Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Boom Town Business Dynamics

Ryan A. Decker, Meagan McCollum, Gregory B. Upton Jr.

2020-081

Please cite this paper as:
Decker, Ryan A., Meagan McCollum, and Gregory B. Upton Jr. (2020). "Boom Town Business Dynamics," Finance and Economics Discussion Series 2020-081. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2020.081.

Boom Town Business Dynamics*

Ryan A. Decker
Meagan McCollum
Gregory B. Upton Jr.

September 2, 2020

Abstract

The shale oil and gas boom in the U.S. provides a unique opportunity to study economic growth in a "boom town" environment, to derive insights about economic expansions more generally, and to obtain clean identification of the causal effects of economic growth on specific margins of business adjustment. The creation of new business establishments—separate from the expansion of existing establishments—accounts for a disproportionate share of the multi-industry employment growth sparked by the shale boom, an intuitive but not inevitable empirical result that is broadly consistent with canonical models of firm dynamics. New firms, in particular, contribute nearly half of the cumulative economic growth resulting from the shale boom.

JEL Codes: L26, M13, E24, R23, Q33, Q35

---

*Decker: Federal Reserve Board (corresponding author; ryan.a.decker@frb.gov). McCollum: Department of Finance, University of Tulsa. Upton: Center for Energy Studies, Louisiana State University. This paper previously circulated under the title "Firm Dynamics and Local Economic Shocks: Evidence from the Shale Oil and Gas Boom." Emily Wisniewski (Fed) and Andrew Smith (LSU) provided excellent research assistance. We thank Shai Bernstein, Jason Brown, Chiara Lacava, Huiyu Li, Nida Cakir Melek, Judith Ricks, Scott Stern, Xiaoqing Zhou, and participants at the 2017 Federal Reserve System Energy Conference, the 2017 SEA conference, ASSA 2018, FMA 2019, the World Congress of Energy and Resource Economists 2018, the George Mason University applied micro group, Union College, the 2018 IFN Stockholm Entrepreneurship and Innovation workshop, the NBER Summer Institute 2019 Entrepreneurship group, the 2018 JPMCC New Directions in Commodities Research symposium, Singapore Management University, University of Hong Kong, University of Georgia, Oberlin College, the Federal Reserve Bank of Richmond, York University, Resources for the Future, and the Hamilton Project at Brookings for helpful comments. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau, the Board of Governors, or its staff. This research was performed at a Federal Statistical Data Center under FSRDC Project Number 1237. All results have been reviewed to ensure that no confidential information is disclosed.

1 Introduction

What does an economic boom town look like? More broadly, when economic growth occurs, *who* does the growing? In response to positive economic shocks, firms can either expand their existing business operations or create new "greenfield" business establishments; alternatively, entrepreneurs may enter with entirely new firms. Canonical models of firm dynamics suggest that the business entry margin plays a critical role in facilitating the economy's aggregate response to economic shocks, but the question is difficult to study empirically due to the paucity of exogenous growth shocks. Yet the question is of critical importance to researchers and policymakers alike. Firm dynamics models in wide use must be disciplined by empirical patterns of business adjustment. Well-designed policy depends on an understanding of the margins of business activity that are most responsive to stimulus; and many of the official statistics policymakers follow necessarily omit entering businesses, a costly (though unavoidable) omission if entry is an important growth margin.[1]

The U.S. shale oil and gas revolution provides a unique opportunity to study the dynamics of "boom towns" experiencing an exogenous economic shock: in response to a rapid expansion of oil and gas activity, areas affected by the shale boom saw significant employment growth in many other local industries. We describe the evolution of the U.S. shale boom towns in terms of the formation and growth of businesses both in and outside of the shale oil and gas industries, contributing a new dimension to our understanding of booming economies and the shale boom specifically. Net new establishments *accounted* for a large share of overall boom town employment growth; additionally, as compared to plausible counterfactuals, new firms and establishments contributed disproportionately to the growth *caused* by the shale boom. The role of new firms is particularly notable a few years after the shale boom began; nearly half of cumulative employment growth from 2006 to 2014 was supplied by firms founded after 2006. To the extent that our findings are generalizable, we contribute not only to literature on boom towns and the various consequences of the shale boom but also to existing literature on the business dynamics foundations of aggregate fluctuations.

While a focus on the shale boom does present challenges in terms of generalizability and external validity, it is nevertheless an important opportunity to seek lessons about economic

[1]Most prominently, the monthly BLS Current Employment Statistics jobs report (commonly referred to as "the establishment survey") relies on a sample of continuing establishments, filling in the estimated job contribution of establishment births with an ARIMA forecasting model which, until recently, jumped off administrative data from a year earlier; see https://www.bls.gov/web/empsit/cesbd.htm. The model was modified somewhat in 2020 to allow current behavior of continuing establishments to influence birth and death estimates.

growth in general. Natural experiments in which growth shocks can be thought of as exogenous are rare, but the shale boom presents one such case. After many years of declining crude oil production in the United States, recent technological developments made the extraction of previously inaccessible energy resources feasible in regions with certain preexisting geological characteristics. Specifically, the advent of horizontal drilling and hydraulic fracturing techniques enabled the exploration and production of oil and gas from "shale" geological formations and led to significant new drilling activity. Because of the nature of these geological formations, an economic "boom" occurred in clearly specified local areas where these previously inaccessible resources could now be profitably extracted. Indeed, many of these areas had no significant oil and gas activity before these discoveries were made. These areas are *Anadarko*, *Appalachia*, *Bakken*, *Eagle Ford*, *Haynesville*, *Niobrara*, and the *Permian Basin*.[2] Economic growth was particularly notable in Anadarko, Bakken, Eagle Ford, and the Permian Basin; while we study all the main shale areas, we direct extra focus on these four "boom town" areas.

We tell our story in two stages. First, we provide a descriptive (yet original) portrayal of the shale areas during the shale boom in terms of the evolution of aggregate activity and industry composition. The shale boom sparked broad-based employment growth, though growth was particularly strong in industries providing goods or services that supply or complement the output of the oil and gas industries; from the perspective of these supply-or-complement industries (and others), the shale boom is a large demand shock. A critical element of the description we provide is a parsimonious but powerful analysis of the relative roles of existing establishments and new establishments in accounting for aggregate employment growth. A significant share of aggregate employment growth occurred through the net addition of new establishments.

Second, we complement our descriptive findings with a more rigorous analysis to better understand the consequences of the shale boom. The shale areas differ from the average U.S. county in important ways, and the coincident nationwide downturn of the Great Recession contaminates simple descriptive analysis. Using rich longitudinal business microdata from the Census Bureau, we implement a difference-in-differences research design: using propensity score matching we construct a control group of counties that are, *ex ante*, similar to the shale counties, and we compare the shale "treatment" group to these controls as the shale boom occurred. This exercise, which relies on the plausibly exogenous interaction of shale technology improvements with preexisting geological traits of specific regions,

[2]These areas are defined by EIA (2019); see Figure 1.

yields estimates of the causal effect of the shale boom on county employment growth. Most importantly, we decompose these estimates into contributions from existing business establishments, new greenfield establishments of existing firms, and entirely new firms.

In our causal analysis, we find that—consistent with our descriptive exercises—new business establishments played a disproportionate role in the employment growth caused by the shale boom, even outside of the oil and gas mining sector. New firms and greenfield establishments of existing firms made similar contributions at an annual frequency, each accounting for between one fifth and one third of annual employment growth. Consistent with existing literature on early lifecycle dynamics of firms, the strong contribution of new firms continues during their first few years of existence such that the overall role of entry for cumulative aggregate growth is enhanced; relative to a plausible counterfactual, seven years after the onset of the oil and gas boom new firms account for almost half of total cumulative jobs created. More broadly, the results highlight the importance of entrepreneurship and the extensive margin of the firm distribution for studying economic fluctuations—emphasizing the relevance of firm heterogeneity and dynamics—while also highlighting the distinction between new firms and greenfield establishments, a distinction that is frequently glossed over in formal treatments.

Sectoral analyses reveal further insights into firm dynamics. In the oil and gas mining sector where employment gains were largest, new firms account for more than one third of the overall employment growth response, and the expansion of existing establishments accounts for the majority of the remainder (with greenfield establishments not playing a significant role). In the complementary construction, transportation, and warehouse industries, new firms account for about one fifth of the employment growth while greenfield establishments of existing firms likewise contribute little. Moreover, we observe a strong relationship of total employment in the oil and gas sector with employment growth—and business entry—in other industries, consistent with the notion that oil and gas booms are associated with employment growth in a wide range of industries.

Our descriptive and causal analyses yield a rich story of boom town economics: As the shale boom struck local areas, new establishments opened in large numbers, transforming the local economic landscape. In some sectors, such as utilities, transportation and warehousing, professional and business services, and education and health services, new establishments were critical contributors to employment growth. Retail trade and leisure and hospitality grew more through expansion of existing businesses. Manufacturing, which may compete with oil and gas businesses for workers and materials, contracted through both net establishment closure and downsizing by existing establishments. Overall, the strong performance of the boom towns relative to a plausible counterfactual was facilitated by disproportionate resilience of firm and establishment entry. The shale boom remade these local economies such that the economic anatomy of the boom towns looked very different after the boom than it did just a few years earlier.

The importance of establishment and firm entry for the boom town growth experience was not theoretically inevitable. Admittedly, the affected counties tended to be small, with fewer workers and businesses than the U.S. average. One might therefore argue that the extensive margin was the *only* way these areas could have been expected to grow, but this would be "begging the question"—assuming that the growth must come from entrants because incumbents cannot sufficiently expand. These areas did have businesses before the boom, and those businesses could have, in principle, grown sufficiently to meet all the needs of the enlarged post-shale economy. It is not difficult to imagine models, such as representative firm models with perfect competition and constant returns to scale production, in which rapid growth of the existing business footprint is precisely what would occur in response to a positive aggregate shock. Whether this latter view or the more nuanced view afforded by richer models is most appropriate is an empirical question—one that we can answer in our quasi-experimental setting. The insights of those richer models are not inherently limited to rural or sparsely populated economies. While generalizing from any specific experiment must always be done with caution, our results are strongly suggestive about the importance of the business entry margin.

2 Theory and Relevant Literature

2.1 The Role of New Businesses in Employment Growth

Models of representative firms[3]—often characterized by perfect competition and constant returns to scale production—give rise to intuition in which economic shocks are accommodated entirely by homogeneous existing firms that scale up or down as necessary. In contrast, models of firm heterogeneity allow for a more realistic firm distribution with entry and exit. Since our contribution is empirical, not theoretical, we do not explicitly describe such a model here; but in Appendix A, we provide a mathematical sketch of a canonical model of this nature for illustrative purposes. A common way to create this more realistic environ-

[3]For the purposes of our *theory* discussion, we use the terms "firm" and "establishment" interchangeably, since standard models do not distinguish between the two. In model terms, the focus is on productive units.

ment is to impose curvature on firms' revenue function, either due to decreasing returns to scale production technology (e.g., through span-of-control limitations as in Lucas (1978)) or to imperfect competition (e.g., through product—or, perhaps, geographic—differentiation).[4] When facing such revenue function curvature, the responsiveness of businesses (in terms of, e.g., employment growth) to profitability shocks is dampened, such that the stock of existing firms do not face strong incentives to grow when faced by positive aggregate shocks. Intuitively, businesses do not want to expand infinitely but instead face constraints on their ability and incentive to grow within a single physical location.

Moreover, models with entry and exit typically use some version of a free entry condition that links the value of even incumbent firms to entry costs; that is, in the face of a positive aggregate shock, entrepreneurs enter the market until the value of operating a firm is driven down to the entry cost. Intuitively, the incentive to create a new business is determined by the amount of revenue available to the market generally, and the value of existing firms is constrained by the threat of entry. In such an environment—one characterized by revenue function curvature and some sort of free entry condition—existing businesses do not grow enough to fully accommodate aggregate shocks, so the resulting increase in aggregate production depends also on growth in the *number* of firms, including through increased entry.[5] In other words, richer models facilitate a role for discrete geographic expansion of existing firms (through greenfield establishments) and/or new business formation by entrepreneurs in macroeconomics.

While the model sketch described above (and detailed in Appendix A) is extremely stylized, the basic intuition can hold in richer, more realistic settings. An important example is Clementi and Palazzo (2016), which explores an enhanced firm dynamics model calibrated to establishment dynamics data and consisting of capital (and capital adjustment costs), a richer distribution of recent entrants, and explicit business cycle dynamics. In their general equilibrium formulation, potential entrants receive advance signals about their productivity and enter when the expected discounted profits from doing so exceed entry costs. Incumbent firms face decreasing returns to scale. Positive aggregate shocks (whether demand or technology) induce significant entry activity for the reasons sketched above.

Importantly for our purposes, in the more fully specified model of Clementi and Palazzo (2016) not only do aggregate shocks boost the share of activity accounted for by entrants,

[4]Other options for generating a firm distribution include static distortions, matching frictions, factor adjustment costs, and heterogeneity of factor prices.

[5]Karahan et al. (2018) make this point eloquently in a balanced growth perspective, showing that under certain assumptions the bulk of labor demand adjustment is accommodated on the entry margin.

but also surviving young firms grow quickly on average such that each cohort of entrants has persistent effects on aggregate employment in early years.[6] The reason for the rapid growth of young businesses is that many young businesses have high marginal products due to the cost of obtaining and adjusting their capital stock. Our paper can be thought of, in part, as a reduced form empirical study of models like Clementi and Palazzo (2016), but our contribution has other dimensions since in reality the prevalence of high marginal products among recent entrants is likely limited to young *firms*, while young *establishments* of existing firms are more likely to enter fully capitalized. One contribution of our paper is to demonstrate the differing dynamics of new firms and greenfield establishments in response to economic shocks, showing that the choice to calibrate firm dynamics models to establishment data is not benign.[7]

Broadly speaking, the theoretical insight that potential entrants and young firms respond disproportionately to shocks is consistent with a large empirical literature finding that new entrants play a disproportionate role in job creation. Decker et al. (2014) note that new entrants account for nearly one fifth of gross job creation annually in the U.S. despite accounting for less than 10% of firms and less than 5% of employment. Haltiwanger et al. (2013) show that the job creation benefits often attributed to small businesses are more accurately attributed to new businesses, inspiring our focus on new firm activity. These insights hold over the business cycle as well; for example, Fort et al. (2013), Pugsley and Sahin (2015), and Sedlacek and Sterk (2017) show that young firms are more cyclically sensitive than older firms. Bernstein et al. (2018) find that new and young firms also respond strongly to commodity price shocks in Brazilian data. Adelino et al. (2017) study local economic shocks and find that new firms disproportionately account for the response of both net and gross employment. One key contribution of our work relative to existing empirical literature is our ability to distinguish between new firms and greenfield establishments.

New firms have often been treated as synonymous with "entrepreneurship" in much relevant literature, largely due to the importance of business age for key job creation and productivity results (Decker et al. (2014), Haltiwanger et al. (2013)). Other concepts of entrepreneurship have been studied in relation to energy booms, however. Gilje and Taillard (2016) examine investment by public and private firms in the natural gas industry and find that publicly traded firms are more responsive to changes in investment opportunities than

[6] This is broadly consistent with Decker et al. (2014), who show that, despite high failure rates of young firms, typical firm cohorts retain 80% of their initial employment impact after five years.

[7] Clementi and Palazzo (2016) is just one of several recent modeling exercises based on an establishment concept; other recent examples include Moreira (2017) and Lee and Mukoyama (2015).

private firms, a finding that may be thought of as contrary to the view that new firms are key but may be supportive of our findings on greenfield establishments. Tsvetkova and Partridge (2017) document limited positive impacts to self-employment (i.e., including nonemployer self-employment) in energy boom towns in 2001–2013 using American Community Study (ACS) data, consistent with previous evidence that resource sector booms may crowd out entrepreneurial activity (Davis and Haltiwanger, 2001; Glaeser et al., 2015; Betz et al., 2015). Our data cover employer businesses, so nonemployer self-employment is outside the scope of our study. We therefore view our work as complementary to Tsvetkova and Partridge (2017) as we add the employer-business side of entrepreneurship, which likely has a stronger association with later economic growth but has somewhat different interpretations in terms of the entrepreneurial occupational choice.[8] In this respect, we add employer entrepreneurship to the list of economic outcomes that have been studied in relation to resource booms, a literature that we review next.

2.2 Economic Effects of Oil and Gas Booms

Beginning with the advent of the modern oil industry in 1859, for a century the U.S. experienced consistent increases in oil production. But in 1970, this age of increasing domestic production reached its end and production began a period of decline that continued for the next four decades. However, during recent years the oil landscape has changed both suddenly and dramatically as illustrated in Figure 2. By 2007, after a long period of declining U.S. production, a technological breakthrough allowed "shale" oil and gas extraction to become economically viable for the first time in history; the "shale boom" was underway.[9]

The advent of horizontal drilling and hydraulic fracturing fundamentally transformed the oil and gas industry such that both oil and natural gas production are currently at record levels. In addition to the innovations that first made shale production economical, subsequent innovations have generated dramatic gains in output and, therefore, magnified the economic consequences of the shale boom (see Decker et al. (2016)). Some innovations mitigate the considerable costs of assembling and disassembling drilling rigs, which can take multiple days and require dozens of heavy trucks; these include pad drilling (in which a rig

[8] Poschke (2018) provides a model with a clear distinction between employer businesses and nonemployer ("own account") self-employment, using cross-country evidence to argue that nonemployer self-employment is best thought of as an occupational choice in response to weak labor markets. It would therefore be unsurprising to find that nonemployers and employers respond differently to shocks.

[9] For the main empirical specifications in this research, the shale boom will begin in 2007 consistent with the time that EIA began tracking shale production (EIA, 2019). We will consider the specific timing of the treatment in an alternative specification.

drills multiple wells from a single spot) and "walking rigs" (which complete one well then transport themselves a few dozen feet to drill the next well). Other innovations increased output by boosting the productive capacity of wells, such as increases in well length, changes in the mix of water versus sand and other proppants in the hydraulic fracturing process, and recompletions of existing wells. Still others improved productivity by improving well site selection and design, such as improvements in the computing tasks associated with exploration. The result of these various innovations is that shale production has become economical under a wide range of market circumstances, including far lower world market prices than was the case two decades ago. It is the interaction of these technological improvements with the preexisting geological characteristics of the shale regions that we use to identify economic booms and the specific role of entry.

