)
Observations	10,682
Sample period	69-12

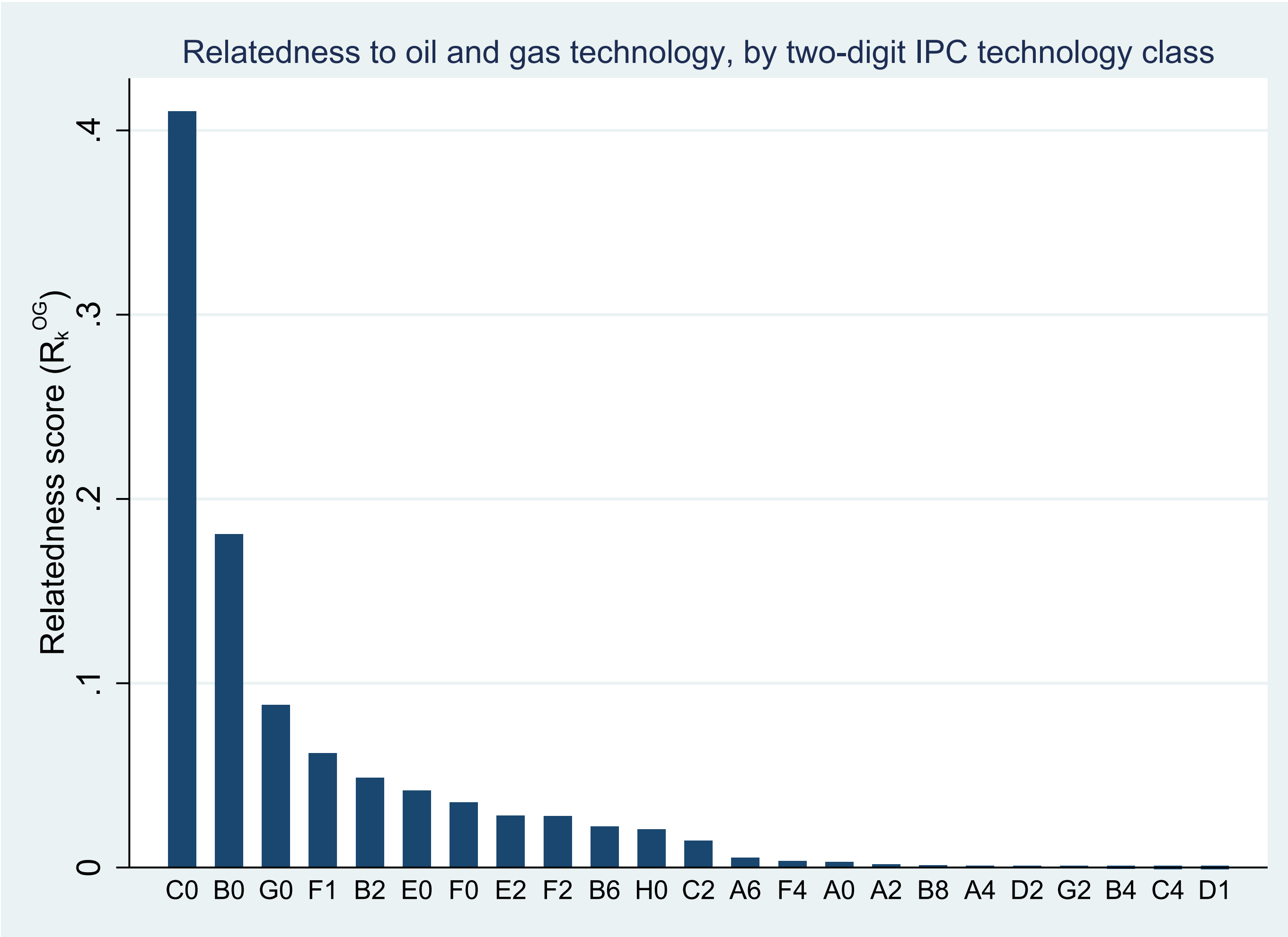
Notes: In this table we analyze the impact of oil and gas booms on employment in oil and gas (and other mining) at the commuting zone level. We estimate Equation (1) but use the oil price rather than national oil and gas employment as shift variable, to avoid simultaneity. Data on the dependent variable are from BEA's Regional Economic Accounts. 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 additional mechanisms

