Measuring the Employment Impacts of Shale Gas Development

Thomas DeLeire

Paul Eliason

Christopher Timmins

August 2014

Abstract

In recent years technological innovations in drilling and shale stimulation have produced a boom in the natural gas extraction industry across the portion of Pennsylvania that is situated on the Marcellus Shale. The development of this resource provides a relevant setting in which to study the effects of a natural resource boom on local labor markets. By employing a distributed lag model to estimate the impact of shale development on the labor market, we are able to identify and compare the short-run and long-run effects on employment and earnings. We also use quantile methods to allow for heterogeneous effects. To control for confounding factors, we employ synthetic controls in a way similar to a difference-in-difference. Our findings indicate that fracking has a positive and substantial effect on total employment but little impact on earnings, suggesting a slack labor market. In addition, we find evidence consistent with "Dutch disease"—a situation where a natural resource boom contributes to a contraction in the traded-goods sector. We also demonstrate how replacing the distributed lag model with a standard assumption that all shale development has an equal impact on employment regardless of timing can produce misleading results. Finally, we document that while on average fracking contributed only about 2% of job growth, in a few counties it was responsible for as much as 10% of total job creation.

Thomas DeLeire, McCourt School of Public Policy, Georgetown University, 37th and O Streets, N.W., 100 Old North, Washington, D.C. 20057 (*thomas.deleire@georgetown.edu*). Paul Eliason (Ph.D. Student) and Christopher Timmins, Department of Economics, Duke University, PO Box 90097, Durham, NC, 20078 (*paul.eliason@duke.edu* and *christopher.timmins@duke.edu*).

1. Introduction

Unconventional methods of extracting natural gas from shale formations have revolutionized the energy landscape in the United States. If fugitive emissions from the extraction process are properly controlled, natural gas provides an attractive alternative to fuels like coal as it emits less carbon and other pollutants per unit of heat produced. Given these advantages, new techniques for drilling and stimulating natural gas wells (e.g., the technique of hydraulic fracturing or "fracking") that dramatically lower production costs have become very popular. There are, however, potential risks to health and the environment that accompany this intensive extraction process. Fracking requires the injection of large quantities of water, accompanied by a mix of potentially harmful chemicals, into the ground at high pressure. This creates a risk of contamination of local water supplies due, for example, to faulty well casing or cement.[1] Homeowners dependent upon private wells are particularly vulnerable to this contamination risk, and research has shown that this risk is capitalized into housing prices.[2]

Despite these risks, fracking has proceeded forward at a rapid pace in many states (e.g., North Dakota, Texas, Louisiana, Colorado and Pennsylvania) and likely will do so in the not-too-distant future in several others (e.g., New York and North Carolina). This is largely a result of (i) lucrative leases signed by property owners in exchange for granting mineral access to drillers, and (ii) the prospects for job creation and increased economic activity. The latter is particularly important as shale resources are often located in economically depressed areas. It is this aspect of shale gas development that we explore in this research.

There are varied reports and claims predicting huge employment from fracking[3]. A projection published by IHS Global Insights, claimed that during 2012 the

[1] SEAB. 2011. Secretary of Energy Advisory Board, Shale Gas Production Subcommittee Second Ninety Day Report, November, 18. U.S. Department of Energy.

[2] In particular, Muehlenbachs, Spiller and Timmins 2014 identified costs from the risk of groundwater contamination as large as 22% of housing value for homes located in close proximity to drilling activity. See also Steck and Timmins (2014), "The Impact of the Fracking Boom on Rents in Pennsylvania."

[3] See "Fracking Nonsense: The Job Myth of Gas Drilling" and the studies cited therein. (http://www.cepr.net/index.php/blogs/cepr-blog/fracking-nonsense-the-job-myth-of-gas-drilling)

unconventional oil and natural gas industries in the US would support 1.7 million jobs.[4] This report predicted that number would increase to 2 million jobs by 2020. Tom Corbett, the current governor of Pennsylvania claimed in a recent opinion piece and in an advertisement for his re-election campaign that fracking has brought over 200,000 new jobs to Pennsylvania alone.[5] A report prepared by the Marcellus Shale Education & Training Center, a collaboration between the Pennsylvania College of Technology and Penn State, claims that fracking supported between 23,385 and 23,884 new jobs in 2009, an early year in development of the shale (Kelsey et al. 2011). Another industry group, the Marcellus Shale Coalition, reported in 2011 that shale development was responsible for 139,889 jobs at the end of 2010. It predicted that this number would grow to 181,335 at the end of 2012 and 256,420 by the end of 2020 (Considine *et al.* 2011). These numbers are all based on forecasts, extrapolations of industry surveys and analyses relying on arbitrary comparison groups.

In this paper we use detailed data from the Pennsylvania Longitudinal Employer-Household Dynamics (LEHD) program to estimate the effect of drilling activity on net job creation over time at the industry and county level. In addition to the magnitude and sign of the effect of drilling on employment, we are interested in the duration and composition of any realized job creation. Which industries experience job growth? Do the new jobs last? These are crucial questions to any analysis of the employment effects of fracking as they address concerns that a booming resource sector may have undesirable long-term consequences. In addition to estimating the marginal effect on the mean, we consider the marginal effects at other points of the distribution by using a quantile regression approach adapted to panel data. This allows us to consider the possibility of heterogeneous effects. To what extent are the effects of drilling uniform and to what extent do they vary across area? By answering these questions we provide a more complete picture of the labor market impacts of the fracking boom on counties in Pennsylvania.

---

[4] "America's New Energy Future: The Unconventional Oil and Gas Revolution and the US Economy." IHS Global Insights. 2012.

[5] See Op-Ed, "Protecting our environment growing our economy" (http://timesleader.com/news/energy-news/1028114/Protecting-our-environment-Growing-our-economy).

2. Background

Much of western Pennsylvania sits on top of the Marcellus Shale formation. This underground rock formation contains a rich reservoir of natural gas that, until recently, was largely untapped because the low permeability of the shale made natural gas extraction economically unviable. A series of technological innovations changed this. One such innovation is hydraulic fracturing or “fracking.” After the well is drilled, water, chemicals and sand are injected into the shale rock with extreme pressure. The pressure fractures the shale surrounding the wellbore and increases permeability. Other major innovations include reductions in the cost of directional drilling. Often, gas deposits are located deep beneath the earth’s surface (usually 5,000-20,000 feet down). After a vertical wellbore is drilled down to depth, horizontal drilling is used to tap greater portions of the broad shale formation. This helps reduce both cost and surface disruptions per unit of natural gas extracted from the shale.

Beginning in the mid-2000s these unconventional methods increased the profitability of natural gas wells. Figure 2.1 shows that after these technologies became available, the number of wells spudded (“spudding” refers to the initialization of the process of drilling a new well) grew quickly. Today, the number of new wells being spudded remains high, though is lower than at its peak at the end of 2011. In 2011, 1,964 new wells were spudded in Pennsylvania alone. Through 2012, 39 (of 67) Pennsylvanian counties had natural gas wells, though the majority of wells were located in 6 counties.