Unsurprisingly, a growing body of work quantifies the economic effects of localized natural resource based booms. While this literature began before the specific shale oil and gas booms of this past decade (Black et al., 2005; Allcott and Keniston, 2014), this new era of shale has created a significant resurgence in this literature in part because of the clean empirical identification afforded by the nature of the shock.

Feyrer et al. (2017) finds that the shale boom specifically created significant economic shocks to local labor markets. Every million dollars of oil and gas extracted is estimated to generate $243,000 in wages, $117,000 in royalty payments, and 2.49 jobs within a 100 mile radius. In total, the authors estimate that the shale boom was associated with 725,000 jobs in aggregate and a 0.5% decrease in the unemployment rate during the Great Recession. Marchand (2012) similarly finds both direct and indirect impacts of the shale boom on employment; for every 10 jobs created in the energy sector, 3 construction, 4.5 retail, and 2 services jobs are created. Agerton et al. (2016) find that one additional rig results in the creation of 31 jobs immediately and 315 jobs in the long-run. Other studies corroborate the positive impact of the shale boom on local labor markets (Weber, 2012; Marchand, 2012; Komarek, 2016; Bartik et al., 2019; Upton and Yu, 2017; McCollum and Upton, 2018; Unel and Jr., 2020).[10] While positive effects associated with the economic activity spurred by drilling and production have been documented extensively, negative effects might also be observed, specifically in the manufacturing sector (Cosgrove et al., 2015; Freeman, 2009).[11]

[10] Due to the oil and natural gas price declines of 2014, there is also an emerging literature on the "bust" side of the cycle that will likely grow in upcoming years. For instance, Brown (2015) finds that elimination of each active rig eliminates 28 jobs in the first month and this increases to 171 jobs eliminated in the long-run.

[11] To be clear, we are interested in short-term boom town effects, in contrast to the large literature on resource endowments and long run economic growth (Sachs and Warner, 2001; van der Ploeg, 2011; Venables, 2016; Alexeev and Conrad, 2009; Michaels, 2010; Smith, 2015; Oliver and Upton, 2019).

Our work adds to this growing body of literature in that ours is the first study to investigate the margins by which the business sector adjusted to the shale boom, directly tying the event to broader questions in firm dynamics and macroeconomics. We find that establishment entry accounts for a disproportionate share of the increases employment growth in shale regions, and new firms (i.e., those born during the shale boom) account for nearly half of total employment gains.

3 Data

We first provide two critical definitions: for our purposes (and consistent with U.S. Census Bureau definitions), an *establishment* is defined as a specific business operating location, while a *firm* is a group of establishments under common ownership or operational control.

We focus on two data sources for our analysis. First, for our descriptive analysis, we use the Census Bureau's County Business Patterns (CBP) dataset, a publicly available annual tabulation of employment, payroll, and establishment counts at the county-by-industry level (with data recorded as of the pay period including March 12 of a given year). CBP allows us to paint a broad picture of the industry and establishment dynamics that followed the shale boom, and its public availability affords flexibility in the number and nature of calculations we can perform. The CBP is based on the Census Bureau's Business Register (see DeSalvo et al. (2016)) and covers the near-universe of private nonfarm business establishments in the U.S. We provide more detail about CBP and, in particular, how we address the problem of disclosure avoidance data censoring in some industry-by-county cells of CBP, in Appendix B.

For our causal analysis we use the Census Bureau's Longitudinal Business Database (LDB), which consists of longitudinal establishment-level microdata covering almost all private non-farm businesses in the U.S. (see Jarmin and Miranda (2002) for extensive detail on the LBD). Like CBP, the LBD is based on the Census Bureau's Business Register, and the two datasets have the same industry scope. Unlike CBP, however, LBD data are confidential, require special sworn status for access, and feature limitations on the number and nature of calculations we can report. But the LBD microdata yield a number of benefits relative to CBP. The LBD provides annual data on establishment location and detailed NAICS industry as well as annual employment counts (also corresponding to the pay period including March 12); importantly for our purposes, the LBD provides *firm* identifiers that allow us to link establishments together as firms and to track firm age; see the appendix for more detail.[12]

4 Describing the shale boom

What happened to shale areas during the shale boom? In this section we *describe* how shale county economies evolved during the boom. Throughout the paper we define the "shale boom" as comprising the years 2007 to 2014; we initiate the boom in 2007 to be consistent with some other literature and because 2007 appears to mark the beginning of significant shale-related expansions in many shale areas. We end our analysis in 2014 at the peak of shale activity; starting in mid-2014, oil prices declined and shale activity slowed until a recovery began in mid-2016 (we leave study of the 2014-2016 "shale bust" to future research). While the shale economies have been described in much previous literature, here we focus particularly on the evolution of industry composition and the relative roles of establishment growth and entry in facilitating shale county employment growth.

4.1 The pre-boom period

We first characterize the shale counties prior to the boom. Panel A of Table 1 reports average employment, establishment counts, and prevalence of oil and gas mining activity for shale and non-shale counties for the 2000-2006 pre-boom period.[13] During 2000-2006, shale counties had total private nonfarm employment of about 21,000, on average, compared with about 38,000 for non-shale U.S. counties. Counties in the four "boom town" areas (Anadarko, Bakken, Eagle Ford, and Permian Basin) were even smaller, with about 9,000 employees on average. The fourth line of the table, "Non-shale control set," refers to non-shale counties available in the set of counties from which our control group will later be drawn. We discuss this restriction more below (Section 5.1), but the group omits counties that are in the same state as, or in a state adjacent to, the control counties; these counties are just slightly larger than U.S. non-shale counties generally. Importantly, while shale counties (and boom town

[12] The LBD is the premier source of business microdata for the U.S., with its detailed longitudinal establishment information and its ownership-based firm concept and linkages. An alternative data source is the publicly available National Establishment Time Series (NETS) based on Dun & Bradstreet data; Crane and Decker (2020) document the limitations of NETS for studying business dynamics.

[13] We do not employ our specific propensity score control groups in this section; statistics for non-shale counties refer to all counties in the U.S. except shale counties and those counties disqualified from our control groups due to adjacent to shale areas. Due to disclosure requirements, we are not able to report the specific counties used in the propensity score matched control group. However, special sworn researchers with approved projects in Federal Statistical Research Data Centers can obtain access to the data use.

counties generally) are smaller than other counties in the U.S., their *establishment* counts are still nontrivial. Boom town counties had, on average, 700 establishments during the 2000-2006 pre-boom period; in principle, there is no reason these existing establishments could not expand sufficiently (in terms of product variety and output) to accommodate the economic boom that followed.

The third and fourth columns show that oil and gas mining activity—establishments classified as NAICS 211 or 213—was not a dominant industry in any counties but did account for a nontrivial share of activity in the pre-boom period, comprising 3.5% and 6.8% of employment in shale counties generally and boom towns, respectively, compared with 1% or less in other counties.

Panel B of Table 1 shows wide variation between the shale plays. Appalachia, which includes metro areas in eastern Ohio and western Pennsylvania, and Niobrara, which includes the Denver metro area, have the largest counties on average, while Bakken, the Permian Basin, and Eagle Ford consist of largely rural counties. Yet even Bakken, with average employment of 3,000, still had 310 establishments on average during the pre-boom period. Oil and gas mining activity was more prevalent in shale areas than the U.S. generally for all plays but Appalachia, with Anadarko and Permian Basin having significant petroleum extraction industries.

Table 2 reports establishment counts by broad sector. Consistent with oil and gas mining activity shares reported in Table 1, Table 2 shows that shale counties tend to have far more mining establishments (NAICS 21) than counties elsewhere in the U.S. No sector appears to be at risk of having no establishments, even in shale counties. Interestingly, the sector that tends to have the lowest establishment counts, utilities (NAICS 22), has a similar average establishment count in shale and non-shale counties. Both shale and non-shale areas have plenty of establishments in industries that are economically distant from oil and gas extraction, such as retail trade (NAICS 44-45) and education, health, leisure, and hospitality services (NAICS 61-72). Again, the shale areas had numerous establishments in the pre-boom period that could, in principle, accommodate economic expansion through organic growth.[14]

4.2 The boom period

The shock that struck shale counties during the shale boom is evident in the sharp rise in oil and gas mining activity during that period, which can be seen on Figure 3 (where employment and establishment counts in oil and gas mining are shown relative to year-2006 levels). While

[14]Table A1 in the appendix reports employment by sector and shale play for 2000-2006.

oil and gas mining activity in non-shale counties moved down then flattened out after 2007 (coinciding with the Great Recession), activity remained elevated in shale areas then rose sharply after 2010; this rise is manifest in both employment and establishment counts and is particularly notable in the boom town areas. Between 2006 and 2014, oil and gas mining employment in boom towns rose by about 150% while establishment counts rose about 50%.

The surge in economic activity was not limited to oil and gas drilling and extraction, however. Figure 4 reports employment and establishment counts for all industries except oil and gas mining. While activity in the shale areas does appear to have been affected by the nationwide recession, the decline in employment and establishment counts was shallower in the shale areas than elsewhere. Moreover, activity in the boom towns rebounded rapidly and exceeded the pre-recession peak by 2012, whereas activity in nonshale areas had yet to recover by 2014. Boom town employment grew by more than 10%, on net, from 2006 to 2014, and boom town establishment counts grew by nearly 10% over the same period. In short, overall economic activity—not just oil and gas drilling and extraction—evolved very differently in the shale areas, and particularly in the boom towns, relative to the rest of the U.S. This economic boom is the primary object of our study.[15]

The rapid post-2006 growth of employment and establishment counts in boom towns reflects considerable underlying heterogeneity across sectors. Figure 5 reports post-2006 growth of employment and establishment counts by broad sector. Unsurprisingly, average county-level mining employment grew substantially—by more than 150% from 2006 to 2014—while establishment counts in the sector grew by more than 50%. The only sector to see an employment *decline* was manufacturing. Aside from mining, the largest gains occurred in sectors providing significant inputs to shale activity: construction (employment and establishment gains of 29% and 7%, respectively) and transportation and warehousing (51% and 40%).[16] Even aside from these shale-adjacent sectors, nontrivial employment gains were seen in utilities (16%), retail trade (12%), professional and business services (7%), education and health services (11%), leisure and hospitality services (22%), and other services (4%). Retail trade saw a modest decline in establishment counts despite employment gains, while professional and business services saw larger gains in establishment counts than in employment.

Note that the employment changes reported on Figure 5 are in absolute terms (i.e., not relative to some other group of counties), reflecting actual employment growth in these

[15]The pattern of employment gains varies widely across plays, as shown on Figure A1 in the appendix.

[16]Note that the transportation and warehousing sector includes both pipelines and the many trucks required for drilling and fracking operations, though railroads (NAICS 482) are out of scope for both CBP and the LBD.

counties. These gains occurred against the backdrop of a nationwide recession in which some sectors saw dramatic employment declines (most notably construction and manufacturing, but also retail trade and other services). Figure A2 in the appendix repeats this exercise for non-shale counties for comparison. From the descriptive perspective of this section, the absolute gains in employment are critical for understanding the actual experience of shale counties; but the comparison to other parts of the country highlights the importance of the more rigorous causal analysis that follows. Before we move to that analysis, however, we now describe simple exercises that more concretely demonstrate the importance of the net establishment entry margin for understanding the growth patterns described above.

4.3 Accounting for net establishment entry

In a strictly descriptive sense, we can assess the importance of the establishment entry margin for the growth described above using a simple accounting decomposition. For brevity we will present results on the boom town plays as a whole. Let N_t be the number of establishments in year t, and let $\overline{emp}_t$ be average establishment size (employees per establishment) in year t. Consider the change in total employment between year 0 and year T, $N_T\overline{emp}_T - N_0\overline{emp}_0$. It can be easily shown that

$$N_T\overline{emp}_T - N_0\overline{emp}_0 = (N_T - N_0)\overline{emp}_0 + N_0(\overline{emp}_t - \overline{emp_0}) + (N_T - N_0)(\overline{emp}_T - \overline{emp}_0) \quad (1)$$

The first term on the right-hand side is the change in total employment accounted for by the change in the number of establishments (holding establishment size constant). The second term is the change in total employment accounted for by the change in the average establishment size (holding the number of establishments constant). The third term is a covariance term which, in practice, is relatively small. It is straightforward to calculate the share of total employment change that is accounted for by each of the components. Since the covariance term is small, for simplicity we distribute it proportionally among the other two terms.[17]

Figure 6 reports the result of this accounting exercise for the boom town plays in industries. The solid red line reports the total change in employment between 2006 and any given year ($N_T\overline{emp}_T - N_0\overline{emp}_0$). The dashed green line reports the change in total employment in which establishment size is held constant ($(N_T - N_0)\overline{emp}_0$); this line indicates the portion of

[17] In all industries excluding oil and gas mining, the covariance term is less than 3%, while it is 24% in the oil and gas mining sector.

employment growth accounted for solely by the change in the number of establishments. The left panel of the figure corresponds to all industries excluding oil and gas mining, while the right panel reports results for oil and gas mining. Among all industries, excluding oil and gas mining, the cumulative employment gain from 2006 to 2014 was about 13%; the employment gain accounted for by the changing establishment count was about 9 percentage points, or roughly two-thirds of the total employment gain. Even in oil and gas mining, where the establishment margin is less important, net establishment entry accounted for almost half of total employment gains. While simple, this exercise is a key contribution of our work and demonstrates the outsized role of the establishment margin for facilitating overall growth.[18]

5 Causal Methodology

We now move from descriptive exercises to an empirical design for estimating and decomposing the causal effect of the shale boom on margins of business growth and entry. For these exercises we rely on LBD data (see Section 3) to compare the shale areas with our propensity matched control groups.

5.1 Treated and Control Areas

The U.S. Energy Information Administration provides monthly data and analysis for regions defined by the agency as shale plays (for example, EIA (2019) is the January 2019 Drilling Productivity Report; each report is accompanied by corresponding data on rigs and output as well as a list mapping specific counties to specific shale plays). Following EIA, we classify counties in the *Anadarko*, *Appalachia*, *Bakken*, *Eagle Ford*, *Haynesville*, *Niobrara*, and *Permian Basin* plays as treated areas. Figure 1 shows a map of where these shale plays are located. For many of our exercises, we focus on a more limited group of four plays: Anadarko, Bakken, Eagle Ford, and Permian Basin. We refer to these plays as the "boom towns," and we focus on them specifically because these four plays experienced the most significant growth of overall employment relative to plausible counterfactuals.

We conduct our main exercises on all counties in all shale areas combined; in other exercises we also study the boom towns as a group as well as each play individually. First, we define a set of control counties that represent a plausible counterfactual. We construct a set of control groups through propensity score matching. The variables on which we match

[18] Figure A3 in the appendix reports this exercise for each NAICS sector.

are cumulative 2000-2006 employment growth as well as 2000-2006 averages of total county employment, the share of firms in the county that are new, the share of employment in the county that is at new firms, the share of employment in the county that is at greenfield establishments of existing firms, and the share of employment in the county that is at oil and gas establishments (NAICS 211, 213, 324, and 325) and construction, transportation, and warehousing establishments (NAICS 23, 48, and 49).[19] In this way, we construct a control group that is similar to the treatment group in terms of new firm activity, greenfield establishment activity, and activity of the oil and gas and related industries in the pre-shale time period. In other words, for each treatment county we find a (single) corresponding "control" county that has similar patterns of business dynamics *ex ante*. We construct a control group for the all-plays treatment group and another control group for the boom towns treatment group, then we create separate control groups for each play individually (i.e., for regressions that include only the boom towns or a single specific play, we redraw the control group with replacement). We will show that these results are robust to placebo tests and alternative control groups.[20]

To reduce the risk of our results being contaminated by spillover effects, counties that are in states with shale activity but that themselves are not included in EIA-defined shale plays were removed from the list of potential control counties.[21] In addition, states that directly border counties with shale activity were removed from the potential control group.[22]

As background, we describe key variables in the treatment and control groups, both before and during the shale boom, in Table 3. We note two important items. First, average employment in the shale counties (i.e. treatment group) increased only modestly (by about 2.4%) between the pre-shale and post-shale time periods, while control counties actually experienced a 4% decline in employment. This is unsurprising given the fact that the United States was in the midst of the Great Recession during the early years of the shale boom.

Second, new firm employment as a share of total employment and young firm employment as a share of total employment declined substantially in both shale and non-shale areas.

[19]Importantly, we do not match on our specific outcomes of interest—the share of employment growth accounted for by business entry margins.

[20]See Appendix C.

[21]For a technical discussion of spillover effects on empirical estimates, see James and Smith (2020) and Feyrer et al. (2020).

[22]After applying these criteria, the potential control group comes from firms located in counties in the following 28 states: AL, AZ, CA, CT, DE, FL, GA, HI, ID, IL, IN, IA, ME, MI, MN, MS, MO, NV, NH, NJ, NC, OR, RI, SC, TN, VT, WA, WI. These criteria closely follow McCollum and Upton (2018) with the exception of Oklahoma, which includes the Anadarko play (but was included in the control group of McCollum and Upton (2018) as EIA had not yet defined the Anadarko play.

New firms' share of employment declined by more than 18% in shale counties and more than 30% in non-shale counties. A similar pattern is observed for young firm employment as a share of total employment, which declined by 16% and 23% in shale and non-shale counties, respectively. But for both new and young firm employment, shale counties experienced a 12 percentage point (and 7 percentage point, respectively) slower decline than non-shale counties. It is not surprising that new and young firm activity declined over this time period given the particular sensitivity of young firms to the business cycle documented in the literature described above; this fact highlights the importance of studying the shale boom with a carefully designed empirical strategy. For the U.S. as a whole, new firm employment was about 2.8% of total employment during 2000-2006 and 2.2% of total employment during 2007-2014, somewhat less than the share in our treatment and control counties.[23]

Next, Table 3 shows employment shares of greenfield establishments (i.e., new establishments of existing firms). Note that existing firms need not have existed in the county of interest beforehand; they could have activity anywhere in the U.S. Thus, some of these firms may have existed in other parts of the country and opened up a new establishment in the shale county. In shale counties, greenfield establishment employment as a share of total employment increased by 3.6%, while greenfield employment shares declined more than 11% in control counties. Thus, from these basic summary statistics, it appears that employment growth from new establishments of existing firms was particularly important during the shale boom and might potentially account for the lion's share of the employment growth. For the U.S. as a whole, greenfield establishment employment was about 3.3% of total employment during 2000-2006 (higher than in our treatment and control counties) and 2.5% of total employment during 2007-2014 (similar to our treatment and control counties).[24]

5.2 Difference in Differences

We employ a simple and intuitive difference-in-differences (DD) estimation strategy for measuring the effect of the shale boom on economic outcomes:

$$y_{ct} = \alpha + \delta(S_{Shale_c} \times Shale_t) + \tau_c + \gamma_t + \varepsilon_{ct} \qquad (2)$$

[23]Data for the total U.S. taken from the Business Dynamics Statistics (BDS), the public-use tabulations of LBD data.