Dependent Variable →	# Non-oil&gas patents		ln(House Price Index)	
Unit of observation →	Commuting zones, three-year periods		Counties, annual	
Sample →	CZs in states with O&G revenue sharing info, and CZs without O&G endowm.		All counties with available data	
	(1)	(2)	(3)	(4)
ln(National oil&gas empl.) × endowment	0.088**	0.111**	0.005	0.005*
	(0.035)	(0.056)	(0.004)	(0.003)
ln(N. O&G E.) × endowm. × O&G revenue share		-0.100*		
		(0.059)		
ln(N. O&G E.) × endowm. × Urban non-metro		0.102**		
		(0.040)		
ln(N. O&G E.) × endowm. × Rural non-metro		0.078		
		(0.196)		
ln(N. O&G E.) × endowm. × Ini. pat-intensity		0.039*		
		(0.023)		
ln(N. O&G E.) × endowm. × Human capital		0.013		
		(0.034)		
ln(N. O&G E.) × endowm. × College density		-0.040		
		(0.030)		
Observations	6,137	6,137	24,507	67,892
Sample period	69-12	69-12	75-12	75-12

Notes: In this table we explore potential public finance (column 2) and inventor wealth effects (columns 3-4). Fixed effects and the clustering of standard errors 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 1, column 3. *O&G revenue share* is scaled by the variable's standard deviation. Standard errors in parentheses are clustered at the state level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

Figure OA1: Relatedness to oil and gas



Notes: This graph plots the relatedness score, R_k^{OG}, of each two-digit IPC technology class. See Section 4.4 for details.

OA2 Details on patent data computations

OA2.1 Patent technology codes and technology classes

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

C/IPC classification system

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

The structure of the C/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 C/IPC 6-digit and C/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:

$$\omega_{ik} = \frac{\sum_l IPC_{ikl}}{\sum_l IPC_{il}}, \tag{2}$$

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 (CIT OECD).

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 (counties). Ideally, we compute weights as

$$\omega_{ir} = \frac{I_{ir}}{I_i} \tag{3}$$

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:

$$\omega_{ir} = \frac{1}{\#r} \tag{4}$$

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 (2).

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

$$g_{it} = 1 - \sum_{k \in K} \left(\frac{cit_{ik}}{cit_i} \right)^2, \tag{5}$$

Note that the g_{it} is undefined if patents i is never cited.[1]

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:

$$g_{it}^{norm} = \frac{g_{it}}{\bar{g}_t}, \tag{6}$$

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 $g_{kt} = \frac{\sum_i \omega_{ik} g_{ik}}{\sum_i \omega_{ik}}$

[1] 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 $k \in K_j$.

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 $g_{it}^{norm} = \frac{0.5}{0.525} = 0.95$.

Notes on practical implementation: Sample selection

Cited patents: all patents with at least one inventor based in the US and with non-NA O&G classification (baseline DWPI classification). Note that patents can be filed to the USPTO and/or other patent offices, although, in practice, patents by US inventors are rarely filed to foreign patent offices only.

Citing patents: *all* patents with non-NA C/IPC codes, excluding Y codes (regardless of where they are filed or where inventors and assignee are located). Note: Y codes are excluded because they are not a category on it's own, but an additional category added by EPO to identify green patents. Patents with Y codes and non-Y codes are kept, but Y codes dropped; patents with Y codes only are dropped.

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:

$$\bar{g}_{tr} = \frac{\sum_r \omega_{ir} g_{itr}}{\sum_r \omega_{ir}} \tag{7}$$

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

$$p_{rt}^{OG} = \sum_i \omega_{ir} pat_{it}^{OG} \tag{8}$$

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

$$p_{krt}^{nOG} = \sum_1 \omega_{ik}\omega_{ir} pat_{ikrt}^{nOG} \tag{9}$$

We can compute p_{krt}^{OG} in the same way. $p_{krt}^{nOG} + p_{krt}^{OG} = p_{krt}$ 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 our key variables: local oil and gas endowment and local patenting. We describe other data sources and variable computations, dataset by dataset, below.

National oil and gas employment

Data are provided by the Bureau of Economic Analysis (BEA), via *SAEMP25: Total full-time and part-time employment by industry.*[2] 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,

[2] https://apps.bea.gov/itable/?ReqID=70&step=1eyJhcHBpZCI6NzAsInN0ZXBzIjpbMSwyOSwyNSwzMV0sImRhdGEiOltbIlRhYmxlSWQiLCIzMCJdLFsiTWFqb3JfQXJlYSIsIjAiXV19

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.