Figure 2.1A

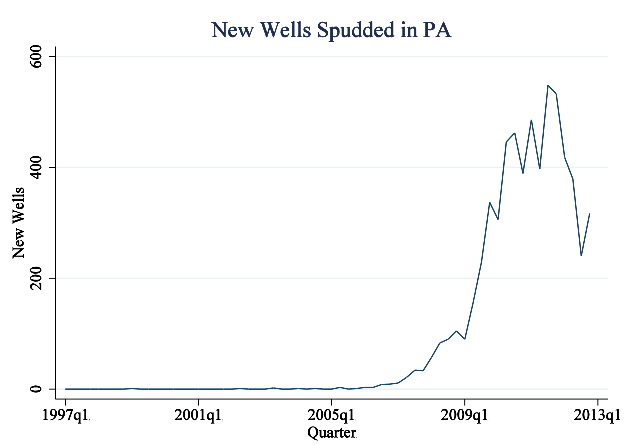
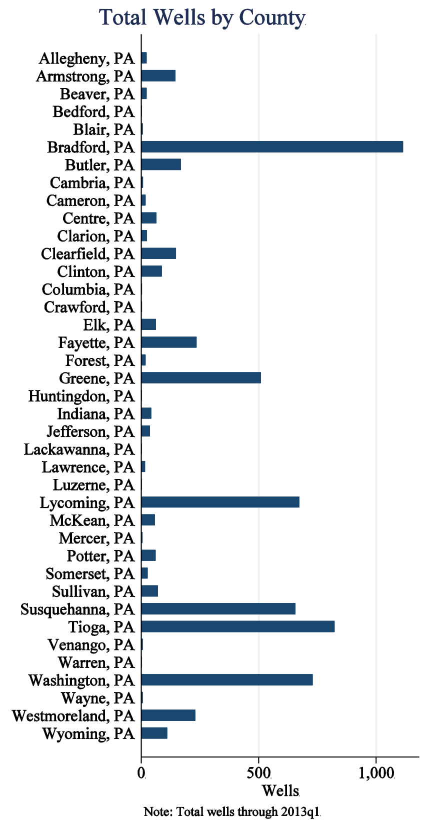
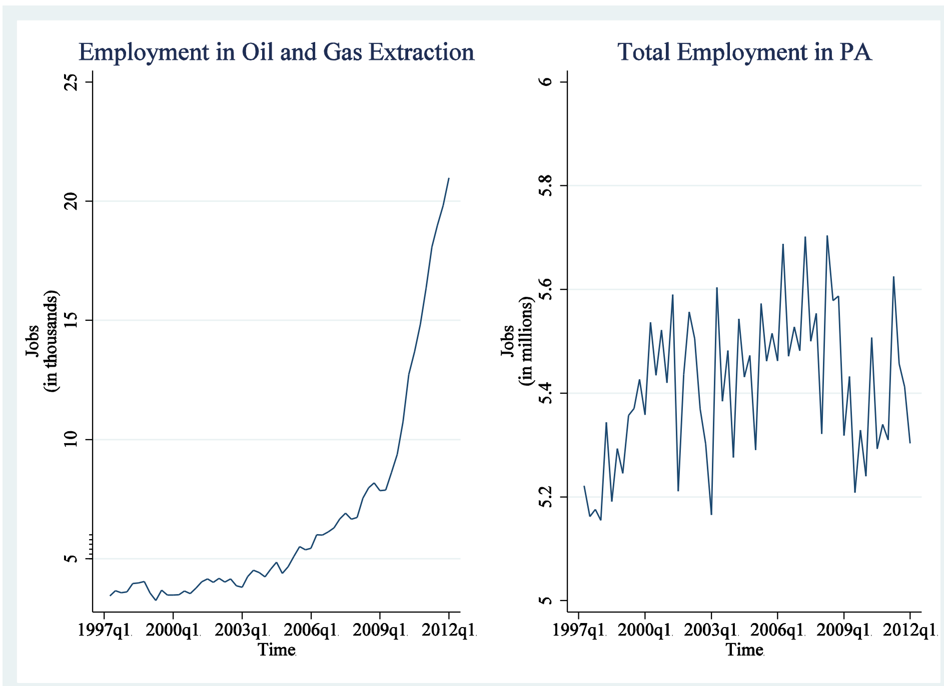


Figure 2.1B



Broadly speaking, there are three channels through which fracking could affect local employment. We call these the "direct", "indirect", and "induced" effects. While we do not attempt to estimate these channels separately, it is instructive to briefly consider each. The direct effect follows from the employment demand generated by the need for workers directly involved in the process of bringing a new well online and maintaining its operation. Drilling and stimulating a new well is a labor intensive task that requires workers across a broad array of skill sets. One report stated that a single Marcellus well requires about 420 individuals from 150 different occupations (Brundage et al. 2011). That report concludes that 80% of this labor is used during the drilling and stimulation stages of the well's life. It also notes that unconventional drilling and stimulation methods require a very specific skill set. During the early stages of development of the Marcellus Shale, no companies based in Pennsylvania had this expertise. Consequently, a large portion (the authors claim as much as 70-80%) of workers directly involved in bringing new wells online were transient out-of-state workers. It is believed that this portion has been declining over time as resident firms acquire these skills, but data on this are scarce. The left panel of Figure 2.2 shows growth in Pennsylvania employment counts in the oil and natural gas extraction industry.

Figure 2.2



Source: Pennsylvania LEHD

Development of the Marcellus Shale is thought to affect local labor markets through other, less direct channels. These indirect effects flow from goods and services that support but are not directly involved in the fracking process. For example, the construction industry may expand in response to the need to build new and maintain existing roads and infrastructure. Wholesalers may experience increased demand if they supply materials to the natural gas extraction industry.

Finally, the induced effect is realized from labor demand shifts associated with new wealth in the community. Transient workers fill hotel rooms and patronize restaurants and retail stores. Property owners receive lease and royalty payments from natural gas companies. This increased wealth may lead them to demand more goods and services. Increased tax revenues may also lead to new public projects and employment opportunities in the public sector. In our analysis, these three channels will be evident in our consideration of the impacts of shale gas development on different industries (e.g., construction versus retail), but we will not attempt to parse them out further.

3. Previous Literature

Earlier in the paper, we cited a number of studies predicting large employment gains as a result of the rapid growth of shale gas in the United States. There have been fewer direct analyses, however, of realized employment impacts in places where hydraulic fracturing has been going on for years. This deficit exists despite there being well-established techniques that have been applied in other settings to evaluate the employment spillover effects from exogenous shocks to local employment, such as those caused by energy booms and the entry of a large production plant. Marchand (2012), for example, examines the differential growth in employment and earnings between local labor markets with and without energy resources in Western Canada. Focusing on periods of booms (1971-1981, 1996-2006) and bust (1981-1991) in energy markets, he finds significant evidence of impacts on employment and earnings in energy sectors and modest impacts in non-energy sectors (particularly construction, retail trade, and services) during boom periods.

Moretti (2010) estimates a long-term employment multiplier—i.e., the long-term change in the number of jobs in tradable and non-tradable sectors generated by an exogenous increase in the number of jobs in the tradable sector. He develops a conceptual framework that predicts an increase in employment in local non-traded sectors (e.g., restaurants, real estate, cleaning services, construction, medical services, and retail) in response to an exogenous shock to employment in a traded goods sector. This response is a simple result of the shock leading to more workers using their additional wages to demand more goods and services in the locality. The size of this effect depends upon consumer preferences for non-tradables, the type of jobs added with the exogenous shock (i.e., skilled v. unskilled), and offsetting equilibrium impacts on wages and prices, which depend upon the elasticities of local labor and housing supply.

While the effect on the local non-tradable sector is theoretically clear in Moretti's framework, the effect on other tradable good sectors is *a priori* unclear. In particular, the city-wide increase in labor costs caused by the exogenous shock hurts employment in other tradable goods sectors. Because tradable goods sectors' prices are determined outside the local area, they do not adjust to reflect local economic conditions. Some of the employment in these sectors will ultimately be shifted to other locations where labor demand is not as high. Moretti tests this conceptual framework using a model similar to ours, which we describe in sub-section 4.1.

Closely related to the concept of resource booms and local labor market shocks is the "Dutch disease" model from Corden and Neary (1982). This model suggests there should be two effects from expansion of resource sector: (1) a resource movement effect—increased labor demand in the resource sector pushes wages up and raises production costs for non-tradable and tradable sectors (causing employment share in each of these sectors to contract) and (2) a spending effect—higher wages drive up local incomes, inducing an increase in demand for tradables and non-tradables for given prices. These effects can be (partially) offsetting for the non-traded sector. The price of inputs may go up, but increased demand for non-traded goods may partially or fully offset this effect. Thus the net effect on the non-traded sector cannot be signed a priori. When demand for traded goods increases, the local firms in this sector do not benefit. These firms face a contraction in employment without an offsetting increase in demand. This contraction may be further exacerbated by a booming non-traded sector, leading to an expectation that this sector will contract.