[24]We calculate greenfield establishment employment for the U.S. as a whole as job creation by establishment births associated with firms with age greater than zero (BDS data).

where y_{ct} is the outcome of interest for county c in year t; in our main results this is an employment growth component described below, though we also report results for employment levels in background exercises. S_{Shale_c} is an indicator variable corresponding to shale counties as defined by EIA (i.e., the treatment group) and is zero for non-shale counties; and τ_c and γ_t are fixed effects for county and year, respectively. $Shale_t$ is an indicator variable that indicates the years during which shale activity occurred; all of the shale plays, and therefore the counties that EIA defines to have shale activity, saw increases in drilling starting around 2007, and this drilling activity continued until the end of 2014.[25] The coefficient δ gives the estimated causal effect of the shale boom on shale counties, controlling for aggregate temporal shocks as well as time-invariant differences across counties that remain after the propensity score matching process. For each model, we estimate standard errors clustered at the county level.[26] We do find evidence of common pre-treatment trends in our treatment and control counties, though we defer exploration of this important issue to our discussion of cumulative effects in section 6.4.

5.3 Outcomes: Growth Rates and Growth Components

We consider several outcome variables in the estimation described by equation (2). In initial background exercises, we study the effect of the shale boom on overall county employment levels (in logs). Our main outcome of interest, however, is annual employment growth. Consider the following growth rate concept:

$$g_{ct} = \frac{emp_{ct} - emp_{ct-1}}{0.5(emp_{ct} + emp_{ct-1})} \tag{3}$$

where c indexes counties, t indexes years, and emp_{ct} is total employment for county c in year t. The growth rate g_{ct} is commonly referred to as the "DHS growth rate" after Davis et al. (1996) and is widely used in the empirical firm dynamics literature; this growth rate concept has the desirable property of facilitating the inclusion of entry and exit. Now consider a

[25] Of course, the exact start time of the boom varies across shale plays. In the initial specification, we specify 2007 as the start date for the shale boom, but we also present year-specific estimated treatment effects by shale play in later investigations. We end the analysis in 2014 because global oil prices dropped immediately thereafter, and therefore the "bust" plausibly began in 2015. Therefore, 2007 to 2014 is the best general time period that can be considered the "boom" or "treatment" time period. We also note that our results are not materially different if we vary the "shale boom" cutoff year by one or two years in either direction.

[26] Our results are broadly robust to clustering by county and year.

related growth rate, commonly referred to as a growth component:

$$g_{ct}^{j} = \frac{emp_{ct}^{j} - emp_{ct-1}^{j}}{0.5(emp_{ct} + emp_{ct-1})} \tag{4}$$

where j indicates a grouping based on firm or establishment ages (and lack of superscript indicates inclusion of all groups). In the case of firms, $j \in J = \{\text{age 0}, \text{age 1-4}, \text{age 5+}\}$, where we define the categories as "new," "young," and "mature." In the case of establishments, $j \in J = \{\text{new firm}, \text{greenfield establishment}, \text{incumbent establistment}\}$. Defined in either way, it is straightforward to show that:

$$\sum_{j \in J} g_{ct}^{j} = g_{ct} \tag{5}$$

Hence, each g_{ct}^{j} is a growth "component" such that the components sum to the overall growth rate. This follows the approach of Adelino et al. (2017) and allows for ease of coefficient interpretation; moreover, for any group, g_{ct}^{j}/g_{ct} gives the share of aggregate (county) employment growth accounted for by group j.

The main outcomes of interest, then, are the share of annual employment growth accounted for by new firms, "young" firms (those with age 1-4), mature firms, greenfield establishments (of existing firms), and incumbent establishments of existing firms. The use of these growth components as dependent variables in our linear regression framework ensures that regression coefficients are additive in the way described above. Importantly, for the firm-based growth components we focus on "organic" growth as in Haltiwanger et al. (2013) and subsequent literature, in which the lagged employment term emp_{ct-1}^{j} is comprised of the lagged employment of all establishments in county c that belong to firms in group j in year t. This approach allows us to abstract from growth driven by merger and acquisition activity. In practice this means that the growth of an establishment that changes firm owners between years $t-1$ and t is assigned to the firm that owns the establishment as of time t.

5.4 Cumulative Effects

Following our main results for annual growth rates, we estimate regressions that will shed light on the roles of various types of businesses in the *cumulative* employment change at the county level. To do this, we construct the following outcome variable:

$$e_{ct} = \frac{emp_{ct}}{emp_{c2006}} \tag{6}$$

where e_{ct} is employment in county c in year t relative to employment in county c in the year 2006. We again create a group-specific version of this variable:

$$e_{ct}^{k} = \frac{emp_{ct}^{k}}{emp_{c2006}} \tag{7}$$

where, we emphasize, k is defined differently from the j-indexed groups described above. In particular, we focus on three k groupings: (1) establishments that entered in year 2006 or before; (2) establishments that entered after 2006 belonging to firms that existed as of 2006 or before; and (3) establishments that entered after 2006 belonging to firms that entered after 2006. That is, for any year t, e_{ct}^{1} gives county c employment of establishments that were incumbents as of year 2007; e_{ct}^{2} gives county c employment of establishments born after 2006 to firms that were incumbents as of year 2007; and e_{ct}^{3} gives county c employment of establishments born after 2006 to firms born after 2006 (and these firms could have been born in any county in the U.S.). In each case, employment is expressed relative to year-2006 total county employment; therefore, the following convenient condition holds:

$$\sum_{k \in \{1,2,3\}} e_{ct}^{k} = e_{ct} \tag{8}$$

Moreover, note that $e_{ct}^{2} = e_{ct}^{3} = 0 \; \forall t \leq 2006$ by construction. We choose the year 2006 consistent with our assumption above that the shale boom began in 2007. The general purpose of this set of dependent variables is to study, for any given year after 2006, how much of the cumulative (post-2006) employment growth in a county is accounted for by establishments that existed prior to the boom, establishments born after the boom to firms that existed before it began, and firms born after the boom. This provides an alternative view of the role of the business entry margin in driving aggregate employment that does not depend on single-year growth rates and allows time for early lifecycle dynamics to play out.

To study these outcomes, we generalize our difference in difference strategy as follows:

$$e_{ct}^{k} = \alpha + \delta_{t}^{k} \times S_{Shale_c} + \gamma_{t}^{k} + \varepsilon_{ct}^{k} \tag{9}$$

where δ_{t}^{k} is the year-specific estimated treatment effect for firms in a given group k, and we abuse notation slightly to include the overall group of all establishments as one of our k groups. Note that we omit county fixed effects in this specification since they are a linear combination of included variables; but recall that our control counties are chosen to be similar to our treatment counties in the pre-2007 period, and employment is scaled by 2006 county employment. The difference of means generated by δ_t^k compares shale counties to control counties in any given year, controlling for aggregate shocks affecting all counties. Conveniently, the set of estimated δ_t^k for each of the establishment groups $k \in \{1, 2, 3\}$ described above will sum to the δ_t associated with overall cumulative employment growth (relative to 2006) so that, again, we can easily calculate the share of aggregate employment growth accounted for by different types of establishments.[27] This set of specifications is useful not only because it facilitates the study of cumulative employment effects but also because it allows us easily to inspect the assumption implicit in our difference in differences framework: common pre-shock trends.

6 Empirical Results

6.1 Total Employment

Table 4 reports background results that simply show the effect of the shale boom on log total employment (i.e., setting y_{ct} from equation (2) equal to log employment) for all shale plays; we include all industries except for oil and gas mining (NAICS 211, 213) to focus on industries *responding to* the shale shock. We present the average treatment effect (corresponding to δ in equation 2). The shale boom is associated with a 6.9% increase in total employment relative to the control group; note that this is an average level comparing all years after the boom to all years before. The full group includes all shale plays in the study; we next break out treatment effects by play.[28] This effect varies in size and significance across shale plays, with the shale boom in the Eagle Ford region estimated to have the largest effect on employment, a 19.8% increase. Additionally, we define the "Boom Town" group to include all observations from Eagle Ford, Permian, Bakken and Anadarko, plays that saw substantial economic growth in the shale era. The shale boom is associated with an increase of 13.1% in total employment for these areas, almost double the effect for all plays combined.[29]

[27] In these exercises, we make no attempt to ensure that growth is "organic" since it is not clear how to interpret organic growth in this context. As such, however, the employment share of post-2006 new firms in any given year can, in principle, include employment of establishments that are older but were acquired by those new firms during the post-2007 period.

[28] In our causal empirical exercises we are unable to report play-level results for Bakken or Anadarko due to confidentially restrictions. However, these two plays are included in the "All" and "Boom Town" groups.

[29] We report employment level results by sector in Appendix C.

6.2 Employment Growth Rates

We now explore our main results by estimating equation 2 with employment growth rates and components as dependent variables. Table 5 reports results where employment growth is expressed in percentage points, again omitting oil and gas mining industries to focus on those industries that respond to the shale boom. First, note that the "Total" column, in which the dependent variable is the growth rate of aggregate (county) employment, is equal to the sum of columns 1, 2, and 4 or, alternatively, the sum of columns 1, 5, and 6. Column 3, which reports the growth component for all firms with age less than 5, is equal to the sum of columns 1 and 2.

Column 7 of Table 5 indicates that the shale boom is associated with a 1.829 percentage point increase in annual employment growth rates at the county level. This is a strong effect but is not surprising in light of the results for log employment just described, which found a 7% increase in the average employment *level*. Column 5 shows that greenfield establishments (new establishments of existing firms) account for 0.34 percentage point of the overall increase; that is, greenfield establishments account for about 18.6% of the increase in net employment growth rates. This is similar to the 0.321 percentage point increase (or 17.6% of the employment growth) in annual employment growth rates contribution (column 1) of new firms. Incumbent establishments account for the remaining growth, slightly less than two thirds of the total increase in employment growth rates. Another way to interpret the 1.829 percentage point of growth is as the sum of new firms, young firms, and mature firms. The contribution of new firms (0.321) and young firms (0.359) together accounts for about 37.2% of the total growth rate. Mature firms, those aged 5+ years, account for 1.149 of the 1.829 percentage points of growth, 62.8% of the total employment growth.

We repeat this exercise using only the "Boom Town" subsample. In these results we see an even greater role for new and young firms, as well as greenfield establishments. For this group, the shale boom is associated with a 3.864 percentage point increase in annual employment growth rates at the county level. Column 5 shows that greenfield establishments contribute 0.709 percentage points (18.3% of the total growth rate), similar to the full sample. However, in the boom towns we find a much larger contribution for new firms (1.182 percentage points or 30.6% of the total growth rate). Together new firms and greenfield establishments contribute just shy of 50% of the total growth in the boom town sample, as compared to approximately one-third in the full sample. Breaking out results by firm age, we find similar results. Together, new and young firms account for 1.854 percentage points of employment growth rate, 48% of the total growth rate. Mature firms (column 4) account for the other

52% of the growth.

The evidence points to an important role for new and young firms. On the one hand, it is important not to understate the role of incumbent firms. By no means do young firms account for the majority of the employment growth response. However, the contribution of new firms and young firms generally is significantly disproportionate relative to their typical share of activity levels (each accounting for less than 5% of employment). Our results are not as dramatic as those found by Adelino et al. (2017), who find that firms age less than two account for 90% of the local employment growth response to local demand shocks, but our results are striking nonetheless showing an important role for new firms, young firms, and greenfield establishments in employment growth.

Among existing firms, employment growth is disproportionately facilitated by greenfield establishment formation. In part this may reflect firms based outside the shale areas newly entering the shale area by creating new establishments. More broadly a comparison of the greenfield establishment coefficient with the new firm coefficient highlights the importance of carefully distinguishing between the two when studying firm dynamics. While the effects are similar among all plays, they are markedly different in boom towns and, as we will show below, in a number of other specifications; more broadly, they reflect fundamentally different economic mechanisms. An incumbent firm, whether starting in or out of a shale play, opens new establishments using the resources of the firm, including supplier relationships, credit access, name recognition and customer base, and workforce. These establishments can enter larger with more upfront job creation than new firms that face particular barriers to credit access, labor market search and matching, upfront investment costs, and supplier and customer acquisition. As such, it is not surprising to see a strong employment role for greenfield establishments, and the disproportionate role for new and young firms is all the more striking in the firms in boom towns. We return to this topic in our discussion of cumulative employment gains.

Table 6 provides more color in terms of industry activity. We first report results for oil and gas mining, which are omitted from Table 5. We also report results for construction, warehousing, and transportation (NAICS 23, 48, and 49), which likely experience direct spillovers from oil and gas activity, and "all other industries" (i.e., all industries except NAICS 211, 213, 23, 48, and 49). The total growth rate effects are strongest in oil and gas mining, followed by the construction, transportation, and warehouse sectors that provide critical inputs to oil and gas activity. We find a smaller, but still economically and statistically significant, effect for the remainder of the sectors, consistent with our estimates of employment level effects above. In the oil and gas mining sector new and young firms account for about 45.8% of overall employment growth, similar to the 44.8% contribution of these firms in the construction, transportation, and warehouse sectors' growth rate, but higher than the 33.5% contribution of new and young firms to the growth rate in all other industries combined. This relationship is similar when only the boom town sample is used.

6.3 Extensions and Robustness

6.3.1 Accounting for Shock Magnitude

The size of the shale boom shock varied across plays and over time within plays. Our differences-in-differences estimate does not account for this heterogeneity. We can therefore gain more insights into the results from Table 5 by allowing effects to vary by the size of the oil and gas boom. For simplicity, we do this by regressing employment growth components (for industries excluding oil and gas mining) on county-level oil and gas mining employment (in logs); while these regressions do not necessarily have a causal interpretation, they do directly relate the shale boom "shock" to its consequences for non-shale industries (alternatively, one may think of this exercise as a way of scaling treatment effects by treatment intensity). That is, we estimate:

$$g_{ct}^{j} = \alpha + \beta \ln emp_{ct}^{211,213} + \tau_c + \gamma_t + \varepsilon_{ct} \qquad (10)$$

where $emp_{ct}^{211,213}$ is employment in NAICS 211 and 213, and fixed effects for county (τ_c) and year (γ_t) are included as before. We estimate this regression on the same sample as that used for Table 5; that is, we include both our treated counties and their matched control group. Table 7 shows the results; since the independent variable is *the log of* oil and gas mining employment, we interpret the total effect (approximately) as follows: a 10% increase in county-level oil and gas mining employment is associated with an increase of *annual* overall employment growth (outside oil and gas mining) of 0.06 percentage points; new firms and greenfield establishments each account for more than a quarter of this overall effect.[30] We view this as a sizeable effect since oil and gas mining employment ultimately grew by roughly 200% in the shale areas.

Comparing these results with Table 5, we find that the role of entry (both new firms and greenfields) is somewhat larger when we account for the magnitude of the shock to the oil and gas mining sector. When restricting the sample to boom towns, this effect is even more pronounced; new firms and greenfields each account for more than 40% of the total

[30]These results are broadly similar if we lag oil and gas mining employment by one year.

employment growth effect. In unreported results, we also find that the entry margin is more important when scaling our differences-in-differences indicator by play-level annual oil and gas revenue (though comparisons with rig counts produce mixed results).

The result that entry is more important when the independent variable (or treatment) is scaled by the size of the oil and gas mining shock is consistent with our theoretical discussion above. Business entry is highly (and disproportionately) responsive to aggregate shocks.

6.3.2 Job Destruction and Exit

A natural question is whether the strong entry responses we document occurred against a backdrop of higher job and business churning generally. In unreported results, we estimate our differences-in-differences specification with job destruction rates and establishment exit rates as dependent variables. Among all plays and among boom plays, we observe modestly negative but not statistically significant effects of the shale boom on both job destruction and establishment exit. In other words, the shale boom apparently did not raise overall job and business churn.

6.3.3 Other Specifications

In exercises that we do not report (to minimize Census Bureau disclosure burden), we find that our main results are not materially affected by including county size (in terms of employment) as a control variable in equation 2; that is, our results are not driven by heterogeneity in county size.

Our results are likewise unaffected by including 2000-2006 growth in FHFA county-level house price indexes as part of the criteria for our propensity score match; recent house price growth may be thought of as a proxy for preexisting financial conditions as well as a proxy for vulnerability to the housing crisis and Great Recession that occurred during the shale boom period. That said, our results are not robust to explicitly controlling for *contemporaneous* house price growth, which is likely to be highly correlated with business entry for reasons other than our topics of interest.

We specify the shale boom as occuring during 2007-2014; in reality, the exact timing of the shale boom is not precise and varies some across regions. Our main results are quantitatively robust to varying the "shale boom" cutoff year by one or two years in either direction; the results are also robust to dropping any specific year in 2000-2014 from the sample and to dropping consecutive pairs of years around the 2007 shale boom cutoff (which is close in timing to the 2007-2009 recession).

6.3.4 Robustness Checks

We describe two direct robustness checks in Appendix C. First, we discard the propensity score match and instead create 20 control groups by selecting counties completely at random (from the set of states that make up the control group choice set) then estimate our main regressions. The results are similar to our main results but differ in certain dimensions that suggest our propensity score match adjusts for importance sources of selection. Separately, we conduct placebo tests that help us rule out the possibility that our main results are due to random chance.

6.4 Cumulative Employment Growth

The foregoing results focus on annual employment growth contributions using annually based definitions of firm and establishment entry. An alternative approach is to focus on cumulative effects over time, as described in Section 5.4.

Table 8 reports effects on cumulative employment, by year and relative to county-level employment in 2006, as described in Section 5.4 and equation (9). Recall that for this purpose we discard the annually based definitions of firm and establishment entry used previously, instead focusing on establishments and firms created *during* versus *prior to* the shale boom broadly. Our discussion focuses on all shale areas, but results for boom towns are shown in Appendix Table A4. The results on Tables 8 and A4 are graphically reported on Figures 7 and A4, respectively; note also that for any given year, the coefficients in columns 1, 2, and 3 sum to column 4. It is important to recall that these specifications include year fixed effects such that coefficients indicate employment relative to control group counties; roughly speaking this is still a difference in differences approach where we compare treatment county employment relative to 2006 to control county employment relative to 2006. The results can be interpreted as growth of group employment between 2006 and a given year, comparing treatment and control counties.