Derwent World Patents Index (DWPI)

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 codes provide a comprehensive coverage of all aspects of the oil and gas industry, excluding competitive products such as coal and peat. They identify activities such as obtaining crude oil and natural gas (C01G, E21B), unit operations (C10G), transportation and storage, petroleum processing (C10G), refining and engineering, gaseous and liquid fuels (C10L), lubricants and lubrication (C10M), and petroleum products other than fuels and lubricants (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.

Population, Employment, Earnings per worker, Personal income per capita, GDP, Oil and gas employment

These county-level data items 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`.

For employment, we use the variable "Total employment (Number of jobs)", which is available in the data set *CAEMP25*, or in *CAINC4*. Earnings per worker is computed as "Wages and salaries" divided by "Wage and salary employment". Both are available in *CAINC4*, same as population (number of persons) and personal income, which we divide by population to arrive at personal income per capita. GDP, in "thousands of chained 2012 dollars", is obtained via *CAGDP1*. Unlike all other variables, which are available from 1969 onwards, GDP data are only available from 2001 onwards.

Oil and gas employment is not available as 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 (obtained from *CAEMP25*) 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. Note that oil and gas employment is not reported for some counties and years, to avoid disclosure of confidential information. When aggregating from the county to the commuting zone level, we treat those observations as being equal to zero. The results are robust to instead using the median value across all counties in the particular year.

County government revenue

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).[3]

Classifying commuting zones into metropolitan versus non-metropolitan etc.

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 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 X). 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 4≤rural-urban code<8. Rural non-metropolitan commuting zones are those with a code≥8.

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

Figure OA2: 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 (See `https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation/#RevisionHistory`) for a description of changes which make pre-2003 data not perfectly comparable to later data) Figure source: `https://wayback.archive-it.org/5923/20110914000642/http://www.ers.usda.gov/Briefing/Rurality/RuralUrbCon/priordescription.htm`

Adult population by educational attainment

Until 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 Table 2), 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 of agglomeration economies, we want to study local adult population by educational attainment 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 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 must resort to data from American Community Survey (ACS). We downloaded ACS data from the Census website `data.census.gov`.[4] Note that ACS has "insufficient coverage for a reliable county estimate in one year" (Weber, 2014), so 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.[5]

College density

In a global sample, Valero and Van Reenen (2019) define college density as the number of universities by capita. However, as the authors point out, "A disadvantage of the "university density" measure is that it does not correct for the size (...) of the university." (p.55) We are able to correct for college size because we have college-level data on the number of employees. Therefore, we define college density as the ratio of college employees to local population, at the commuting zone level.

College-level employment data are provided in the data set *Homeland Infrastructure Foundation-*

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

[5] Comparing values across the census and five-yearly ACS averages is feasible, see for example Weber (2014).

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 data set 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 data set 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. The data set 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.

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 are robust to instead using the median number of college employees.

Creative class workers

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, which we aggregate to the commuting zone level.

Classifying industries as upstream versus downstream to oil and gas

We use the 2007 U.S. Input-Output tables of the *Bureau of Economic Analysis* (BEA) to identify upstream plants. The data set 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 (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:

$$Upstream_j = \frac{Sales_{j,OG}}{\sum_j Sales_j} + \sum_{-J} \left[\frac{Sales_{j,-j}}{\sum_j Sales_j} \times \frac{Sales_{-j,OG}}{\sum_j Sales_{-j,j}} \right] \in [0,1]$$

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

We then walk from the BEA code to the four-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 four-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

For each of 457 four-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`.[6]

Based on the 456 sectors that are represented in our patent-by-industry data (see Section 4.4.2) and in the Holmes-Stevens data, we compute the median distance elasticity. This median 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.

Note that it is for a reason that we do not apply Allcott and Keniston's cutoff in our analysis. The reason is that there is relatively little patenting activity in industries with a distance elasticity larger than 0.8. This could imply that if we defined a highly- and a lowly-traded industry group using this cutoff, and evaluated heterogeneity in terms of patenting during booms, our results might merely reflect differences in statistical variation within the highly- versus the lowly-traded bin. It is better to define the median cutoff, which includes more industries in the lowly-traded bin, or to define terciles, which essentially splits Allcott and Keniston's highly-traded bin into two separate bins, where the variation in patenting across the most-traded and the medium-traded bin is not overly distinct.

House prices

We use a county-level house price index in our analysis, which reaches back as far as 1975 for some counties. The index is based on all transactions in a given year and is made available by the Federal Housing Finance Agency (FHFA). The data can be accessed at `https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx#qexe`, under *Additional Data* → *Annual House Price Indexes* → Counties.

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

County splits

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

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.

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.

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

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.

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.

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

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