Black, McKinnish and Sanders (2005) look at local shocks associated with a resource boom or bust. In the 1970's, there were a series of oil shocks driven by political turmoil in the Middle East. These shocks drove up the price of coal in the US, and created a positive economic shock in areas that had coal resources to exploit. By 1983, prices had dropped and the boom turned into a bust. Since oil shocks came from outside the local community, it is reasonable to think that they would be uncorrelated with other local unobservable shocks, and therefore provide an exogenous source of variation with which to identify effects that ripple throughout the local labor market. How were non-mining sectors affected by shocks to the mining sector? How did these effects differ between sectors producing local goods and those producing traded goods? The answers to these questions have broader implications for plant closures, government incentives for locating a business, and other local economic shocks on other sectors. Measuring these effects is difficult, because we typically don't have a valid counterfactual (i.e., what would have happened had the plant not been built). Black, McKinnish and Sanders (2005) look within the Appalachian region and compare counties with and without coal resources—using the latter as a control group in a difference-in-differences analysis. Producers of local goods will experience an increase in labor costs, but they will also experience an increase in demand. On the whole, this could lead to an increase in employment. Local firms that sell their output nationally or internationally, however, may be expected to contract as they experience an increase in labor costs but no offsetting increase in demand. The authors find evidence of modest employment spillovers into sectors with locally traded goods but not into sectors with nationally traded goods. In particular, one mining job created during the boom period creates 0.174 local sector jobs, while one mining job lost during the bust period destroys 0.349 local sector jobs. No evidence of any ripple effects (positive or negative) is found for traded goods sectors, suggesting no evidence of the "Dutch disease". Significant effects on earnings, poverty, and the age and gender composition of the population are found, however, using the same difference-in-differences strategy.[6]

Fetzer (2014) uses variation in the location of exploitable shale deposits to identify impacts of shale gas development on employment in different sectors. By

[6] This analysis raises an interesting question about the use of national resource shocks to identify local impacts—particularly those associated with industries that sell nationally (or internationally). If the whole country is undergoing an oil price shock, might we think manufacturers who sell nationally would contract their employment anyway? When oil prices drop, the resulting boost to the national economy might lead them to increase their sales for reasons that have nothing to do with drops in local wages.

relying on cross-sectional variation in drilling activity (observed in 2012), this paper looks for long-run changes in employment that may be attributed to technological innovations that made previously unviable deposits viable. Despite rising labor costs, he does not find evidence of Dutch disease (i.e., contraction) in the tradable goods sector, while the non-tradable goods sector does contract. He claims that this arises because cheaper energy is providing a source of local comparative advantage. This should be particularly true in places with pipeline constraints—binding outflow capacity forces the extracted gas to be consumed locally, forcing down local gas prices. He also shows that a drop in energy prices is enough to offset increased labor costs, which can explain why the local non-traded goods sector contracts while the tradable goods sector does not.

Allcott and Keniston (2014) also look for evidence of the Dutch disease as a result of resource booms. Using panel data on US counties going back to the 1960's, they develop a county level measure of resource abundance. Using an empirical strategy similar to that of Bartik (1991), they then exploit national resource booms and busts (measured by national level employment in oil and gas). In particular, they estimate the effect of those shocks, multiplied by county-level resource abundance, on county-level economic measures, including employment in oil and gas, other sectors, wages, and factor productivity. They use a reduced form, and without the accompanying first-stage regression, one cannot draw structural interpretations. The results do indicate, however, that a resource boom that doubles national employment in oil and gas will increase total employment in a county with one standard deviation larger oil and gas endowment by 3.5 percent. Wages also rise, suggesting the possibility for Dutch disease. There is, however, no such evidence—manufacturing employment, revenues, number of establishments and capital investments turn out to be pro-cyclical with oil and gas. Exploring this result further, the authors find evidence that manufacturers who benefit most are those that are upstream of the oil and gas sector, and benefit by supplying it with inputs, and those that sell directly to the local population. Those that show no benefits are those with low transportation costs, which can sell more easily in other markets. However, even these firms do not suffer costs, as their labor appears to be less substitutable with that used in the resource sector.

Finally, Maniloff and Mastromonaco (2014) use a panel dataset containing economic outcome data such as employment and wages as well as oil and natural gas data to quantify the local economic impacts of fracking activity across the United States. Their empirical strategy uses a first-differenced model to estimate the effect of changes in shale development between 2000 and 2010 on changes in an economic outcome over the same period. Relying on this low-frequency variation they find significant, though somewhat small impacts of shale development on economic outcomes. Breaking down their results by industry, they focus on the traded-goods sector in order to look for evidence of Dutch disease. They find no statistically significant effect of shale development on wages or employment in this sector.

3. Data

3.1 Pennsylvania LEHD

Our data on employment and earnings outcomes come from the Longitudinal Employer-Household Dynamics (LEHD) program. LEHD is a product of a collaborative effort between the Census Bureau and state unemployment insurance agencies. It uses unemployment insurance administrative records, other administrative data and data from censuses and surveys to link employers to employees and create extremely detailed data on local labor markets. Although the source data of LEHD is at the individual/firm level, we use a publicly available product called the Quarterly Workforce Indicators (QWI). QWI is a panel data set containing 30 quarterly employment aggregates at the county level within very specific industries. These indicators include data on worker flows (hires, separations, turnover,…), employment levels, and earnings data. The data have been aggregated from the individual /firm level up to the county/industry level at a quarterly frequency, allowing the researcher to identify labor market trends within or across industries and over time. One outcome of particular interest is net job flows, that is, to investigate how many jobs are added to local economies because of the growth in the fracking industry Net job flows for a firm are consistent with the following identity:

$$Firm\,Job\,Change_{it} = Firm\,Job\,Gains_{it} - Firm\,Job\,Loss_{it} = Hires_{it} - Separations_{it}$$ [7]

Other relevant outcomes include earnings and employment levels.

There is an important caveat related to measurement error in these data that applies specifically to the fracking industry in Pennsylvania. While the history of shallow oil and gas extraction in Pennsylvania goes back a long time, the Marcellus Shale was not developed intensely until hydraulic fracturing methods were developed and made available in the mid-2000s. The "unconventional" techniques required to develop shale resources, such as directional drilling and stimulation of oil/gas production, require very specialized skills. Very few Pennsylvanian companies have this expertise and it is very common for energy companies and contractors to be brought in from out of state to develop the shale. As mentioned above, one estimate places the portion of shale exploration and development employees from out of state at 70-80% (Brundage *et al.* 2011). Many of these workers are transient. It is also presumed that over time, as resident PA employees and firms acquire experience with these unconventional techniques, this portion will decrease. Because the LEHD sampling is based on state UI records, it is unclear what portion of these out-of-state workers are included in the PA data. Therefore, we must be aware of the possibility that, in the drilling and gas extraction industries, the Pennsylvania data may systematically undercount the number of workers.

Industries are classified according to the North American Industry Classification System (NAICS). Employees at an establishment are classified into industries based on the primary form of business at that establishment. We focus on the most general NAICS codes: those denoted by two-digit codes. These include: construction, manufacturing, and mining, quarrying, and oil & gas extraction. The Pennsylvania QWI includes employment figures for all NAICS codes that are present in the state. Table 3.1A contains a summary of quarterly net job flows by industry for the 39 counties on the shale.