First, consider column 4 of Table 8 (and boom town corollary Table A4), which reports the cumulative gain in total employment (in treatment versus control counties). Prior to 2007, total employment is flat and close to zero (and not statistically significant), lending support to the assumption underlying our main difference-in-differences result that treatment and control counties have similar pre-treatment trends. After 2006, total employment rises, becoming statistically significant in 2008. By 2014, total employment in treatment counties has risen 17.2% in all areas since 2006 (relative to controls). The results are even more striking when we examine the boom town group; by 2014 there is a relative increase of 36.5% in total employment in the treated counties since 2006.

In column 1 we present results for establishments that were born prior to 2007 (that is, these establishments were incumbents when the shale boom began). We find a positive and significant effect of the shale boom for these establishments from 2009 onward. For example, in the year 2009, we find that employment among these pre-2007 establishment cohorts has risen 4.56% (3.82% in boom towns) relative to total employment in 2006. This effect peaks at 6.67% in 2013 (12.74% in boom towns) before attenuating slightly to 5.47% in 2014 (11.09% in boom towns). If we divide the 2014 coefficient in column 1 by the 2014 coefficient in column 4, we find that these pre-2007 establishment cohorts account for about one third of the total post-2006 rise in employment in shale areas (relative to control counties). The remaining two thirds of the rise is therefore attributable to establishments born after 2006.

In column 2 we present results for greenfield establishments, that is, those opened in 2007 or later by firms that existed prior to 2007. We see a positive and significant result beginning in 2008 (an increase in 0.96% of 2006 total employment) that strengthen annually to the end of the sample in 2014 (5.47% relative increase). Similar patterns, but larger magnitudes are observed in boom towns. This net job creation among new establishments of preexisting firms accounts for about one quarter of the cumulative gain in total employment as of 2014.

In column 3 we examine the effect of the shale boom on employment in new firms, that is, firms started in 2007 or later. Roughly speaking, these are firms that were created after the shale boom began. In these results, we again observe a positive treatment effect in 2010, consistent with the fact that new firms tend to start small, but by 2014 this group has a larger relative increase in employment (7.6%) than either of the other two groups. This net job creation among post-2006 firms accounts for 44% of total shale area employment growth relative to the counterfactual.

As noted above, new establishments (either born to preexisting firms or new firms) account for about two thirds of the total employment gain. One other important implication arises from these results: while increased employment among new firms does not become statistically significant until two years after employment at greenfield establishments of preexisting older firms (2008 vs. 2010), new firm employment surpasses greenfield employment two years later in 2012. This is our most striking finding about the difference between new firms and greenfield establishments: new firms start smaller but grow rapidly, consistent with a theory in which greenfield establishments, born with the advantage of existing firm ownership, begin their lifecycle better capitalized or with a stronger customer base than do young firms. New cohorts of young firms grow rapidly, however, likely as a result of a few extremely fast growers as documented by Decker et al. (2014). An important implication for theory is that modelers should not conflate firms and establishments.

We graphically report these year-specific effects for all plays on Figure 7 and separately for each play in the appendix in Figure A4. The results do vary notably by play; Eagle Ford and Permian Basin look similar to the overall results described above. The gas-heavy plays—Haynesville and Appalachia—show small overall employment effects and a less-consistent story about firm and establishment entry. Modest preexisting trend differences between treatment and control groups are sometimes evident in these areas, though the differences are rarely statistically significant. In short, however, the cumulative results suggest that areas in which the shale boom generated large economic expansions saw an important role for entry, with new firms ultimately accounting for the largest share of activity gains.

7 Conclusion

The U.S. shale boom has given rise to a large literature studying the economic effects of natural resource shocks. We add to this literature by studying the effects of the shale boom on new firms and establishments, adding entrepreneurship and business creation to the list of economic outcomes that are stimulated by natural resource production (i.e., natural resource booms do not appear to only benefit existing business establishments). Waves of business formation transformed the economic geography of local economies.

Our results also have significant implications for the study of macroeconomics. In particular, a large literature in firm dynamics focuses on the role of new business creation in the response of the aggregate economy to broad economic shocks. We show that the growth of aggregate employment in response to the shale boom is disproportionately accounted for by new firms and new establishments of existing firms. At an annual frequency, new firms and greenfield establishments each account for about one fifth of overall employment gains resulting from the shale shock, while establishment entry accounts for about two thirds of cumulative gains throughout the boom period. This finding lends strong support to models of firm dynamics in which, under standard assumptions, the entry margin accounts for a large share of aggregate adjustment. Further, though, our results point to important differences between new firms and new establishments of existing firms (greenfield establishments). New firms appear to start small but, as a cohort, grow rapidly. New establishments of incumbent firms appear to start out larger, with a more gradual growth trajectory.

These differences between firms and establishments have important implications for theories of firm dynamics. New firms are likely more constrained than greenfield establishments in terms of initial investment costs and the challenges associated with building a workforce, establishing supplier relationships, and building a customer base (Moreira (2017), Foster et al. (2016)). But the importance of firms born after the boom increases over time such that new firms ultimately contribute more to cumulative employment gains. These results shed additional light on the dynamics of young businesses and their importance for aggregate adjustment, presenting important facts with which models of firm dynamics must grapple.

The disproportionate role of business entry also has implications for the measurement of economic activity. Workhorse surveys of businesses, such as the "payroll survey" of the BLS (Current Employment Statistics) or the Census Bureau's Monthly Retail Trade Survey (a critical input for GDP estimates) rely on continuing businesses to track changes in aggregate economic activity. Shocks to aggregate economic activity may be measured poorly in the absence of timely measures of business formation.[31]

[31] Importantly, the Census Bureau recently developed high-frequency, timely measures of business applications in the Business Formation Statistics.

References

Adelino, Manuel, Song Ma, and David T. Robinson, "Firm Age, Investment Opportunities, and Job Creation," *The Journal of Finance*, 2017, *72* (3).

Agerton, Mark and Gregory B. Upton, "Decomposing crude price differentials: Domestic shipping constraint or the crude oil export ban?," *The Energy Journal*, 2019, *40* (3), 155–172.

_ **, Peter. R. Hartley, Kenneth B. Medlock III, and Ted Temzelides**, "Employment impacts of upstream oil and gas investment in the United States," *Energy Economics*, 2016, *62*, 171–180.

Alexeev, Michael and Robert Conrad, "The Elusive Curse of Oil," *Review of Economics and Statistics*, 2009, *91*, 586–598.

Allcott, Hunt and Daniel Keniston, "Dutch disease or agglomeration? The local economic effects of natural resource booms in modern America," Working Paper, National Bureau of Economic Research 2014.

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

Bernstein, Shai, Emanuele Colonnelli, Davide Malacrino, and Tim McQuade, "Who Creates New Firms when Local Opportunities Arise?," Technical Report 2018.

Betz, Michael R., Mark D. Partridge, Michael Farren, and Linda Lobao, "Coal mining, economic development, and the natural resources curse," *Energy Economics*, 2015, *50*, 105 – 116.

Black, Dan, Terra McKinnish, and Seth Sanders, "The economic impact of the coal boom and bust," *The Economic Journal*, 2005, *115* (503), 449–476.

Brown, Jason P, "The response of employment to changes in oil and gas exploration and drilling," *Economic Review-Federal Reserve Bank of Kansas City*, 2015, pp. 57–81.

Clementi, Gian Luca and Berardino Palazzo, "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations," *American Economic Journal: Macroeconomics*, 2016, *8* (3), 1–41.

Cosgrove, Brendan M., Daniel R. LaFave, Sahan T. M. Dissanayake, and Michael R. Donihue, "The Economic Impact of Shale Gas Development: A Natural Experiment along the New York / Pennsylvania Border," *Agricultural and Resource Economics Review*, 2015, *44* (2), 20–39.

Crane, Leland D. and Ryan A. Decker, "Research with Private Sector Business Microdata: The Case of NETS/D&B," *Working paper*, 2020.

Davis, Steven J and John Haltiwanger, "Sectoral job creation and destruction responses to oil price changes," *Journal of Monetary Economics*, 2001, *48* (3), 465–512.

_ , John C Haltiwanger, and Scott Schuh, *Job Creation and Destruction*, The MIT Press, 1996.

Decker, Ryan A., Aaron Flaaen, and Maria D. Tito, "Unraveling the Oil Conundrum: Productivity Improvements and Cost Declines in the U.S. Shale Oil Industry," FEDS Notes March 2016. http://dx.doi.org/10.17016/2380-7172.1736.

_ , John Haltiwanger, Ron S. Jarmin, and Javier Miranda, "Changing Business Dynamism and Productivity: Shocks vs. Responsiveness," *American Economic Review*, Forthcoming.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda, "The role of entrepreneurship in US job creation and economic dynamism," *The Journal of Economic Perspectives*, 2014, *28* (3), 3–24.

DeSalvo, Bethany, Frank Limehouse, and Shawn D. Klimek, "Documenting the Business Register and Related Economic Business Data," Technical Report, Center for Economic Studies working paper CES-WP-16-17 2016.

Dismukes, David E., Dek Terrell, and Gregory B. Upton, "2020 Gulf Coast Energy Outlook," Technical Report, Louisiana State University 2019.

EIA, "Drilling Productivity Report: For key tight oil and shale gas regions," Technical Report, U.S. Energy Information Administration, https://www.eia.gov/petroleum/drilling/archive/2019/01/pdf/dpr-full.pdf 2019.

Feyrer, James, Erin Mansur, and Bruce Sacerdote, "Measuring Geographic Spillovers: A Reply," *American Economic Review*, 2020, *Forthcoming.*

__ , **Erin T Mansur, and Bruce Sacerdote**, "Geographic dispersion of economic shocks: Evidence from the fracking revolution," *The American Economic Review*, 2017, *107* (4), 1313–1334.

Fort, Teresa C, John Haltiwanger, Ron S Jarmin, and Javier Miranda, "How firms respond to business cycles: The role of firm age and firm size," *IMF Economic Review*, 2013, *61* (3), 520–559.

Foster, Lucia, John Haltiwanger, and CHad Syverson, "The Slow Growth of New Plants: Learning about Demand?," *Economica*, 2016, *83* (329), 91–129.

Freeman, Donald G., "The 'Resource Curse' and regional US development," *Applied Economics Letters*, 2009, *16* (5), 527–530.

Gilje, Erik P and Jerome P Taillard, "Do private firms invest differently than public firms? Taking cues from the natural gas industry," *The Journal of Finance*, 2016, *71* (4), 1733–1778.

Glaeser, Edward L, Sari Pekkala Kerr, and William R Kerr, "Entrepreneurship and urban growth: An empirical assessment with historical mines," *Review of Economics and Statistics*, 2015, *97* (2), 498–520.

Haltiwanger, John, Ron S Jarmin, and Javier Miranda, "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, 2013, *95* (2), 347–361.

James, Alex and Brock Smith, "Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution: Comment," *American Economic Review*, 2020, *Forthcoming.*

Jarmin, Ron S. and Javier Miranda, "The Longitudinal Business Database," Technical Report, Center for Economic Studies 2002.

Karahan, Fatih, Benjamin Pugsley, and Aysegul Sahin, "Demographic Origins of the Startup Deficit," *Working paper*, 2018.

Komarek, Timothy M, "Labor market dynamics and the unconventional natural gas boom: Evidence from the Marcellus region," *Resource and Energy Economics*, 2016, *45*, 1–17.

Lee, Yoonsoo and Toshihiko Mukoyama, "Entry and Exit of Manufacturing Plants over the Business Cycle," *European Economic Review*, July 2015, *77*, 20–27.

Lucas, Robert, "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 1978, *9* (2), 508–523.

Marchand, Joseph, "Local labor market impacts of energy boom-bust-boom in Western Canada," *Journal of Urban Economics*, 2012, *71* (1), 165–174.

McCollum, Meagan and Gregory B. Upton, "Local Labor Market Shocks and Residential Mortgage Payments: Evidence from Shale Oil and Gas Booms," *Resource and Energy Economics*, 2018, *53*, 162–197.

Michaels, Guy, "The Long-term Consequences of Resource-Based Specialisation," *The Economic Journal*, 2010, *121*, 31–57.

Moreira, Sara, "Firm Dynamics, Persistent Effects of Entry Conditions, and Business Cycles," Technical Report 2017.

Oliver, Matthew and Gregory B. Upton, "Are Energy Endowed Countries Responsible for Conditional Convergence?," *USAEE Working Paper*, 2019, *414.*

Poschke, Markus, "Wage Employment, Unemployment and Self-Employment Across Countries," Technical Report, McGill University 2018.

Pugsley, Benjamin W and Aysegul Sahin, "Grown-up business cycles," *US Census Bureau Center for Economic Studies Paper No. CES-WP-15-33*, 2015.

Sachs, Jeffrey D. and Andrew M. Warner, "Natural Resources and Economic Development: The curse of natural resources," *European Economic Review*, 2001, *45*, 827–838.

Sedlacek, Petr and Vincent Sterk, "The Growth Potential of Startups over the Business Cycle," *American Economic Review*, 2017, *107* (10), 3182–3210.

Smith, Brock, "The resource curse exorcised: Evidence from a panel of countries," *Journal of Development Economics*, 2015, *116*, 57–73.

Tsvetkova, Alexandra and Mark Partridge, "The shale revolution and entrepreneurship: An assessment of the relationship between energy sector expansion and small business entrepreneurship in US counties," *Energy*, 2017, *141* (15), 423–434.

Unel, Bulent and Gregory B. Upton Jr., "Effects of the Shale Boom on Entrepreneurship in the U.S.," *USAEE Working Paper*, 2020, *20* (461).

Upton, Gregory B. and Han Yu, "Local Labor Market Shocks and Employment and Earnings Differentials: Evidence from Shale Oil and Gas Booms.," *Working Paper*, 2017.

van der Ploeg, Frederick, "Natural Resources: Curse or Blessing?," *Journal of Economic Literature*, 2011, *49* (2), 366–420.

Venables, Anthony J., "Using Natural Resources for Development: Why Has It Proven So Difficult?," *Journal of Economic Perspectives*, 2016, *30*, 168–184.

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

8 Tables and Figures

Table 1: Pre-Boom Labor Market Summary Statistics

	Employment (counts)	Establishments (counts)	Oil & gas mining share of employment (%)	Oil & gas mining share of establishments (%)
Panel A: Treated and Control Areas				
Shale counties	20,800	1,440	3.5	2.8
Boom towns	8,800	700	6.8	4.9
Non-shale counties	38,200	2,410	0.7	0.9
Non-shale control set	43,800	2,790	0.2	0.3
Panel B: Major Shale Plays				
Anadarko	17,600	1,290	5.5	5.4
Appalachia	30,600	1,950	0.9	1.1
Bakken	3,000	310	3.6	2.8
Eagle Ford	8,100	690	4.1	2.9
Haynesville	17,600	1,170	2.0	1.9
Niobrara	30,500	2,380	2.5	2.3
Permian Basin	6,600	530	9.7	6.2

Note: Average county-level figures by play and NAICS sector for 2000-2006. Employment and establishment counts rounded to nearest 100 and 10, respectively. Oil & gas mining includes NAICS 211 and 213. Boom towns include counties in Anadarko, Bakken, Eagle Ford, and Permian Basin. Non-shale counties include all U.S. counties outside shale areas. Non-shale control set includes all counties except those in shale states and states adjacent to shale counties (see text). Source: County Business Patterns

Table 2: Pre-boom Average County Establishments by Sector

	Min- ing	Util- ities	Con- struction	Man- ufacturing	Ret- ail	Transp- ortation & ware	Prof. Services & biz	Educ- ation & Health	Leis- ure & Hosp.	Other Services
Panel A: Treated and Control Areas										
All Shale Counties	21	6	147	66	234	46	199	160	137	161
Boom Towns	25	4	62	25	115	30	90	72	63	74
All Non-Shale	6	6	244	113	368	66	399	260	227	240
Potential Controls	4	6	287	137	423	77	460	299	262	271
Panel B: Major Shale Plays										
Anadarko	43	6	118	49	200	38	200	146	106	134
Appalachia	14	7	196	102	333	59	245	236	201	242
Bakken	8	3	31	9	55	14	30	25	36	36
Eagle Ford	12	5	63	24	124	63	72	66	64	69
Haynesville	27	11	98	52	212	38	148	129	93	131
Niobrara	25	6	304	91	318	60	456	225	201	207
Permian	27	4	45	20	89	19	60	53	48	59

Note: Average county-level establishment counts by play and NAICS sector for 2000-2006. Sectors are mining; utilities; construction; manufacturing; retail trade; transportation and warehousing; professional and business services; education and health; leisure and hospitality; and other services. Residual sectors omitted. Boom towns include counties in Anadarko, Bakken, Eagle Ford, and Permian Basin. Non-shale counties include all U.S. counties outside shale areas. Non-shale control set includes all counties except those in shale states and states adjacent to shale counties (see text). Source: County Business Patterns

Table 3: Summary Statistics - Baseline Sample

	Treatment Group (T)			Control Group (C)			
	Pre 2007	Post 2007	$\%\Delta T$	Pre 2007	Post 2007	$\%\Delta C$	$\%\Delta T$ - $\%\Delta C$
Total Employment	21,000	21,500	2.38%	25,000	24,000	-4.00%	6.38%
New Firm Share of Emp.	3.3%	2.7%	-18.18%	3.3%	2.3%	-30.30%	12.12%
Young Firm Share of Emp.	10.9%	9.2%	-15.60%	10.6%	8.2%	-22.64%	7.04%
Greenfield Share of Emp.	2.7%	2.8%	3.57%	2.6%	2.3%	-11.54%	15.11%
Oil and Gas Share of Emp.	4.0%	6.2%	55.00%	0.2%	0.2%	0.00%	55.00%
Exit Rate of Firms	5.6%	4.8%	-16.67%	5.5%	4.8%	-12.73	-3.94%
Job Destruction Rate	14.6%	13.9%	-4.79%	14.2%	13.6%	-9.04%	4.25%

Averages of annual data for treatment and control groups. Pre-2007 period is 2000-2006. Post-2007 period is 2007-2014. Total employment in counts. "New Firm Share of Employment" is employment associated with new firms as a share of total county employment. "Young Firm Share of Employment" is employment associated with young firms (age 1 to 4) as a share of total employment. "Greenfield Share of Employment" is the share of employment associated with new establishments of existing firms. "Oil and Gas Share of Employment" is the share of total employment in the oil and gas industry (NAICS 211, 213, 324, and 325). Exit Rate of Firms is employment weighted. Job Destruction rate is the gross measure of all jobs destroyed divided by total employment, using a 2 year average. Source: Longitudinal Business Database