[7] See "LED: Quarterly Workforce Indicators 101", available at: http://lehd.ces.census.gov/applications/qwi_online/

Table 3.1A: Summary Statistics: Net Job Flows by Industry, 1997-2011

	Count	Mean	Std Dev
Total Employment	2301	267.3068	2734.203
Agriculture, Forestry, Fishing and Hunting	2236	.6690519	34.23508
Mining, Quarrying, and Oil and Gas Extraction	2077	8.70053	79.20391
Utilities	2155	-.0709977	82.59438
Construction	2301	24.54846	488.0518
Manufacturing	2301	-20.48805	355.0948
Wholesale Trade	2300	8.675652	112.7131
Retail Trade	2301	30.98827	636.5299
Transportation and Warehousing	2287	11.10582	203.2477
Information	2271	.7058565	199.0753
Finance and Insurance	2287	5.836467	450.1261
Real Estate and Rental and Leasing	2214	3.316621	73.75559
Professional, Scientific, and Technical Services	2263	20.83031	205.2115
Management of Companies and Enterprises	1932	1.766563	177.5202
Administration and Support and Waste Management and Remediation Services	2237	25.10237	378.6373
Educational Services	2297	24.22203	753.6621
Health Care and Social Assistance	2301	60.31595	1045.031
Arts, Entertainment and Recreation	2210	22.06516	515.0533
Accommodation and Food Services	2301	26.35072	366.2785
Other Services	2301	10.76271	133.971
Public Administration	2301	5.480661	263.5775
Observations	2310		

Note: Sample includes only counties on the Marcellus Shale.

Table 3.1B: Summary Statistics: Average Monthly Earnings by Industry, 1997-2011

	count	Mean	sd
Total Employment	2340	2652.245	355.5052
Agriculture, Forestry, Fishing and Hunting	2340	1753.468	526.2942
Mining, Quarrying, and Oil and Gas Extraction	2326	3606.32	1454.65
Utilities	2339	4512.11	1374.507
Construction	2340	2826.31	509.5366
Manufacturing	2340	3310.409	636.4161
Wholesale Trade	2340	2993.995	820.2565
Retail Trade	2340	1568.049	258.1799
Transportation and Warehousing	2340	2230.468	563.882
Information	2340	2371.067	876.848
Finance and Insurance	2340	2886.411	678.6422
Real Estate and Rental and Leasing	2336	1824.88	820.5267
Professional, Scientific, and Technical Services	2340	2869.698	997.7314
Management of Companies and Enterprises	2339	2882.882	1948.379
Administration and Support and Waste Management and Remediation Services	2332	1665.078	556.5345
Educational Services	2340	2755.038	518.981
Health Care and Social Assistance	2340	2315.187	454.1971
Arts, Entertainment and Recreation	2332	1025.335	355.0262
Accommodation and Food Services	2340	802.4885	180.4559
Other Services	2340	1350.069	387.7079
Public Administration	2340	2449.15	562.2804
Observations	2340		

Note: Sample includes only counties on the Marcellus Shale.

Standard deviations are very large across all industries. This is being driven mainly by cross-sectional variation. The sample includes Allegheny County, home of Pittsburgh and the largest population center in the region (over 1.2 million), as well as Cameron County, which according to the Census is home to 5,085 individuals. The discrepancies in sample size reflect missing data. In these cases, most of the missing data was suppressed by the US Census Bureau because it did not meet their confidentiality requirements.

The period covered by these data is economically complex, containing years of sustained economic growth, the financial collapse, and subsequent recession. However, the positive means for most industries suggest that, on average, this was a period of expansion. The notable exception is in manufacturing; on average, counties lost 20.5 manufacturing jobs per quarter over this period.

Table 3.1B presents similar summary statistics for average monthly earnings. The Mining, Quarrying, and Oil and Gas Extraction industry is among the highest in average monthly earnings. Earnings in Accommodation and Food Services is the lowest. Some of this difference comes from the fact that the data do not differentiate between full-time and part-time workers.

3.2 Drilling Activity

We obtain a detailed panel data set on wells in Pennsylvania from DrillingInfo, an oil and gas industry analytics and data provider. The data include detailed information on the location and spud date of each well. The dataset includes and differentiates between both vertical and horizontal wells—the vast majority are horizontal wells. These wells have a significantly more labor intensive drilling process and are likely to have larger employment spillover effects. For the purposes of this analysis, we aggregate all well spuds up to the level of the county and quarter.

3.3 County Level Covariates

We also incorporate county characteristics into the estimation procedure. These include population density, median household income, percent of population with a college degree, median age, unemployment rate, and county sectoral mix. These data all come from the *County and City Data Book*, 2000, a publication of the US Census Bureau.

4. Empirical Methods

In order to explore the local labor market implications of shale gas development, we estimate a series of econometric models. We build off of the simple and intuitive model proposed by Moretti (2010):

$$dY_{j,c,t} = \alpha + \beta dY_{i,c,t} + \tau_t + \varepsilon_{j,c,t} \tag{1}$$

$dY_{j,c,t}$ is the number of net jobs flows for industry *j*, county *c*, over quarter *t*. This model is used to describe an employment spillover into industry *j* from an exogenous shock to net job flows in industry *i* (e.g., employment shocks in the mining, quarrying, and oil and natural gas extraction industry). α is a constant describing the mean net job flows for industry *j* and τ_t is a period fixed effect. Both are common to all counties.

When applied to study the effects of shale gas development on local labor market outcomes, equation (1) is limited in a number of ways. First, the net job flow data from the natural gas extraction industry may suffer from systematic measurement error, which would cause our estimates of β to overstate their true values. Second, our aim is to study the short-, medium-, and long-run effects of shale gas development on labor market outcomes. Equation (1), however, only allows for uniform contemporaneous employment spillovers between industries. Third, since net job flows for industries *j* and *i* are determined simultaneously, the model will suffer from endogeneity.

We first replace equation (1) with a distributed lag model:

$$dY_{j,c,t} = \beta_0 dW_{c,\mathrm{t}} + \beta_1 dW_{c,t-1} + \beta_2 dW_{c,t-2} + ... + \theta_t Z_c + \upsilon_t \lambda_c + \varepsilon_{j,c,t} \tag{2}$$

where $dW_{c,t}$ is the number of new wells drilled in county *c* during period *t*. While similar in spirit to equation (1), this specification addresses many of the shortcomings of our empirical setting. Using data on new wells rather than net job flows in the natural gas extraction industry circumvents the issue of systematic measurement error in that industry except when that is the outcome industry. In addition, using new wells as a regressor broadens the number of channels through which development of shale resources may impact employment. Equation (1) takes changes in the boom industry as given. It fundamentally cannot include in the estimation a significant portion of the direct effect on employment. Equation (2), on the other hand, will estimate a net effect of the three channels through which an additional well stimulates employment in an industry.

Endogeneity is still a concern because the location and timing of new wells is nonrandom. However, by controlling for observable covariates Z_c and a fairly general set of unobservable covariates, we can be confident that $dW_{c,t}$ is conditionally uncorrelated with the residual ($\varepsilon_{j,c,t}$). This set of unobservables includes time-varying effects that are common across counties and also county-specific unobservables (λ_c) that have time-varying factor loadings (υ_t). Note that controls for time-invariant and county specific effects as well as effects that are constant across counties but vary over time are special cases of this factor loading specification. We use synthetic control methods to control for these potentially confounding variables as discussed below.

A distinguishing feature of our empirical strategy is the inclusion of a series of new well lags on the right-hand side. These lags allow us to track the impact of a single new well on industry employment over time. Identification of the parameters $\{\beta_0, \beta_1, ...\}$ will enable a clean differentiation between short-run and long-run effects.