Table 4: Impact of Shale on (Log) Employment - All Industries Except Oil and Gas Mining

	(1) All	(2) Boom Towns	(3) Appalachia	(4) Eagle Ford	(5) Haynesville	(6) Niobrara	(7) Permian
$\hat{\delta}$	0.069***	0.131***	0.014	0.198***	-0.008	0.0025	0.091***
	(0.012)	(0.023)	(0.012)	(0.054)	(-0.038)	(0.035)	(0.035)
N	9,420	3,780	3,780	690	750	1,110	1,620

Oil and gas mining sector (NAICS 211, 213) omitted. Dependent variable natural log of total employment excluding oil and gas mining employment in all regressions. County clustered standard errors shown. Treatment time period post 2007. Treated areas include all counties with shale oil and/or gas production as defined by EIA Drilling Productivity Reports. Control counties chosen using propensity score match from national sample in non-shale states. "Boom Town" is a combination of Permian, Anadarko, Eagle Ford, and Bakken plays. Bakken and Anadarko results are included in the "All" group but are not able to be reported individually due to data confidentially constraints. Parameters estimated with OLS. Source: Longitudinal Business Database

Table 5: Impact of Shale on Annual Employment Growth Components - All Industries Except Oil and Gas Mining

	(1) New Firms	(2) Young Firms	(3) New & Young Firms (1)+(2)	(4) Mature Firms	(5) Greenfield Estabs	(6) Incumbent Estabs	(7) Total (1)+(2)+(4) (1)+(5)+(6)
Panel A. All Areas							
$\hat{\delta}$	0.321***	0.359***	0.680***	1.149***	0.340**	1.167***	1.829***
	(0.122)	(0.123)	(0.144)	(0.274)	(0.146)	(0.301)	(0.335)
N	9,420	9,420	9,420	9,420	9,420	9,420	9,420
Share of Total	17.6%	19.6%	37.2%	62.8%	18.6%	63.8%	100%
Panel B. Boom Towns							
$\hat{\delta}$	1.182***	0.672**	1.852***	2.010***	0.709**	1.972***	3.864***
	(0.285)	(0.286)	(0.308)	(0.496)	(0.275)	(0.607)	(0.669)
N	3,780	3,780	3,780	3,780	3,780	3,780	3,780
Share of Total	30.6%	17.4%	47.9%	52.0%	18.3%	51%	100%

Oil and gas mining sector (NAICS 211, 213) omitted. Dependent variable growth component in all regressions. County clustered standard errors shown. Treatment time period post 2007. Treated areas include all counties with shale oil and/or gas production as defined by EIA Drilling Productivity Reports. Control counties chosen using propensity score match from national sample in non-shale states. "Boom Town" is a combination of Permian, Anadarko, Eagle Ford, and Bakken plays. Parameters estimated with OLS. New firm age (in years) =0, young =1-4, old = 5+. Columns 1+2=3, columns 1+2+4=7, and columns 1+5+6=7. Source: Longitudinal Business Database

Table 6: Impact of Shale on Annual Employment Growth Components - By Industry

	(1) New Firms	(2) Young Firms	(3) New & Young Firms (1) + (2)	(4) Mature Firms	(5) Greenfield Estabs	(6) Incumbent Estabs	(7) Total (1)+(2)+(4) (1)+(5)+(6)
Panel A. All Areas							
Oil & Gas Mining (NAICS 211, 213)							
$\hat{\delta}$	7.84***	0.94	8.78***	10.37***	2.71	8.60**	19.15***
	(2.81)	(2.06)	(3.19)	(3.15)	(2.07)	(3.44)	(4.41)
N	5,133	5,133	5,133	5,133	5,133	5,133	5,133
Share of total	40.9%	4.9%	45.8%	54.2%	14.2%	44.9%	100%
Construction, Warehousing, & Transportation (NAICS 23, 48, 49)							
$\hat{\delta}$	0.97***	1.07***	2.04***	2.50***	0.22	3.36***	4.55***
	(0.37)	(0.32)	(0.41)	(0.64)	(0.29)	(0.69)	(0.79)
N	9,382	9,382	9,382	9,382	9,382	9,382	9,382
Share of total	21.3%	23.5%	44.8%	54.9%	4.8%	73.8%	100%
All Other Industries							
$\hat{\delta}$	0.18	0.28**	0.46***	0.91***	0.25*	0.94***	1.37***
	(0.13)	(0.12)	(0.14	(0.26)	(0.14)	(0.29)	(0.30)
N	9,420	9,420	9,420	9,420	9,420	9,420	9,420
Share of total	13.1%	20.4%	33.6%	66.4%	18.2%	68.6%	100%
Panel B. Boom Towns							
Oil & Gas Mining (NAICS 211, 213)							
$\hat{\delta}$	7.61***	2.45	10.06**	6.45*	3.99	4.90	16.51***
	(3.77)	(2.72)	(4.78)	(3.76)	(3.13)	(5.00)	(6.31)
N	2,125	2,125	2,125	2,125	2,125	2,125	2,125
Share of total	46.1%	14.8%	60.9%	39.1%	24.2%	29.7%	100%
Construction & Transportation (NAICS 23, 48, 49)							
$\hat{\delta}$	2.76***	2.25***	5.01***	5.28***	1.27**	6.26***	10.28***
	(0.85)	(0.69)	(0.93)	(1.25)	(0.59)	(1.33)	(1.61)
N	3,752	3,752	3,752	3,752	3,752	3,752	3,752
Share of total	26.8%	21.9%	48.7%	51.4%	12.4%	60.9%	100%
All Other Industries							
$\hat{\delta}$	0.86***	0.44	1.30***	1.22***	0.42	1.24**	2.52***
	(0.29)	(0.28)	(0.30)	(0.47)	(0.26)	(0.59)	(0.61)
N	3,780	3,780	3,780	3,780	3,780	3,780	3,780
Share of total	34.1%	17.5%	51.6%	48.4%	16.7%	49.2%	100%

County clustered standard errors shown. Treatment time period post 2007. Treated areas include all counties with shale oil and/or gas production as defined by EIA Drilling Productivity Reports. Control counties chosen using propensity score match from national sample in non-shale states. "Boom Town" is a combination of Permian, Anadarko, Eagle Ford, and Bakken plays. Parameters estimated with OLS. New firm age (in years) =0, young =1-4, old = 5+. Source: Longitudinal Business Database

Table 7: Relationship between oil & gas mining employment and non-oil & gas employment growth

	(1) New Firms	(2) Young Firms	(3) New & Young Firms (1)+(2)	(4) Mature Firms	(5) Greenfield Estabs	(6) Incumbent Estabs	(7) Total (1)+(2)+(4) (1)+(5)+(6)
Panel A. All Areas							
$\ln emp^{211,213}$	0.142***	0.006	0.148	0.409***	0.163***	0.251	0.557***
	(0.050)	(0.118)	(0.136)	(0.111)	(0.053)	(0.181)	(0.206)
N	9,420	9,420	9,420	9,420	9,420	9,420	9,420
Share of Total	25.5%	1.1%	26.6%	73.4%	29.3%	45.1%	100%
Panel B. Boom Towns							
$\ln emp^{211,213}$	0.513***	-0.142	0.371	0.812***	0.475***	0.195	1.183*
	(0.156)	(0.384)	(0.443)	(0.284)	(0.128)	(0.552)	(0.625)
N	3,780	3,780	3,780	3,780	3,780	3,780	3,780
Share of Total	43.4%	12.0%	31.4%	68.6%	40.2%	16.5%	100%

Dependent variable growth component in all regressions. County clustered standard errors shown. Treatment time period post 2007. Treated areas include all counties with shale oil and/or gas production as defined by EIA Drilling Productivity Reports. Control counties chosen using propensity score match from national sample in non-shale states. "Boom Town" is a combination of Permian, Appalachian, Eagle Ford, and Bakken plays. Parameters estimated with OLS. New firm age (in years) =0, young =1-4, old = 5+. Columns 1+2=3, columns 1+2+4=7, and columns 1+5+6=7. Source: Longitudinal Business Database

Table 8: Impact of Shale on Employment Ratio - All Industries Except Oil and Gas Mining

	(1) Pre-2007 establishments *(Incumbent Estabs)*	(2) New establishments 2007 to pre-2007 firms *(Greenfield Estabs)*	(3) New establishments of firms born 2007 and later *(New Firms)*	(4) Total *(All)*
All Areas				
δ_{2000}	0.0138	0	0	0.0138
	(0.0129)			(0.0129)
δ_{2001}	0.0143	0	0	0.0143
	(0.0119)			(0.0119)
δ_{2002}	0.0190*	0	0	0.0190*
	(0.0101)			(0.0101)
δ_{2003}	0.0086	0	0	0.0086
	(0.0103)			(0.0103)
δ_{2004}	0.0095	0	0	0.0095
	(0.0083)			(0.0083)
δ_{2005}	0.0007	0	0	0.0007
	(0.0062)			(0.0062)
δ_{2007}	0.039	0.0037	-0.0027	0.005
	(0.0062)	(0.0082)	(0.0023)	(0.0077)
δ_{2008}	0.0129	0.0082**	-0.0142	0.0174**
	(0.0009)	(0.0036)	(0.0142)	(0.0174)
δ_{2009}	0.0456***	0.0096**	-0.0038	0.0099***
	(0.0084)	(0.0053)	(0.0048)	(0.0106)
δ_{2010}	0.0467***	0.0135**	0.0088*	0.0691***
	(0.0084)	(0.0053)	(0.0048)	(0.0106)
δ_{2011}	0.0558***	0.0211***	0.0169***	0.0938***
	(0.0095)	(0.0068)	(0.0053)	(0.0134)
δ_{2012}	0.0604***	0.0234***	0.0430***	0.1268***
	(0.0105)	(0.0076)	(0.0088)	(0.0187)
δ_{2013}	0.0667***	0.0348***	0.0553***	0.1568***
	(0.0112)	(0.0093)	(0.0115)	(0.0235)***
δ_{2014}	0.0547***	0.0417***	0.0760***	0.1724***
	(0.0114)	(0.0100)	(0.0144)	(0.0273)
N	9,420	9,420	9,420	9,420

Oil and gas sector (NAICS 211, 213) omitted. County clustered standard errors shown. Base year 2006 and therefore not shown in table. Treated areas include all counties with shale oil and/or gas production as defined by EIA Drilling Productivity Reports. Control counties chosen using propensity score match from national sample in non-shale states. Parameters estimated with OLS. New firm age (in years) =0, young =1-4, old = 5+. Employment ratio is defined as the ratio of a given group's employment in a given year to the total county employment for that group in the base year of 2006. Pre-treatment period is 2000-2006 and post-treatment period is 2007-2014. Source: Longitudinal Business Database

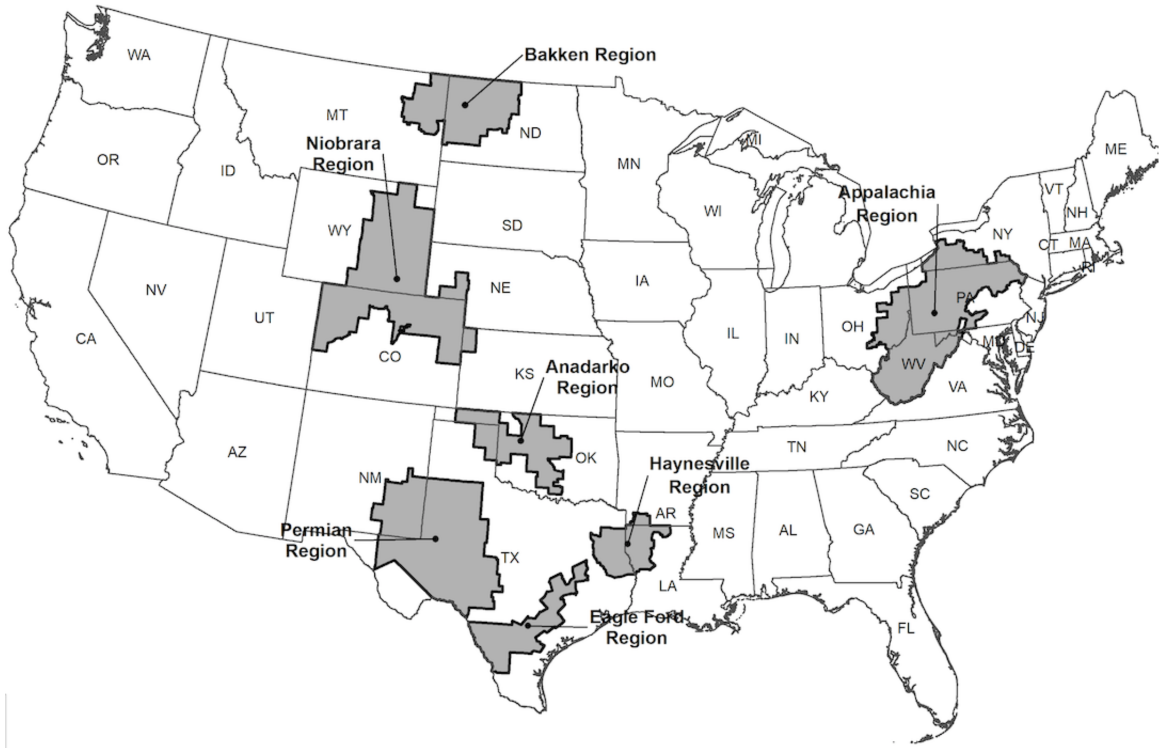


Figure 1: U.S. Shale Plays

Source: U.S. Energy Information Administration. Drilling Productivity Reports

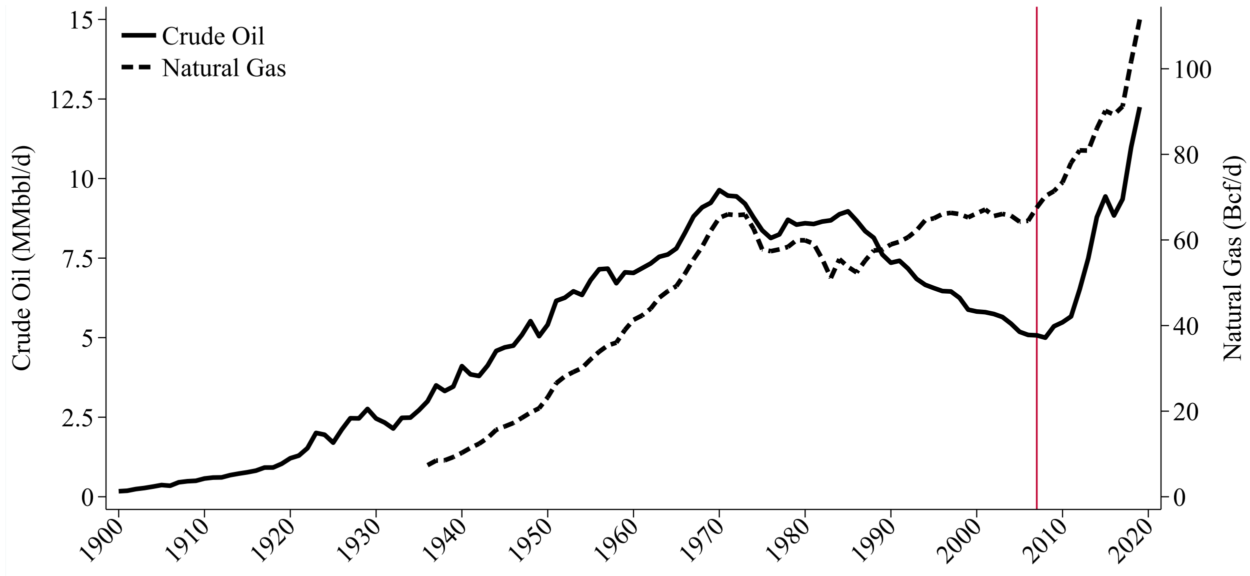
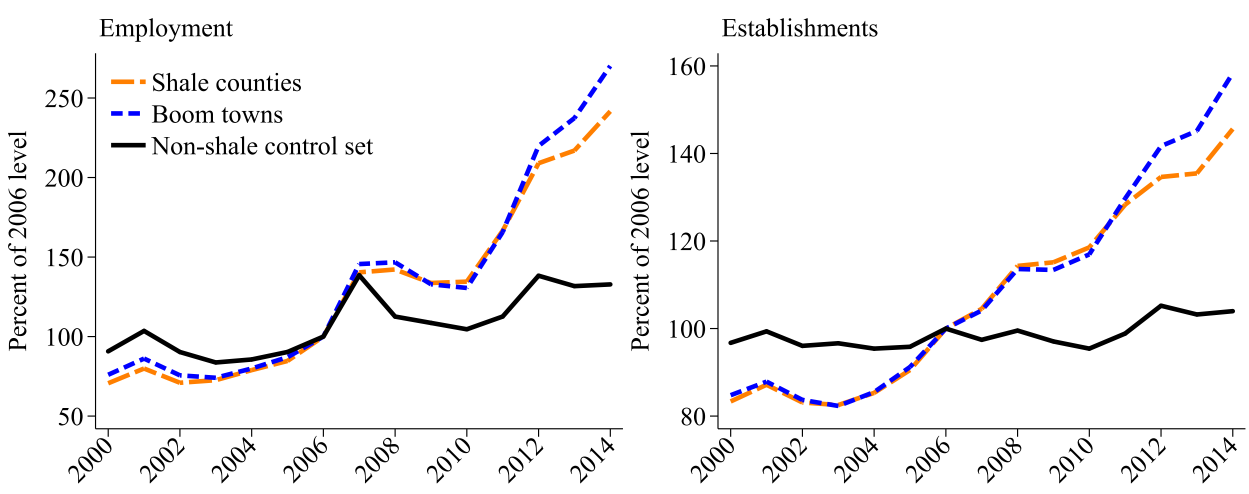