A standard empirical approach for estimating the impact of a natural resource boom on local economic activity is to use a model based on first-differences over a time-horizon lasting a set number of years. (For examples specific to oil and natural gas see Allcott and Keniston (2014), Maniloff and Mastromonaco (2014), Marchand (2012) and Weber (2013)). In the context of shale development, these strategies typically estimate the total effect of shale development (e.g., by using an indicator of treatment, total oil and gas production, total wells, etc.) on the total change in the outcome of interest from the before the boom through the end of the boom. In this context, those strategies could be represented by the following model:

$$dY_{j,c,2000-2010} = \gamma dW_{c,2000-2010} + \theta_t Z_c + \upsilon_t \lambda_c + \varepsilon_{j,c,t} \tag{3}$$

Equation (4.3) can be interpreted as quantifying the mean effect of each new well installed during the interval from 2000 to 2010 on the change in employment from 2000 to 2010. By pooling new wells across time this coarse approach ignores when the well was installed and how long it has been operating. It identifies only an average of the impacts of each new well on the total change in employment over the entire horizon. The resulting estimate cannot be classified as a short-run or long-run effect because it is actually an average of them both—if new wells affect employment in a time-varying manner then the estimate for γ will depend on the timing of the new well installations. For example, if (as is the case) most of the new wells are spudded toward the end of the interval considered, then the estimate will pick up more of the short-run effects. If the bulk of the new wells were spudded towards the beginning of the interval then the estimate will include more of the long-run effect.

Inclusion of the distributed lag terms in equation (2) improves upon the standard approach described in (3) by allowing for the estimation of a rich set of policy-relevant parameters. These parameters let the researcher differentiate between employment effects that are contemporaneous with spudding and those that are lagged. The main drawback to inclusion of a large (and theoretically infinite) number of lagged terms is the identification challenge associated with the large number of covariates. We address this in the following section.

4.1.1 Identification of Distributed Lag Terms

Equation (2) is an example of a distributed lag model. Distributed lag models are appealing because they do not impose an arbitrary cutoff for the impulse response, but rather allow the data to dictate how quickly it goes to zero. However, allowing for an infinite number of RHS parameters complicates identification. We use a polynomial inverse lag (PIL) model to impose structure on the parameters estimates for these lagged effects in order to achieve identification.

The PIL model imposes structure on the β_i 's by assuming they have the following form:

$$\beta_i = \sum_{j=2}^{n} \frac{a_j}{(1+k*i)^j} \tag{4}$$

$\{a_j\}_{j=2}^{n}$ are structural parameters that describe the impulse response to the introduction of a new well (i.e. how does employment increase over time). Note that the *j* subscript runs from 2 to *n*. Unlike other distributed lag models, this model allows for a very flexible response. Using *n*=3 allows for diminishing effects. Using $n \geq 4$ allows for effects that are non-monotonic over time. In any case, the PIL model is designed so that the lagged effects eventually go to zero. When we estimate this model, we focus on the case where *n* = 4 because it balances the constraint of a limited sample size with the desirability of a flexible impulse response function. The *k* in the denominator is a parameter that determines how quickly the imposed polynomial term plays out and the lagged effects go to zero. Rather than explicitly estimate this parameter simultaneously with the rest of the parameters, we estimate the model conditioning on a fixed value of *k.* We repeat this for a range of integer values for *k* and pick the one that best fit the data (highest R-squared), which was k = 10.

In order to see how the model works, we write it out explicitly for the case of $n = 4$ and $k = 1$.

$$\beta_i = \frac{a_2}{(1+i)^2} + \frac{a_3}{(1+i)^3} + \frac{a_4}{(1+i)^4} \tag{5}$$

Incorporating this expression into the original estimation equation:

$$dY_{j,c,t} = [a_2 + a_3 + a_4]dW_{c,t} + \left[\frac{a_2}{2^2} + \frac{a_3}{2^3} + \frac{a_4}{2^4}\right]dW_{c,t-1} + \left[\frac{a_2}{3^2} + \frac{a_3}{3^3} + \frac{a_4}{3^4}\right]dW_{c,t-2} + \left[\frac{a_2}{4^2} + \frac{a_3}{4^3} + \frac{a_4}{4^4}\right]dW_{c,t-3} + ... + \varepsilon_{j,c,t} \quad (6)$$

Rearranging terms yields:

$$\begin{aligned} dY_{j,c,t} &= a_2(dW_{c,t} + \frac{dW_{c,t-1}}{2^2} + \frac{dW_{c,t-2}}{3^2} + \frac{dW_{c,t-3}}{4^2} + ...) + \\ &\quad a_3(dW_{c,t} + \frac{dW_{c,t-1}}{2^3} + \frac{dW_{c,t-2}}{3^3} + \frac{dW_{c,t-3}}{4^3} + ...) + \\ &\quad a_4(dW_{c,t} + \frac{dW_{c,t-1}}{2^4} + \frac{dW_{c,t-2}}{3^4} + \frac{dW_{c,t-3}}{4^4} + ...) + ... + \varepsilon_{j,c,t} \quad (7) \\ &= a_2\Omega_{2,t} + a_3\Omega_{3,t} + a_4\Omega_{4,t} + ... + \varepsilon_{j,c,t} \end{aligned}$$

Typically, finite representations of the $\Omega_{j,t}$'s are used to approximate the actual values. However, since the total number of lags with positive numbers of fracked wells in PA is finite, we can use our data to calculate the true values of $\Omega_{j,t}$'s:

$$\begin{aligned} \Omega_{2,t} &= \left[dW_{c,t} + \frac{dW_{c,t-1}}{2^2} + \frac{dW_{c,t-2}}{3^2} + ... + \frac{dW_{c,T_{0,c}}}{(t - T_{0,c} + 1)^2}\right] \\ \Omega_{3,t} &= \left[dW_{c,t} + \frac{dW_{c,t-1}}{2^3} + \frac{dW_{c,t-2}}{3^3} + ... + \frac{dW_{c,T_{0,c}}}{(t - T_{0,c} + 1)^3}\right] \quad (8) \\ \Omega_{4,t} &= \left[dW_{c,t} + \frac{dW_{c,t-1}}{2^4} + \frac{dW_{c,t-2}}{3^4} + ... + \frac{dW_{c,T_{0,c}}}{(t - T_{0,c} + 1)^4}\right] \end{aligned}$$

where $T_{0,c}$ denotes the first period in which a well was drilled in county *c*. In other words, for any county *c* and for all $r > 0$, $dW_{c,T_{0,c}-r} = 0$. From here it is a simple matter of estimating equation (7) and backing out the parameters in (2).

4.1.3 Controlling for Unobservables Using Synthetic Controls

To protect against correlation between new wells and the error term in factor model (4.2) want to control for unobservables that are county-specific and have time-varying effects. Suppose the counterfactual outcomes (i.e., the labor market outcomes that would have been realized in the absence of shale development), can be described by the factor model:

$$dY^0_{j,c,t} = \theta_t Z_c + \upsilon_t \lambda_c + \varepsilon^0_{j,c,t} \tag{9}$$

With data for $dY^0_{j,c,t}$ in hand, controlling for unobservables would be a simple matter of differencing. However, this is a case of the classic missing data problem: we do not observe both $dY_{j,c,t}$ and $dY^0_{j,c,t}$ simultaneously. As a feasible alternative, we estimate $dY^0_{j,c,t}$ following an approach outline by Abadie, Diamond and Hainmueller (2010) (ADH).

We introduce some notation before outlining their approach. Let $t = 1, 2, ..., \mathrm{T}_0$ be the set of time periods before fracking began. Let *J* denote the set of counties with shale resources that can potentially be developed and let *S* denote the "donor" group, that is, the set of counties without shale resources. For each county in *J*, we will recover a synthetic county that will be a weighted average of the counties from the donor pool. Fix a county, $i \in J$. Suppose there is a set of weights $W^* = \{w^*_s\}_{s \in S}$ such that:

$$\begin{aligned} &\sum_{s \in S} w^*_s dY_{j,s,t} = dY_{j,i,t} \ \forall\, t = 1, 2, ..., T_0 \qquad \textit{and} \\ &\sum_{s \in S} w^*_s Z_s = Z_i \end{aligned} \tag{10}$$

The key result of ADH is that if (10) holds and (2) and (9) describe the true model, then under standard conditions[8]:

$$dY_{j,i,t}^{0} - \sum_{s \in S} w_s^* dY_{j,s,t} = 0 \tag{11}$$

in expectation for $t > T_0$. Denote $d\tilde{Y}_{j,s,t}^{0} \equiv \sum_{s \in S} w_s^* dY_{j,s,t}$. Thus $d\tilde{Y}_{j,s,t}^{0}$ is a natural estimator for $dY_{j,i,t}^{0}$. Estimation of $dY_{j,i,t}^{0}$ then becomes a matter of finding W^*. Since a set of weights satisfying (10) might not exist, we use the weights that minimize the differences between the weighted averages and the outcome/covariates of county *i* so that (10) come as close to holding as possible.

This approach relies on models (2) and (9), which make two key assumptions. The first assumption is that shale gas development has no effect on the outcome before development of the shale starts. By choosing a conservative T_0 this assumption is plausible. The second assumption is that there are no spillovers from counties on the shale into donor counties. As a robustness check we restrict the donor pool to include counties within PA that are not on the shale and also that do not border any counties on the shale. It is reasonable to think that any spillovers would be concentrated in counties that are geographically proximal to the shale development. For a second robustness check, we use a donor pool comprised of New York counties that also sit on the Marcellus Shale. These counties are similar to the PA counties geologically and economically. The main difference between these New York counties and those in Pennsylvania experiencing the fracking boom is that New York has a state-wide moratorium on fracking. Thus, New York seems to offer a promising set of possible donor counties for generating the synthetic controls.

Our baseline approach is to use the estimated synthetic controls in a framework

[8] These "standard conditions" include: 1) $\sum_{t=1}^{T_0} \upsilon_t' \upsilon_t$ is non-singular, 2) terms $\varepsilon_{j,c,t}$ are independent across time, 3) $\varepsilon_{j,c,t}$ are also mean-independent of $\{Z_c, \lambda_c\}_{c \in S}$ and 4) for some even *p*, the *p*-th moments of $|\varepsilon_{j,c,t}|$ exist for $c \in S$ and $t \in \{1,...,T_0\}$.

similar in spirit to difference-in-difference in order to control for both observed and unobserved confounding variables (including unobservables with time-varying factor loadings). Explicitly, we use OLS to estimate:

$$dY_{j,c,t} - d\tilde{Y}^{0}_{j,c,t} \;=\; \Omega_{2,t}a_2 + \Omega_{3,t}a_3 + \Omega_{4,t}a_4 + (\varepsilon_{j,c,t} - \varepsilon^{0}_{j,c,t}) \qquad (12)$$

It is then a simple exercise to back out the original β parameters. The resulting estimate for β_l can be interpreted as the marginal effect on mean net job flows in industry *j*, county *c*, during period *t*+*l* from a new well spudded during period *t*. For ease of interpretation we will report the cumulative net job flows effect: $c_l = \beta_0 + ... + \beta_l$. Since net job flows are additive, c_l represents the total net job change, or the total employment effect from the time the new well was drilled through *l* quarters later.

4.1.4 Inference

Abadie and Imbens (2008) show that using the traditional bootstrap to perform inference for nearest-neighbor matching estimators with a fixed number of neighbors will produce biased standard errors. The traditional bootstrap requires that the limiting distribution of the statistic being estimated be smooth. Matching estimators have highly non-smooth distributions since they are functions of the distribution of the underlying data. While our use of synthetic controls differs from the nearest-neighbor matching estimator, our estimators do share this property of non-smooth limiting distributions that are functions of the entire distribution of the data. As a precaution, we use the subsampling without replacement variant on the bootstrap from Politis and Romano (1994). This is the approach suggested in Abadie and Imbens (2008) for use with matching estimators. This approach relies on very weak assumptions. Subsampling only requires that the statistic being estimated has a limiting distribution—not that it is smooth.

Subsampling is a very simple and intuitive variant on the bootstrap. The two differences are that the number of observations drawn be less than the size of the original data and that the subsampling is done without replacement. Politis, Romano and Wolf (1999) provide a data-driven approach to choosing the subsample size. In addition to providing valid confidence intervals, the subsampling method allows us to compute a bias-reduced estimator. Following the notation of Politis, Romano and Wolf, the empirical bias of a parameter is:

$$Bias(\hat{\theta}) = \sqrt{\frac{b}{n}}\,(\bar{\hat{\theta}}^{*} - \hat{\theta}) \tag{13}$$

$\bar{\hat{\theta}}^{*}$ is the mean of the subsampled estimates, b is the subsample size and n is the original sample size.[9] In the results, we report only these bias-corrected estimates: $\hat{\theta}_{BC} = \hat{\theta} - Bias(\hat{\theta})$.

4.2 *Means, Medians and other Quantiles*

So far equation (12) is used to estimate the mean marginal effect of fracking on employment. While useful, estimating the mean effect is a coarse approach that may hide important heterogeneous effects of natural gas development across localities. Quantile regression, pioneered by Koenker and Basset (1978), allows us to look beyond mean effects and examine changes in other aspects of the distribution. Allowing marginal effects to vary by quantile may provide evidence about the presence of winners and losers within the distribution: some communities may see growth while others experience contraction. To get at this potential heterogeneity, we use our panel data together with a quantile regression framework similar to an approach developed by Chen and Kahn (2008). The main difference between their approach and our own is that they take first-differences over time, while our first-differences are between actual data and the synthetic control.

[9] See Politis, Romano, and Wolf (1999), section 5.4 (and example 3.4.3).

For reasons pointed out in Koenker and Hallock (2001) and demonstrated by Gamper-Rabindran *et al.* (2010), applying a quantile regression approach directly to equation (12) is not a valid approach. To see this, consider the case where the model described in equations (2) and (9) is linearly heteroskedastic:

$$\begin{aligned} dY_{j,c,t} &= \mathrm{X}_{j,c,t}\mathrm{B} + \upsilon_t \lambda_c + X_{j,c,t}\psi\varepsilon_{j,c,t} \\ dY^0_{j,c,t} &= \mathrm{X}^0_{j,c,t}\mathrm{B} + \upsilon_t \lambda_c + \mathrm{X}^0_{j,c,t}\psi\varepsilon^0_{j,c,t} \end{aligned} \tag{14}$$

Where the notation for covariates and parameters has been consolidated using:

- $X_{j,c,t} = \begin{bmatrix} dW_{c,t} & dW_{c,t-1} & \dots & Z_c \end{bmatrix}$
- $X^0_{j,c,t} = \begin{bmatrix} 0 & 0 & \dots & Z_c \end{bmatrix}$
- $B = \begin{bmatrix} \beta_0 & \beta_1 & \dots & \theta_t \end{bmatrix}^T$

This is a generalization of the original model because it allows the variance of the residual to vary with the observable RHS covariates. Let $q_\theta(.)$ be the conditional quantile function corresponding to the θ-th percentile. Taking differences using equation (14), applying the conditional quantile function, and simplifying, yields:

$$\begin{aligned} q_\theta(dY_{j,c,t} - d\tilde{Y}^0_{j,c,t} \mid X_{j,c,t}, X^0_{j,c,t}) \; = \; & \beta_0 dW_{c,t} + \beta_1 dW_{c,t-1} + \dots \\ & + q_0(X_{j,c,t}\psi\varepsilon_{j,c,t} - X^0_{j,c,t}\psi\varepsilon^0_{j,c,t} \mid X_{j,c,t}, X^0_{j,c,t}) \end{aligned} \tag{15}$$

As pointed out by Koenker and Hallock (2001), quantiles of convolutions of random variables are typically difficult to work with. In general, we cannot interchange difference and quantile operators (Gramper-Rabindran *et al.* 2010) and estimation becomes intractable.