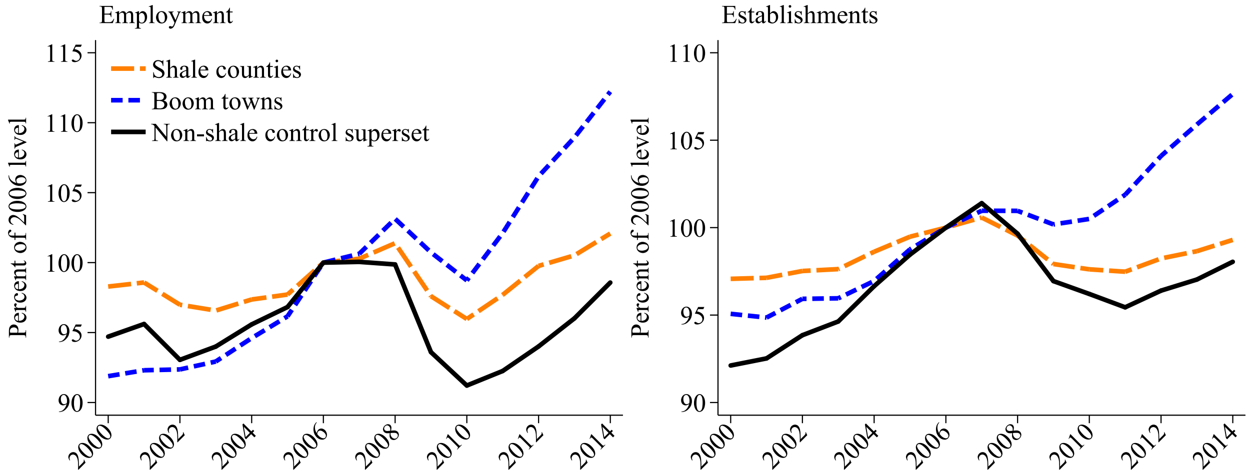
Figure 2: Historical U.S. Crude Production

Source: U.S. Energy Information Administration. U.S. Crude Oil and Natural Gas Production



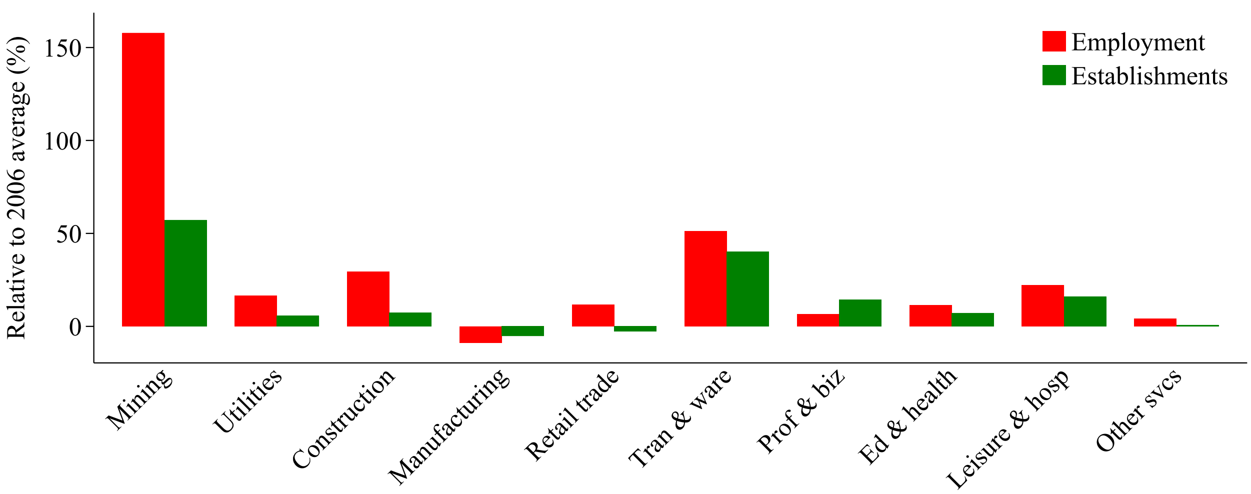
Note: Average county-level employment and establishment counts scaled by 2006 averages.
Boom towns are Anadarko, Bakken, Eagle Ford, and Permian Basin.
Source: County Business Patterns. NAICS 211 and 213. States that include or border shale counties omitted.

Figure 3: Oil and gas mining activity, shale vs. non-shale counties



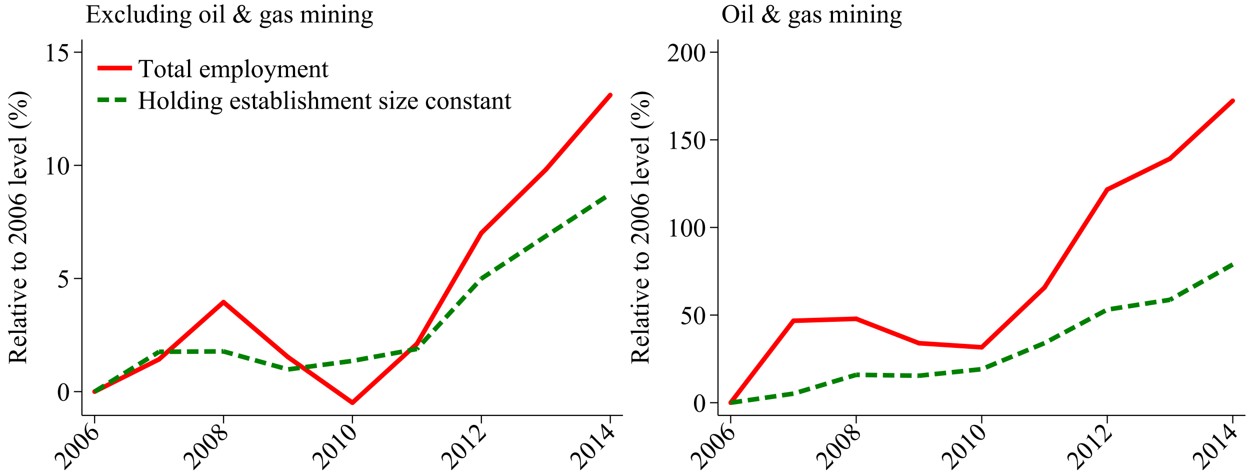
Note: Average county-level employment and establishment counts scaled by 2006 averages.
Boom towns are Anadarko, Bakken, Eagle Ford, and Permian Basin.
Source: County Business Patterns. Excludes NAICS 211 and 213. States that include or border shale counties omitted.

Figure 4: Activity excluding oil and gas mining, shale vs. non-shale counties



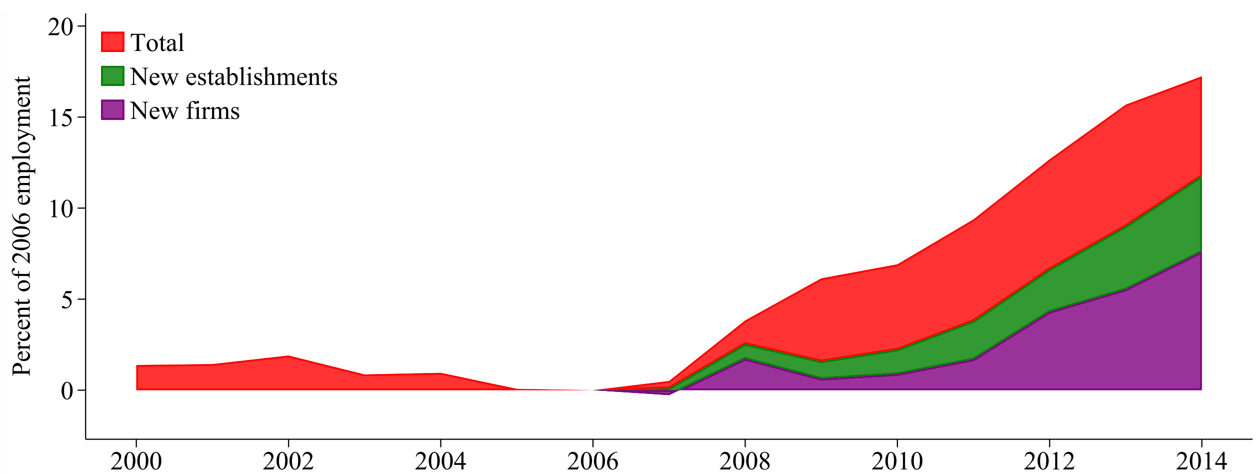
Note: Average county-level employment and establishment counts relative to average 2006 level. Boom towns are Anadarko, Bakken, Eagle Ford, and Permian Basin. Residual industries omitted.
Source: County Business Patterns. NAICS sectors.

Figure 5: Boom town activity by sector, 2006-2014



Note: Total employment relative to year-2006 level. Covariance term shared out proportionally.
Boom towns are Anadarko, Bakken, Eagle Ford, and Permian Basin.
Source: County Business Patterns. Oil & gas mining consists of NAICS 211 and 213.

Figure 6: Margins of employment growth in boom towns



Employment scaled by 2006 county employment. Regression compares treatment and control counties with year effects. New establishments are establishments born after 2006. New firms are firms born after 2006.
Author calculations from LBD. Oil and gas mining (NAICS 211, 213) excluded.
Total employment is statistically significant from 2008 on.

Figure 7: Employment treatment effects by year: All regions

A Appendix A: Model

In this section we briefly sketch model intuition behind our theoretical framework. Clementi and Palazzo (2016) construct a fully featured model of firm dynamics for studying the cyclical properties of business entry. Here we describe a simplified version of that model to explore key results. The differences between our model here and that of Clementi and Palazzo (2016) are (a) we omit capital from the model and (b) we study a simple transition path exercise rather than implementing full stochastic aggregate risk and business cycle exercises. We also initially differ by shutting down ex ante heterogeneity of entrants, but we expand our investigation to include ex ante heterogeneity further below. While we do calibrate the model, we take much of our calibration from existing literature and focus primarily on the qualitative results.

Firms face idiosyncratic productivity draws z and an aggregate productivity state A. Idiosyncratic productivity evolves according to $\ln z' = \rho_z \ln z + \sigma_z \varepsilon_z'$ where $\varepsilon_z \sim N(0,1)$; this yields a conditional distribution of z' given by $H(z'|z)$. Firms produce using technology Azn^{α}, where α governs revenue curvature (which we interpret here as decreasing returns to scale); firms discount profits with factor β and face a spot market for labor with wage w and labor supply curve $L_s(w) = w^{\gamma}$ (with $\gamma > 0$). Continuing firms must pay a fixed operating cost c_f; the operating cost is not persistent, and $c_f \sim LN(\mu_c, \sigma_c)$.

Under these assumptions, entry is determined by the free entry condition:

$$c_e = \mathbb{E}_{z'} V(z'; A, w), \tag{11}$$

where c_e is the entry cost and $V(z'; A, w)$ is the value function of an operating firm. The mass of entrants is determined in equilibrium such that the average firm value is pinned down to the entry cost. Upon entry, new entrants receive productivity draws consistent with the unconditional productivity distribution of incumbent firms.

The timing of the model is as follows. At the beginning of the period, incumbents observe their productivity z then hire labor and produce. Incumbents, following production, draw their operating cost c_f then choose whether to continue or exit; at the same time, the mass of entrants is determined, and entrants pay the entry cost c_e. Then the next period begins.

The incumbents' problem is as follows. First, the incumbent faces a static profit maximization problem yielding the following first-order condition for labor demand:

$$n(z; A, w) = \left(\frac{w}{\alpha A z} \right)^{\frac{1}{\alpha - 1}}. \tag{12}$$

This yields a profit function $\pi(z; A, w)$.

Here we briefly digress to illustrate the central intuition of incumbent behavior in the model. Suppose A increases by x percent. Our interest is in the growth of firm-level employment in a comparative statics view:

$$g = \frac{n(z; (1 + x)A, w) - n(z; A, w)}{n(z; A, w)} \tag{13}$$

where we hold the wage constant (i.e., partial equilibrium). In this simple environment it is straightforward to show that

$$g = (1 + x)^{\frac{1}{1-\alpha}} - 1, \tag{14}$$

that is, the firm's employment growth response is a function only of x and α. Importantly for our study, the absolute value of the growth rate is increasing in α or, equivalently, decreasing in the curvature of the revenue function. Revenue function curvature dampens the response of incumbents to shocks.[32] Equivalently, revenue function curvature compresses the distribution of labor demand across firms of different productivity realizations. This implies that aggregate shocks affect the number of firms, not just the size of preexisting firms.

Returning to the model environment, the value of an incumbent at the beginning of a period is given by:

$$V(z; A, w) = \pi(z; A, w) + \beta \mathbb{E}_{c_f} \left[\max\{0, \mathbb{E}_{z'|z} V(z'; A, w) - c_f\} \right] \tag{15}$$

This optimization problem yields an exit rule such that firms choose to exit when the expected value of the firm is negative (where exit provides a payoff of zero, as shown in the internal maximization operator); this results in a threshold rule such that incumbents exit when $z \leq z^*(A, w)$.

The recursive competitive equilibrium is defined as follows. $V(z; A, w), n(z; A, w)$, and the associated exit rule arising from the threshold z^* solve the incumbents' problem, and the mass

[32] Decker et al. (Forthcoming) show that this result holds even in a more fully specified and calibrated model with labor adjustment costs.

of entrants M is such that the free entry condition (11) holds with equality; the distribution of new entrants is given by $E(z') = M^* H(z')$. The labor market clears; that is, $w^\gamma = \int n(z; A, w) d\Gamma(z)$, where $\Gamma(z)$ is the measure of producing firms (distributed over z). Finally, the measure of firms evolves according to $\Gamma'(z') = \int \int_{c_f} \int_{z^*}^{\infty} d\Gamma(z) dG(c_f) dH(z'|z) + E(z')$. The latter condition simply illustrates that the new distribution of firms reflects the distribution of incumbents that chose not to exit, appropriately transitioned to updated productivity draws, plus the mass and distribution of new entrants .

We calibrate the model as reported on Table A5 in the column labeled "Model 1"; this calibration mostly follows Clementi and Palazzo (2016) except that we choose μ_c (the operating cost distribution mean) to target an entry rate of 9% (that is, entrants account for 9% of firms), consistent with Business Dynamics Statistics data from the early 2000s.

We solve the steady state of the model by starting with guesses for the entry mass M and the wage w, solving value functions and policy functions (via value function iteration), iterating to a stationary distribution where $\Gamma' = \Gamma$, checking labor market clearing, revising the wage until the market clears, then revising the entry mass M until the free entry condition holds. We consider two steady states; in the baseline steady state we set $A = 1$, and in the expansion steady state we set $A = 1.2$ (these choices are arbitrary, designed only to illustrate qualitative dynamics). We then study a transition from the baseline to the expansion state. In period 0, the economy is in the baseline steady state with no expectation for change. In period 1, firms learn that A will transition from 1 to 1.2 effective the beginning of period 2, after which the economy will converge to the steady state associated with $A = 1.2$ and no expectation of change. This exercise is illustrated on the top left panel of Figure A5. The positive aggregate shock (solid blue line) causes a permanent increase in the number of firms (dotted black line); this rise in the firm count is facilitated by a surge in entry (dashed red line), including the employment-weighted entry rate (dot-dashed green line). This result (surging entry and employment-weighted entry) is robust to a wide range of parameterizations.

We next generalize the model slightly to allow ex ante heterogeneity among entrants. At any time there exists a mass M_p of potential entrants. Each potential entrant receives a *signal* about their productivity given by $q \sim Pareto(\min(z), \xi)$. The signal q relates to productivity on entry with the conditional distribution $H(z'|q)$; that is, productivity on entry follows $\ln z' = \rho_z \ln q + \sigma_z \varepsilon'_z$. While it is not strictly necessary that the distribution of potential entrants' signals differ from the distribution of incumbents' productivity, doing so makes it possible to match the number and size of entrant firms to the data. While incumbent firms are producing, potential entrants observe their signal q and choose whether to enter for production in the next period. The potential entrants' problem is solved simply by choosing to enter when the free entry condition holds:

$$\beta \mathbb{E}_{z'|q} V(z'; A, w) \geq c_e \tag{16}$$

As is common in models of this class, this free entry condition yields an entry rule such that potential entrants choose to enter if and only if $q \geq q^*(A, w)$, where $q^*(A, w)$ is a threshold value dependent on the aggregate state. This threshold rule differs in important ways from the simpler free entry condition given by (11); in particular, the threshold rule does not hold with equality and, therefore, has less stark implications for the value of existing firms. Additionally, the productivity distribution of new entrants differs from that of continuing incumbents due to the signal distribution; this is necessary for matching the firm size distribution (as noted by Clementi and Palazzo (2016)), but it creates different dynamics for the employment share of entrants. On Table A5, the column "Model 2" reports calibration details for this model generalization.

We conduct the same transition path exercise as above, reported on the top right panel of Figure A5. The solid blue line reports the path of aggregate productivity. The dotted black line shows that, as in the previous experiment, the improvement in aggregate conditions causes a rise in the number of firms as existing firms do not grow enough to accommodate the shock. The red dashed line shows that, in this calibration, the rise in the number of firms is facilitated in part by a surge in entry. However, unlike the previous experiment, the green dot-dashed line shows that the employment share of entrants does not rise. This is the result of ex ante heterogeneity and quality signals; in this setup, the rise in entry is driven by a decline in the threshold for the productivity signal above which entry is profitable. This induces a selection mechanism in which the positive aggregate shock allows lower-quality entrepreneurs to enter; upon entering, their employment is lower than the minimum productivity of entrants during the initial stationary state.

The exercises from our more general model still support the notion that aggregate shocks are accommodated, at least in part, by a rise in entry. However, even this result is heavily influenced by calibration. For example, the bottom left panel of Figure A5 reports the same experiment except that the revenue curvature parameter α is set at 0.7 (rather than 0.8); in this experiment, even the unweighted entry rate responds negatively to the shock (note that the number of firms still rises, facilitated by a lower exit rate). The bottom right panel of Figure A5 shows that the entry rate effect can be reduced by lowering the labor supply elasticity to $\gamma = 1$ (from $\gamma = 2$). Future research might further explore calibration considerations in relation to our empirical results.

B Appendix B: Data

B.1 County Business Patterns

County Business Patterns (CBP) is based on the Census Bureau's Business Register and covers almost all private employer establishments in the U.S. Non-employers—those businesses without employees for Social Security Administration purposes—are excluded; however, the employer universe covers potentially all legal forms of organization, including sole proprietors (among whom are many employers). See DeSalvo et al. (2016) for details on the Business Register data underlying CBP.