To overcome this, we apply the approach outlined by Chen and Khan (2008). Rather than applying the conditional quantile function to the differenced equation, apply it to (14):

$$q_\theta(dY_{j,c,t} \mid \mathrm{X}_{j,c,t}, \mathrm{X}^0_{j,c,t}) = \mathrm{X}_{j,c,t}\mathrm{B} + \upsilon_t \phi(\mathrm{X}_{j,c,t}, \mathrm{X}^0_{j,c,t}) + X_{j,c,t}\psi\rho_\theta$$
$$q_\theta(dY^0_{j,c,t} \mid \mathrm{X}_{j,c,t}, \mathrm{X}^0_{j,c,t}) = \mathrm{X}^0_{j,c,t}\mathrm{B} + \upsilon_t \phi(\mathrm{X}_{j,c,t}, \mathrm{X}^0_{j,c,t}) + \mathrm{X}^0_{j,c,t}\psi\rho_\theta \qquad (16)$$

We have imposed the substitution $\lambda_c = \phi(X_{j,c,t}, X^0_{j,c,t})$, where $\phi(.)$ is an unknown function of the regressors from both the treated and the control models.[10] Allowing the county effect to be a function of covariates is a generalization of the traditional random effects approach. Differencing these equations now produces:

$$\begin{aligned} q_\theta(dY_{j,c,t} \mid X_{j,c,t}, X^0_{j,c,t}) - q_\theta(dY^0_{j,c,t} \mid X_{j,c,t}, X^0_{j,c,t}) & \\ = \beta_0 dW_{c,t} + \beta_1 dW_{c,t-1} + \ldots + (dW_{c,t} + dW_{c,t-1} + \ldots)\psi\rho_\theta & \qquad (17) \\ = (\beta_0 + \psi\rho_\theta) dW_{c,t} + (\beta_1 + \psi\rho_\theta) dW_{c,t-1} + \ldots & \end{aligned}$$

Thus quantile regression allows the econometrician to exploit heteroskedasticity and allow for marginal effects to vary by quintile.

Since we do not actually have the quantile values required to make estimation of equation (17) feasible, we will adopt the two-stage estimation procedure provided by Chen and Kahn (2008). The first-stage involves estimating (16) to fit predicted quantile values $\hat{q}_\theta(dY_{j,c,t} \mid \mathrm{X}_{j,c,t}, \mathrm{X}^0_{j,c,t})$ and $\hat{q}_\theta(dY^0_{j,c,t} \mid \mathrm{X}_{j,c,t}, \mathrm{X}^0_{j,c,t})$. The second stage then uses the first-difference of these predicted values to estimate (17). This approach is very simple to perform and can be done in STATA and other readily available statistical software packages (Gramper-Rabindran *et al.* 2010).

4.3 The Standard Approach

To highlight the implications and advantages of employing the PIL model, we also estimate a version of the standard model, as described by equation (3). Specifically, we estimate:

---

[10] Ideally, we would estimate $\phi(.)$ nonparametrically. However, due to data limitations we have to impose some functional form assumptions. In the current version, this is specified as a linear function.

$$dY_{j,c,2010-2000} - d\tilde{Y}^0_{j,c,2010-2010} = \gamma dW_{c,2010-2000} + (\varepsilon_{j,c} - \varepsilon^0_{j,c})\,. \tag{18}$$

While we refer to this specification as the "standard approach", in reality it is the standard approach augmented by the use of synthetic controls in the spirit of difference-in-difference.

5. Results

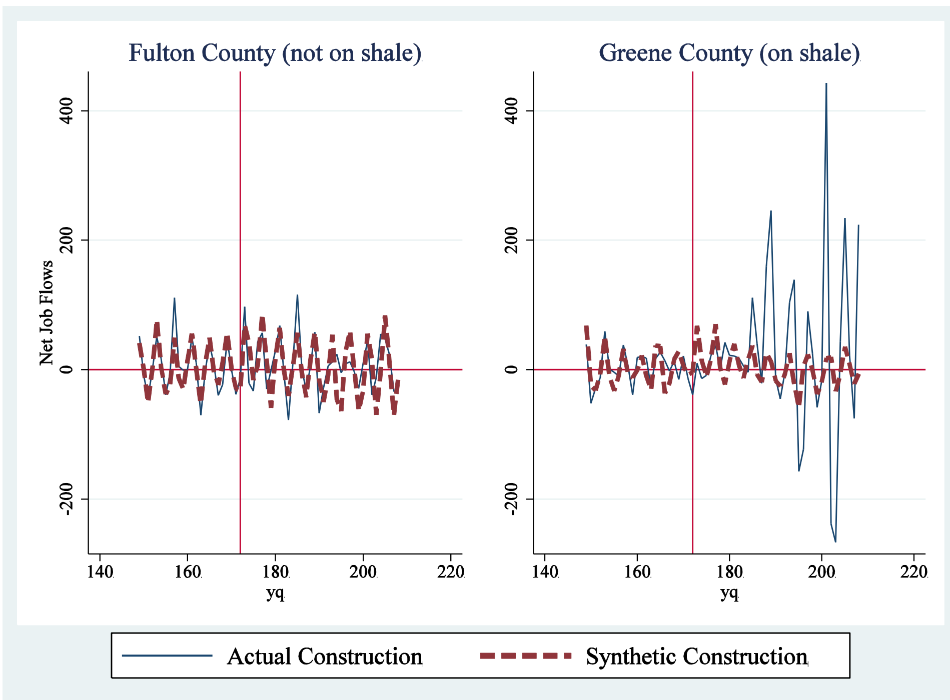
5.1 Comparison Groups

The goal of the synthetic controls is to find the convex combination of donor counties that best approximates the counterfactual for each county that has fracking activity. Before using these synthetic controls in pooled-regression procedures, we need to evaluate the goodness-of-fit for these counterfactual approximations. Of course, since we lack the true data for the counterfactual, we cannot directly test how well the synthetic procedure reproduces the truth. The best we can do is to examine some of the assumptions that the procedure relies on. One necessary condition that the procedure relies on is that the weights, W^*, should come reasonably close to satisfying (10). While this condition is not sufficient for the synthetic controls being unbiased, it is testable. The other condition that must hold is that the data-generating processes described in (2) and (9) must be the true processes—which is not testable.

To assess the fit achieved by the weights in equation (10), we consider two counties: Fulton and Greene. Fulton County is not on the Marcellus Shale and, consequently, is not "treated" by having shale resources. For this reason estimating the synthetic control for Fulton is what ADH refers to as a "placebo". Even after fracking activity begins we expect the synthetic net job flows for Fulton to mimic the true net job flows for Fulton. Greene County is on the shale and experienced significant fracking activity. Equation (10), requires that weights be found that fit the synthetic control to the actual outcome in the pre-fracking period. We follow ADH to produce the synthetic controls for net job flows in the construction industry for Fulton and Greene counties.

Figure 5.1 compares the resulting synthetic control to the true net job flows data for the construction industry in Fulton and Greene counties. While it is difficult to pinpoint exactly when unconventional-drilling techniques made drilling on the Marcellus Shale economically viable, a weak assumption is that it was sometime after 2002. For both counties, the synthetic control and the real data are close throughout the pre-fracking period. This suggests that the condition outlined in (10) comes close to holding and gives credence to the synthetic control as an unbiased estimate of the counterfactual. The synthetic control for Fulton County, the placebo, continues to closely track the real data even throughout the period of time when fracking was becoming popular on the Marcellus Shale. Greene County, on the other hand, experienced a great deal of natural gas development. While the synthetic control closely tracks the true data in the pre-fracking period, it diverges from the true data during the fracking period. These deviations can be interpreted as the effect of fracking on construction net job flows in Greene County.

Figure 5.1



5.2 Mining, Quarrying and Oil and Gas Extraction

In this section, we report results for the mining, quarrying and oil and gas extraction industry. Employment effects in this industry can largely be attributed to the direct effect channel. It does not, however, provide an estimate for the complete direct effect. Many tasks directly involved in drilling and fracking a well can be contracted out to firms that are classified in other industries. For example, millions of gallons of water are required to frack a well. This water is typically brought in by contracted trucking companies that are probably classified to be in the transportation industry. In addition, the effect in this industry should not be considered a pure (if partial) direct effect. There may be jobs counted in this industry that are not directly involved in natural gas but rather were induced by increased local economic activity.