CBP covers the universe of private business establishments, excluding only the following NAICS industries: 111 and 112 (crop and animal production), 482 (rail transportation), 491 (Postal Service), 525110, 525120, 525190 (pension, health, welfare, and vacation funds), 525920 (trusts, estates, and agency accounts), 814 (private households), and 92 (public administration). Government-owned businesses in the following NAICS industries are included: 4248 (wholesale liquor establishments), 44531 (retail liquor stores), 511130 (book publishers), 522120 (federally-chartered savings institutions), 522130 (federally-chartered credit unions), and 622 (hospitals).[33]

While establishment counts are published for all industry-by-county cells in CBP data, employment counts are suppressed in some cells. In these cases, a size range is reported instead of a precise employment count. We use these size range reports to impute employment to suppressed cells; we first impute employment for any suppressed county-level observations, then we impute employment for suppressed county-by-sector observations.

We impute suppressed county-level employment as follows. Within a given year, we first categorize all non-suppressed counties into size bins that correspond to the size bins reported for suppressed counties. We then obtain average actual county employment by size bin (among non-suppressed counties) and populate the employment variable for suppressed counties with the average employment of non-suppressed counties that have reported employment within the corresponding size bin.[34] That is, we estimate that suppressed counties

[33] See https://www.census.gov/programs-surveys/cbp/technical-documentation/methodology.html.

[34] If there are no counties with employment in the indicated size bin, we assign the midpoint of the size bin for all bins except the top bin, for which we assign the lower bound of the bin.

have employment equal to the average employment of non-suppressed counties in the same size class. Next, we sum up total U.S. employment for the year by adding total employment among non-suppressed counties with total estimated employment among suppressed counties. We compare this estimated total to actual reported total U.S. employment for the year (which is available in separate national-level CBP files). Observing the discrepancy between true national employment and our national estimate based on our initial imputation for suppressed counties, we then modify our estimated employment for suppressed counties by sharing out the discrepancy proportionally (based on each county's estimated share of total suppressed employment). The result is our final estimate of employment in each suppressed county. Our imputation method therefore assumes that true county employment for suppressed counties is distributed among employment size bins in a manner similar to the employment distribution of non-suppressed counties, but adjusted to ensure that county employment adds up to true national employment. Observations in which county-level employment is suppressed comprise no more than 0.2% of employment, depending on the year; prior to 2011 imputed observations never account for more than 0.1% of employment.

With populated county-level employment values in hand (whether true or imputed), we next impute employment for suppressed county-by-sector cells (where sectors are defined by two-digit NAICS codes). We proceed in a fashion similar to our county employment imputation method, but our imputation now uses sector-specific averages. Specifically: within a given year *and sector,* we obtain average employment by employment size bin among non-suppressed county-by-sector cells then apply that average to each suppressed cell according to its reported employment bin. After doing this for each sector, we add up sector employment (that is, the sum of total employment among non-suppressed cells and estimated employment among suppressed cells) *by county* and compare the estimated county employment to true reported county employment (or, in the case of counties in which county employment was suppressed, we compare to estimated total county employment as constructed above). Observing the discrepancy between total reported county employment and total county employment based on estimated sector cells, we then adjust our estimates for suppressed cells by sharing out the county-level discrepancy in manner proportional to the initial estimates. This method ensures that sector-level employment within counties adds up to total county employment appropriately. Observations in which county-by-sector employment is suppressed comprise between 1.1 and 3.2% of employment (after imputation), depending on the year (the share generally increases over time).

Finally, since some of our exercises involve narrower industry groups requiring 3-digit

NAICS aggregations, we also impute data for certain county-by-3-digit-NAICS cells (211, 213, 324, and 325). For these we proceed in similar fashion to our approach described above: by year, we estimate cell employment based on the nationwide average of cell employment for each specific 3-digit industry. We then adjust these estimates by aggregating to the county-by-sector level. That is, we adjust estimates for cells of NAICS 211 and 213 by aggregating to county NAICS 21 (mining) employment, and we adjust estimates for cells of NAICS 324 and 325 by aggregating to county NAICS 31-33 (manufacturing) employment. Note that this method requires us to determine suppression and impute for all 3-digit naics industries within these two sectors (21 and 31-33). Observations in which county-by-industry employment for the 3-digit industries in NAICS 21 and 22 is suppressed comprise between 2.6 and 3.5% of employment (after imputation), depending on the year (the share generally increases over time).

B.2 Longitudinal Business Database

The Longitudinal Business Database (LBD), like CBP, is based on the Census Bureau's Business Register. The two datasets also share the same industry scope. Jarmin and Miranda (2002) describe the construction of the LBD. Critically for our purposes, the LBD consists of establishment-level data with longitudinal establishment identifiers. The data also include a firm identifier linking establishments under common ownership or operational control; importantly, this firm identifier is superior to simple tax identifiers (i.e., EINs), since some firms have multiple EINS. Industry codes correspond to establishments. For our purposes it is not necessary to assign an industry code to firms; all industry categories are based on establishment industry (and, as such, industry characteristics of "new firms" actually reflect the industry characteristics of establishments of new firms in a given county).

Consistent with much of the literature (e.g., Haltiwanger et al. (2013)), we define an *establishment* birth as the first year in which an establishment has positive employment, and we determine *firm* age as follows: when a firm identifier first appears in the data, it is assigned the age of its oldest establishment; thereafter, the firm ages naturally each year.[35]

[35]The establishment-level longitudinal linkages in the LBD are generally considered to be of high integrity. Unfortunately, the longitudinal linkages of the LBD's firm identifiers are less reliable and are therefore a source of measurement error. Nevertheless, we follow much recent literature in proceeding with firm age concepts that rely on the LBD firm identifier; these concepts are made more robust by the popular method, which we adopt, of assigning firm age based on establishment age at the firm's first appearance.

C Appendix C: Additional Results and Robustness Checks

C.1 Effects on Total Employment by Sector

Section 6.1 reports diff-in-diff results for (the log of) total employment. Here we examine the impact of the shale boom on employment across industries for both all areas and our boom town group. These results are presented in Table A2. We first study the oil and gas sector inclusive of both oil and gas mining (NAICS 211, 213) and related manufacturing (324, 325). This broad oil and gas sector saw average employment increased by almost 50% as a result of the shale boom; however, we find that this effect was driven entirely by the narrower oil and gas mining sector, which gained 70%, while the related manufacturing industries gains were not statistically significant. This latter finding reflects the fact that while significant downstream investments occurred in response to the shale boom, much of this investment was in areas with historical presence of these industries, not necessarily in new areas where extraction is now occurring,[36] therefore spurring significant investment in transportation infrastructure (Agerton and Upton, 2019).

Also presented in Table A2, we find that employment outside of the oil and gas mining sector was also significantly affected, with impacts differing significantly across industries. For instance, construction, transportation and warehousing increased by 21.9%, while retail trade and leisure and hospitality experienced 3.6% and 7.3% increases, respectively with some other sectors such as utilities, professional business services and other services not statistically significantly impacted.

Additionally, we present results for each industry for only the "Boom Towns" sample. We find that similar industries are impacted for this sub-sample, but the magnitude of these effects is greater. Additionally, in contrast to the full results, we find a marginally statistically significant result (16%) for the oil and gas manufacturing sector; however, the magnitude of this effect is relatively small compared to the results for the oil and gas mining sector.

C.2 Alternative Control Groups

Our main results—and the causal interpretation thereof—depend on our propensity-matched control group. We first test the sensitivity of our results to alternative control groups by randomly choosing 20 control groups (rather than relying on our propensity score matching

[36] Dismukes et al. (2019) estimates that over $110 billion in refining and chemical announcement occurred in Texas and Louisiana during the shale boom, but is mostly located near the Gulf Coast, not in the regions where the shale production actually occurred.

algorithm). The counties in these groups are drawn (with replacement) from the U.S. broadly, with the exception of counties close to our treatment counties (as noted above). We estimate our employment growth (by firm age) regressions with each control group; Table A6 shows the minimum, median, and maximum coefficients obtained from these 20 random control groups along with the propensity score match control group estimated treatment effects (i.e., repeated from Table 5).

The random control group exercises are generally supportive of our main results while pointing to the importance of our propensity score approach for generating causal inference. Column 7 of Table A6 reports coefficients for overall employment growth.

Broadly speaking, though, the random control group exercises support our main results and do not raise any concerns about our research design. The shale boom is plausibly exogenous to the patterns of business entry we study (particularly in industries outside oil and gas mining).

C.3 Placebo Tests

We also perform two placebo tests. We randomly assign observations to the control and treatment groups in two ways. First, we estimate our model only using the treated observations (i.e., counties in shale plays) but randomly assigning the observations to be "treatment" or "control". Second, we repeat this exercise using only the control observations (i.e., counties included in our propensity matched control group). Results of placebo tests for employment by shale play are presented in Table A7. None of the 14 coefficients in this table is statistically significant. Broadly speaking, our placebo tests are supportive of our identification strategy.

D Appendix D: Supplemental Tables and Figures

Table A1: Average county establishment counts by sector, 2000-2006

	Min	Uti	Con	Man	Ret	Tran & ware	Prof & biz	Educ & heal	Leis & hosp	Oth svcs
Shale counties	320	160	1,180	2,790	3,060	700	2,630	3,720	2,250	1,050
Boom towns	330	80	480	800	1,370	360	1,030	1,520	1,040	490
Non-shale counties	130	210	2,200	4,780	4,970	1,320	6,500	5,960	4,100	1,790
Non-shale control set	80	230	2,560	5,840	5,760	1,530	7,470	6,430	4,830	1,980
Anadarko	440	130	900	1,870	2,450	600	2,610	2,880	1,940	990
Appalachia	310	250	1,450	4,790	4,460	1,060	3,570	6,060	3,160	1,510
Bakken	150	70	130	140	560	70	240	670	380	160
Eagle Ford	190	80	480	760	1,500	640	650	1,440	1,040	390
Haynesville	310	160	920	2,870	2,570	460	1,920	3,370	2,020	980
Niobrara	360	140	2,800	2,740	4,420	840	5,440	3,540	3,480	1,380
Permian Basin	400	50	390	500	1,050	210	640	1,140	810	400

Average county-level employment by play and NAICS sector for 2000-2006. Sectors are mining; utilities; construction; manufacturing; retail trade; transportation and warehousing; professional and business services; education and health; leisure and hospitality; and other services. Residual sectors omitted. Boom towns include counties in Anadarko, Bakken, Eagle Ford, and Permian Basin. Non-shale counties include all U.S. counties outside shale areas. Non-shale control set includes all counties except those in shale states and states adjacent to shale counties (see text). Source: County Business Patterns

Table A2: Impact of Shale on Employment by Industry- All Plays

	(1)	(2)	(3)	(4)	(5)	(6)
	All	Upstream	Oil & Gas-			Const., Trans.
	Oil & Gas	Oil & Gas	Manufacturing	Mining	Utilities	& Warehousing
NAICS	211, 214, 324, 325	211, 213	324, 325	21	22	23,48,49
Panel A. All Areas						
$\hat{\delta}$	0.476***	0.701***	0.039	0.621***	0.015	0.219***
	(0.067)	(0.057)	(0.066)	(0.060)	(0.038)	(0.028)
N	9,420	9,420	9,420	9,420	9,420	9,420
Panel B. Boom Towns						
$\hat{\delta}$	0.593***	0.602***	0.160*	0.726***	0.085	0.414***
	(0.093)	(0.076)	(0.095)	(0.089)	(0.059)	(0.055)
N	3,780	3,780	3,780	3,780	3,780	3,780

	(7)	(8)	(9)	(10)	(11)	(12)
		Retail	Prof. Business	Education &	Leisure &,	Other
	Manufacturing	Trade	Services	Health Services	Hospitality	Services
NAICS	31, 32, 33	44,45	54, 55, 56	61, 62	71, 72	81
Panel A continued. All Areas						
$\hat{\delta}$	0.107***	0.036***	0.012	-0.022	0.073***	0.023
	(0.035)	(0.013)	(0.035)	(0.021)	(0.020)	(0.017)
N	9,420	9,420	9,420	9,420	9,420	9,420
Panel B continued. Boom Towns						
$\hat{\delta}$	0.180***	0.085***	0.067	-0.031	0.110***	0.032
	(0.064)	(0.025)	(0.062)	(0.046)	(0.040)	(0.032)
N	3,780	3,780	3,780	3,780	3,780	3,780

Dependent variable natural log of total employment in all regressions. County clustered standard errors shown. Treatment time period post 2007. Treated areas include all counties with shale oil and/or gas production as defined by EIA Drilling Productivity Reports. Control counties chosen using propensity score match from national sample in non-shale states. "Boom Town" is a combination of Permian, Anadarko, Eagle Ford, and Bakken plays. Parameters estimated with OLS. Source: Longitudinal Business Database

Table A3: Impact of Shale on Employment by Firm Age - by Shale Play

	(1) New Firms	(2) Young Firms	(3) New & Young Firms (1) + (2)	(4) Mature Firms	(5) Greenfield Estabs	(6) Incumbent Estabs	(7) Total (1) + (2) + (4) (1) + (5) + (6)
Appalacia							
$\hat{\sigma}$	-0.325***	0.134	-0.192	-0.542	0.008	-0.416	-0.733**
	(0.116)	(0.089)	(0.126)	(0.274)	(0.169)	(0.321)	(0.194)
N	3,780	3,780	3,780	3,780	3,780	3,780	3,780
Eagle Ford							
$\hat{\sigma}$	0.873	0.116	0.989	1.479	1.723*	-0.128	2.468
	(0.624)	(0.580)	(0.603)	(1.06)	(0.837)	(1.169)	(1.499)
N	690	690	690	690	690	690	690
Haynesville							
$\hat{\sigma}$	0.325	0.376	0.701	1.794***	0.368	1.802**	2.495***
	(0.341)	(0.272)	(0.510)	(0.687)	(0.305)	(0.768)	(0.880)
N	750	750	750	750	750	750	750
Niobrara							
$\hat{\sigma}$	0.227	0.048	0.275	0.133	0.721	-0.541	0.408
	(0.426)	(0.307)	(0.437)	(1.071)	(0.462)	(1.108)	(1.238)
N	1,100	1,100	1,100	1,100	1,100	1,100	1,100
Permian							
$\hat{\sigma}$	0.699	0.722	1.421**	1.933**	0.899*	1.756	3.354***
	(0.485)	(0.530)	(0.570)	(0.939)	(0.479)	(1.126)	(1.196)
N	1,620	1,620	1,620	1,620	1,620	1,620	1,620

County clustered standard errors shown. Treatment time period post 2007. Treated areas include all counties with shale oil and/or gas production as defined by EIA Drilling Productivity Reports. Control counties chosen using propensity score match from national sample in non-shale states. Parameters estimated with OLS. New firm age (in years) =0, young =1-4, old = 5+. Due to data confidentially constraints, we are unable to report individual play results for Bakken and Anadarko. Source: Longitudinal Business Database

Table A4: Impact of Shale on Employment Ratio

	(1) Pre-2007 establishments *(Incumbent Estabs)*	(2) New establishments 2007 to pre-2007 firms *(Greenfield Estabs)*	(3) New establishments of firms born 2007 and later *(New Firms)*	(4) Total *(All)*
Boom Towns				
δ_{2000}	0.0206	0	0	0.0206
	(0.0240)			(0.0240)
δ_{2001}	0.0206	0	0	0.0206
	(0.0229)			(0.0229)
δ_{2002}	0.0268	0	0	0.0268
	(0.0193)			(0.0193)
δ_{2003}	0.0087	0	0	0.0087
	(0.0212)			(0.0212)
δ_{2004}	0.0010	0	0	0.010
	(0.0164)			(0.0164)
δ_{2005}	0.0030	0	0	0.0030
	(0.0119)			(0.0119)
δ_{2007}	0.0114	0.0038	-0.0058	0.0094
	(0.0106)	(0.0094)	(0.0047)	(0.0151)
δ_{2008}	0.0382***	0.0104	0.0369	0.0855**
	(0.0127)	(0.0072)	(0.0350)	(0.0382)
δ_{2009}	0.0832***	0.0175**	0.0136**	0.1143***
	(0.0161)	(0.0080)	(0.0057)	(0.0185)
δ_{2010}	0.0808***	0.0226**	0.0145*	0.1179***
	(0.0153)	(0.0107)	(0.0077)	(0.0186)
δ_{2011}	0.0992***	0.0413***	0.0294***	0.1699***
	(0.0181)	(0.0146)	(0.0093)	(0.0259)
δ_{2012}	0.1090***	0.0563***	0.0828***	0.2481***
	(0.0199)	(0.0161)	(0.0188)	(0.0396)
δ_{2013}	0.1274***	0.0821***	0.1178***	0.3272***
	(0.0212)	(0.0203)	(0.0257)	(0.0509)***
δ_{2014}	0.1109***	0.0963***	0.1577***	0.3649***
	(0.0221)	(0.0219)	(0.0328)	(0.0603)
N	3,780	3,780	3,780	3,780

County clustered standard errors shown. Base year 2006 and therefore not shown in table. Treated areas include all counties with shale oil and/or gas production as defined by EIA Drilling Productivity Reports. Control counties chosen using propensity score match from national sample in non-shale states. "Boom Town" is a combination of Permian, Anadarko, Eagle Ford, and Bakken plays. Parameters estimated with OLS. New firm age (in years) =0, young =1-4, old = 5+. Employment ratio is defined as the ratio of a given group's employment in a given year to the total county employment for that group in the base year of 2006. Pre-treatment period is 2000-2006 and post-treatment period is 2007-2014. Source: Longitudinal Business Database

Table A5: Calibration Details

Parameter	Description	Model 1	Model 2
β	Discount factor	0.96	0.96
α	Returns to scale	0.8	0.8
γ	Labor supply elasticity	2	2
ρ_z	Firm TFP persistence	0.55	0.55
σ_z	Firm TFP dispersion	0.22	0.22
μ_c	Fixed operating cost mean	-6.7	-6.7
σ_c	Fixed operating cost dispersion	0.9	0.9
c_e	Entry cost	e^{μ_c}	$3e^{\mu_c}$
ξ	Entrant signal shape		2.69

Table A6: Comparison of Estimated Treatment Effects by Firm Age

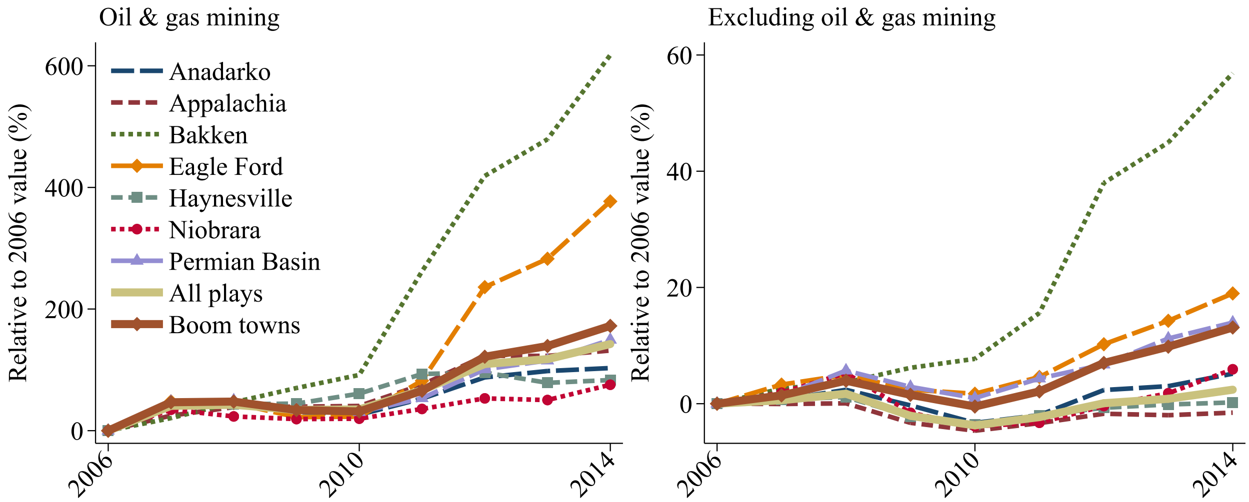
	(1) New Firms	(2) Young Firms	(3) New & Young Firms (1) + (2)	(4) Mature Firms	(5) Greenfield Estabs	(6) Incumbent Estabs	(7) Total (1)+(2)+(4) (1)+(5)+(6)
Propensity Score Match Control Group							
$\hat{\delta}$	0.321	0.359	0.680	1.149	0.340	1.167	1.829
Random Control Groups							
$\hat{\delta}_{minimum}$	0.254	0.103	0.458	0.575	0.522	0.151	1.163
$\hat{\delta}_{median}$	0.355	0.256	0.597	0.769	0.641	0.338	1.339
$\hat{\delta}_{maximum}$	0.480	0.367	0.763	1.038	0.801	0.717	1.707

Treatment time period post 2007. Treated areas include all counties with shale oil and/or gas production as defined by EIA Drilling Productivity Reports. Control counties chosen using propensity score match from national sample in non-shale states. Data across all industries and shale plays is used. Propensity score match group coefficient estimates are from Table 5. Revenue is expressed in hundreds of millions of dollars and rig count is expressed in hundreds of rigs. Parameters estimated with OLS. New firm age (in years) =0, young =1-4, old = 5+. Columns 1+2=3, columns 1+2+4=7, and columns 1+5+6=7. Source: Longitudinal Business Database

Table A7: Placebo Tests

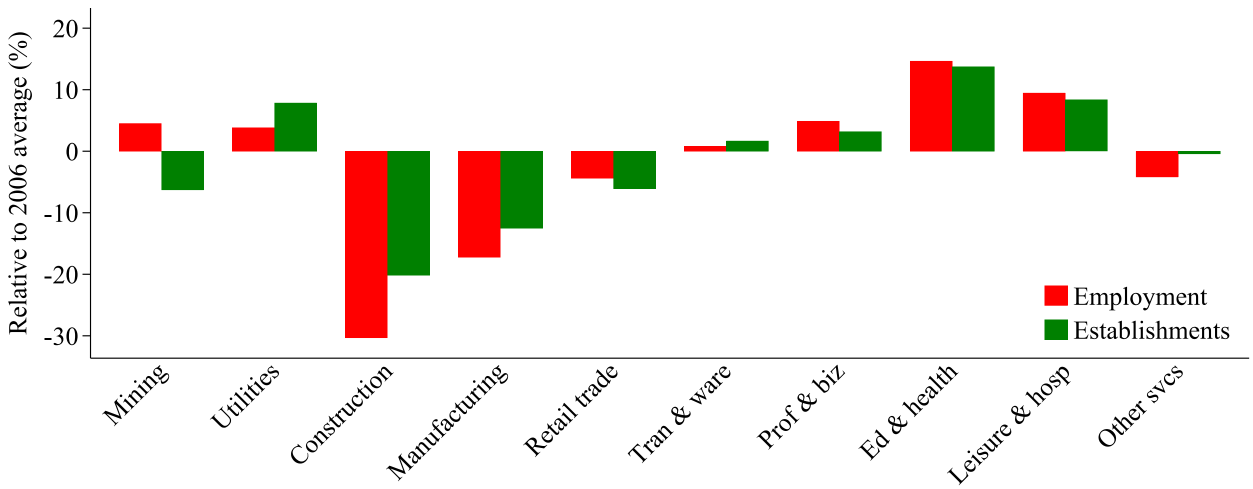
	(1) New Firm	(2) Young Firm	(3) New & Young Firm (1) + (2)	(4) Old Firm	(5) New- Existing	(6) Old-Existing	(7) Total (1)+(2)+(4) (1)+(5)+(6)
Panel A: Treatment Placebo							
$\hat{\delta}$	-0.136	-0.067	-0.203	-0.398	0.241	-0.706	-0.601
	(0.197)	(0.214)	(0.233)	(0.422)	(0.232)	(0.497)	(0.534)
N	4,710	4,710	4,710	4,710	4,710	4,710	4,710
Panel B: Control Placebo							
$\hat{\delta}$	-0.055	-0.073	-0.128	-0.368	0.041	-0.483	-0.497
	(0.145)	(0.120)	(0.170)	(0.351)	(0.177)	(0.339)	(0.405)
N	4,710	4,710	4,710	4,710	4,710	4,710	4,710

Dependent variable natural log of total employment in all regressions. County clustered standard errors shown. Treatment time period post 2007. Treated areas include all counties with shale oil and/or gas production as defined by EIA Drilling Productivity Reports. Control counties chosen using propensity score match from national sample in non-shale states. Parameters estimated with OLS. New firm age (in years) =0, young =1-4, old = 5+. Source: Longidutinal Business Database



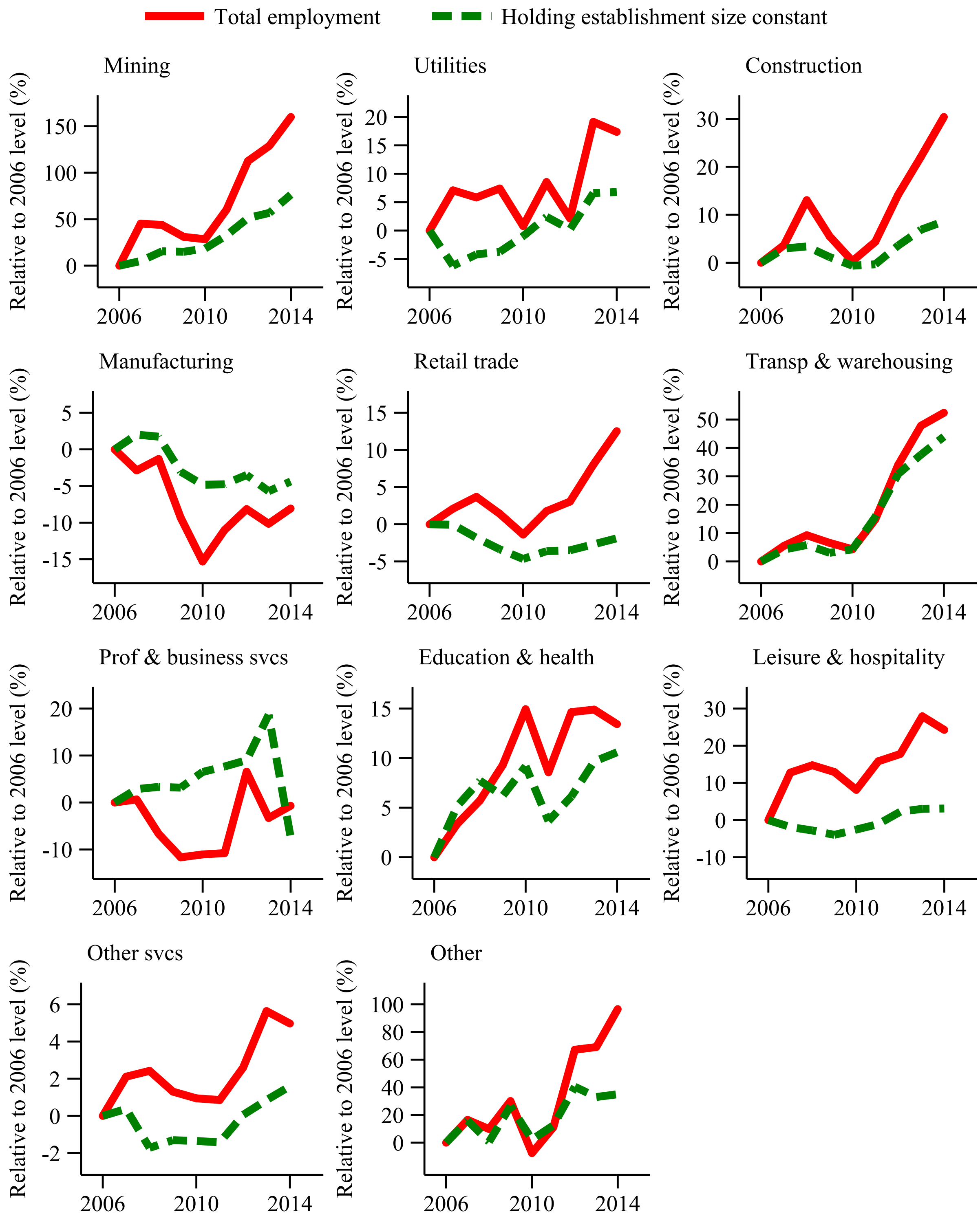
Note: Play-level employment relative to 2006. Boom towns are Anadarko, Bakken, Eagle Ford, and Permian Basin.
Source: County Business Patterns. Oil & gas mining is NAICS 211 and 213.

Figure A1: Employment gains by shale play



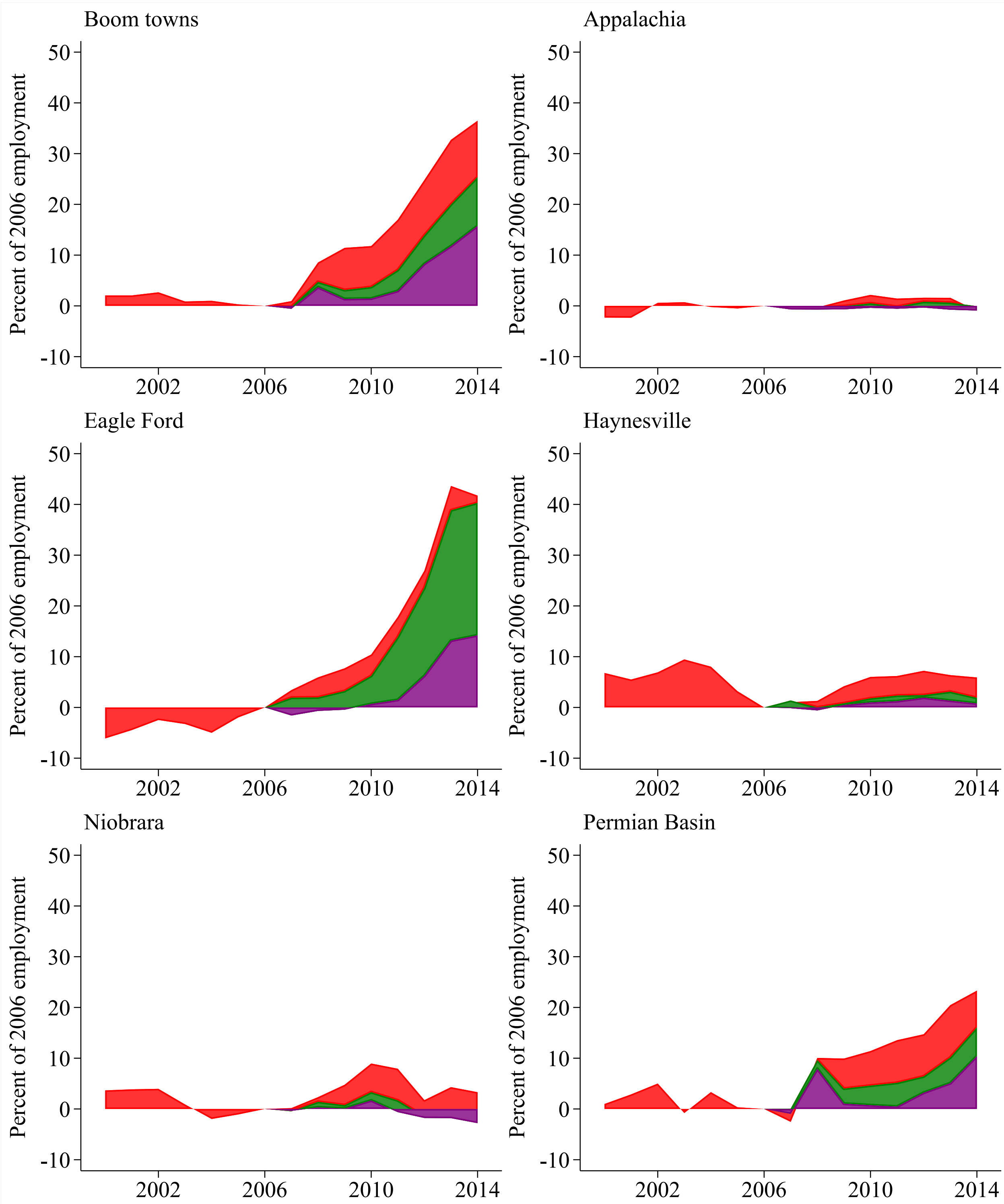
Note: Average county-level employment and establishment counts relative to average 2006 level. Excludes states with shale counties and states adjacent to shale counties. Residual industries omitted.
Source: County Business Patterns. NAICS sectors.

Figure A2: Control superset activity by sector, 2006-2014



Note: Total employment relative to year-2006 level. Covariance term shared out proportionally.
Includes Anadarko, Bakken, Eagle Ford, and Permian Basin.
Source: County Business Patterns. NAICS sectors.

Figure A3: Margins of employment growth in boom towns by sector



Employment scaled by 2006 county employment. Regression compares treatment and control counties with year effects. New establishments are establishments born after 2006. New firms are firms born after 2006. 'Boom towns' are Permian, Anadarko, Eagle Ford, and Bakken plays. Author calculations from LBD. Oil and gas mining (NAICS 211, 213) excluded.

Figure A4: Employment treatment effects by year

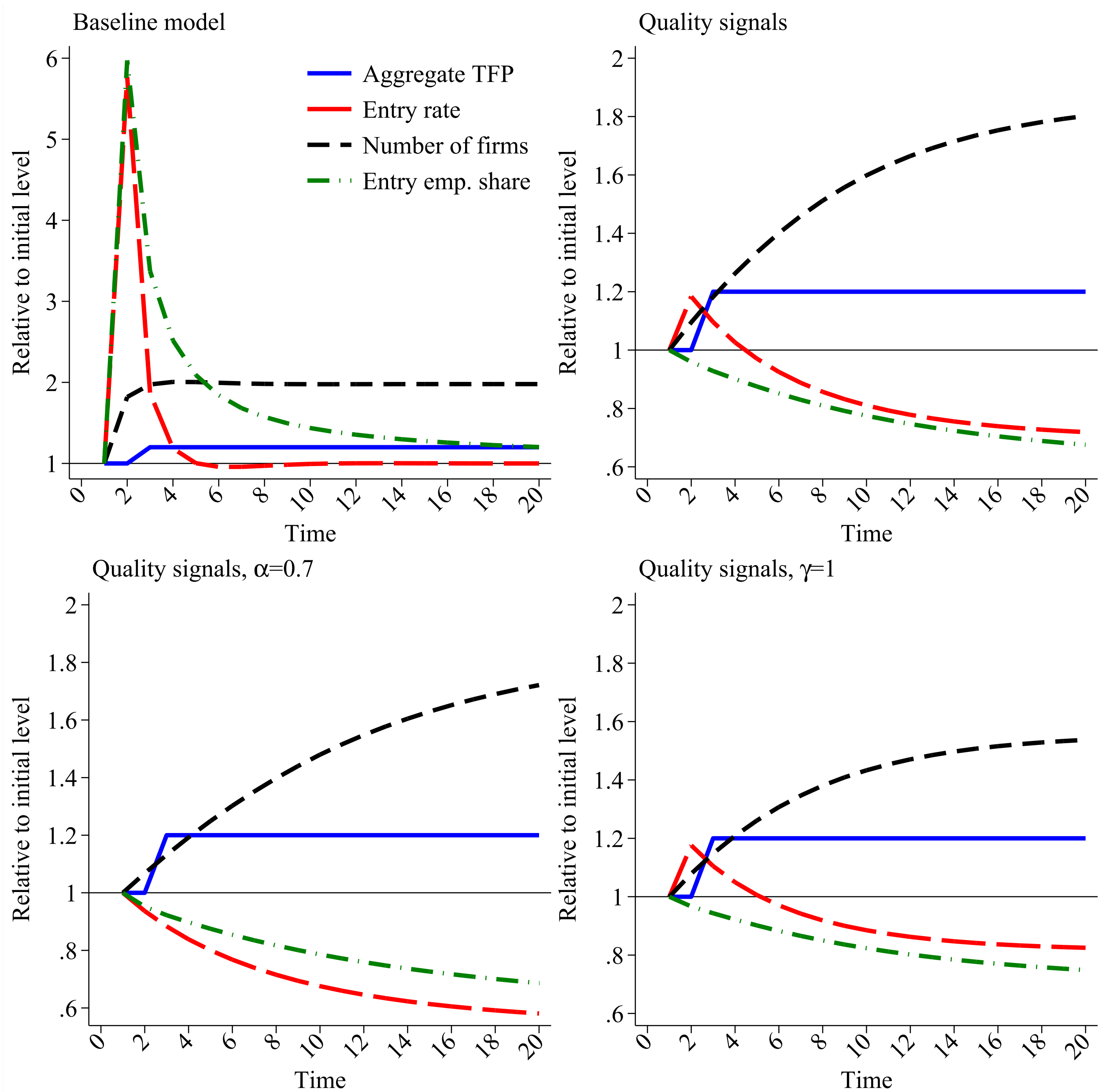


Figure A5: Model dynamics after aggregate productivity increase