Figure 5.2 shows the estimated employment effect that one new well has on employment in the mining, quarrying and oil and natural gas industry. The y-axis is the estimated cumulative net job flow effect ($\beta_0 + \beta_1 + ... + \beta_t$) from a new well after the number of lags corresponding to the x-axis. The solid black lines represent the point estimates for these cumulative effects from our headline specification using the PIL model. The solid gray area is a 95% confidence interval. The dashed line indicates the estimate using the standard approach described in equation (18). First consider the mean effect panel in the top left of Figure 5.2. The PIL model suggests that fracking activity brings 3 new jobs contemporaneous with drilling a well. The contour shows that most of these jobs disappear immediately after the drilling is complete. Two years out, however, 1.4 new jobs persist. These jobs are likely involved in maintenance, production or administration of the industry. The point estimate from the standard approach is 0.5. While somewhat smaller than the PIL model suggests, the two results are statistically indistinguishable. Both estimates are economically small.

Breaking out the quantile effects shows interesting heterogeneity in the employment effect of fracking activities on jobs in the mining industry. At the tenth percentile, there is still a positive net job flows effect in the period contemporaneous with drilling. However, it is small and after two or three quarters of statistically significant decreases in mining employment the long run effect is close to zero. At the 25th percentile, there is virtually no effect over the first two years. The effects on the median, 75th percentile and 90th percentile show positive effects at quarter 0 (though insignificant in the case of the median). The cumulative net jobs flowing from fracking activity increases in the case of the median and the 75th percentile and remains level in the case of the 90th percentile as we go through the first two years after the new well was drilled. The upper quantiles of 75 and 90 percent also experience much more employment growth. Two years out from spudding, each well at the 75th percentile results in 4.7 new jobs while each well at the 90th percentile results in 5.8 new jobs. The quantile results also illustrate two distinct profiles for the effect of a new well on jobs. First is one where there is an increase in employment simultaneous with drilling but this increase is short lived. Recall from background discussion that the unconventional drilling and fracking processes require specialized skill sets that were initially only help by a small number of out-of-state firms. Consequently, this estimate for jobs contemporaneous to drilling may be biased-downward as they are being filled with transient out-of-state workers who may not appear in our data (although, from the point of view of impacts on local employment, these jobs are not relevant). The second profile is one that has an increase simultaneous with drilling and this increase persists and even increases in the long run. These long-run jobs should be properly counted while the short run jobs may be undercounted. It is possible that these production jobs accrue in the upper part of the distribution.

Figure 5.2

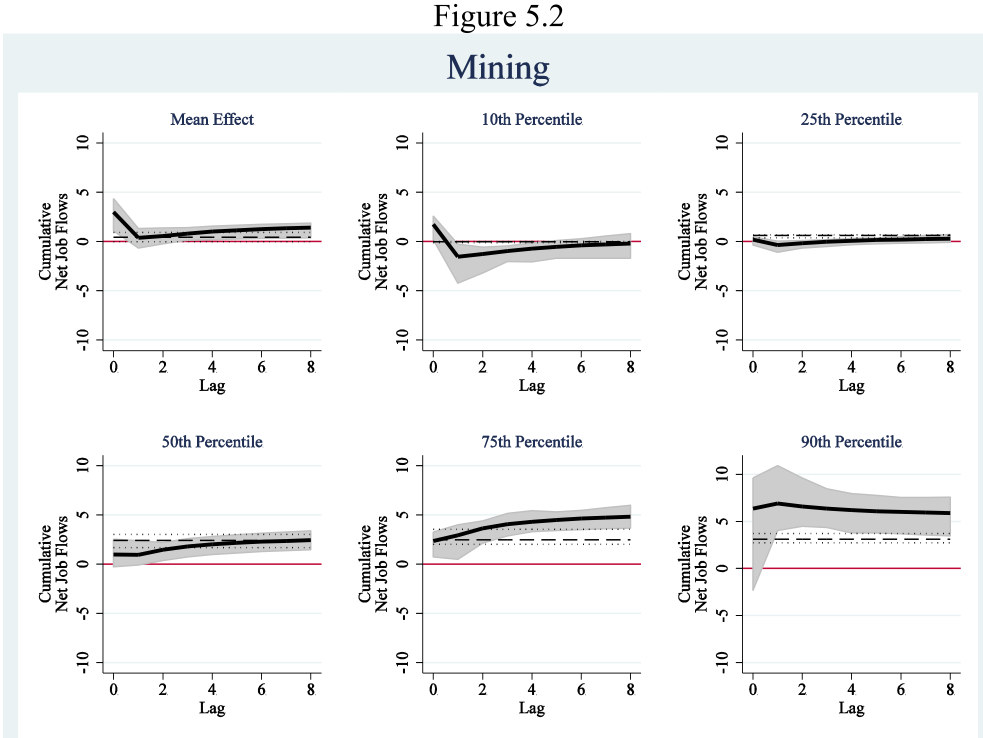
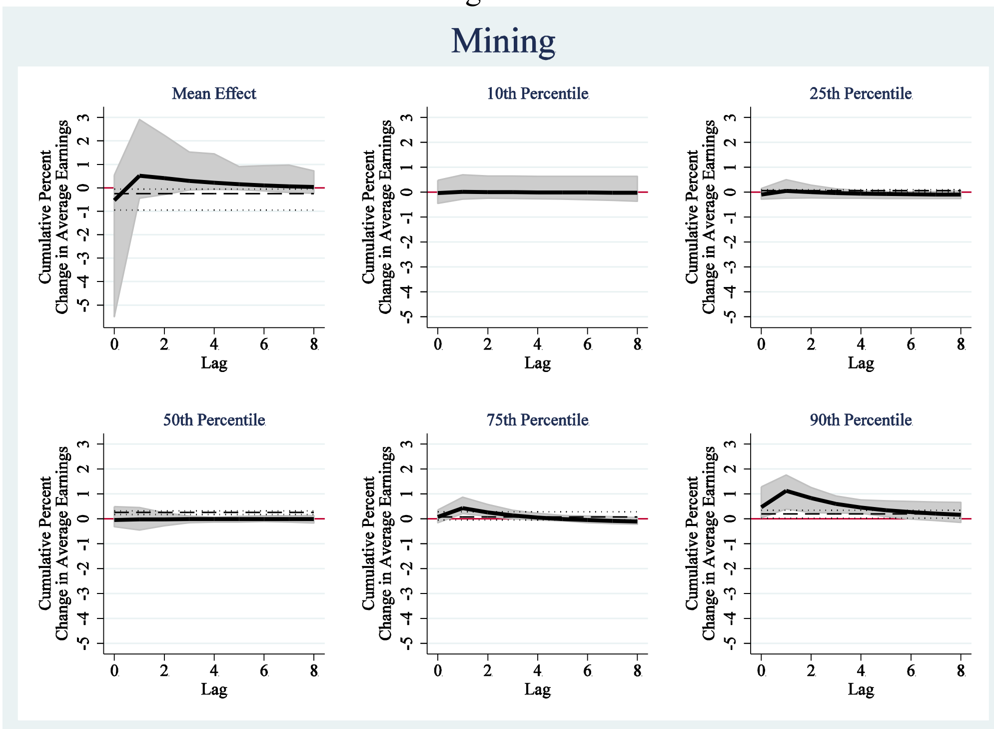


Figure 5.3



Note: For both figures 5.2 and 5.3 the solid black line indicates the cumulative net job flow estimates found using the PIL model (12) and the gray area represents the corresponding 95% bootstrapped confidence interval. The dashed line represents the estimate from the simple first-difference model and the dotted lines are the corresponding 95% confidence interval.

Looking at Figure 5.3, there is no evidence of a positive earnings effect on the mean or the lower half of the distribution. However, there are positive earnings effects on the 75th and 90th percentiles. At the peak, these effects are .42% and 1.12% per well. While we cannot identify the factor driving this heterogeneity, it is instructive to consider that most of these counties have multiple new wells installed each quarter indicating potentially large earnings effects, at least in the short run. Over the long run as jobs transition from temporary drilling jobs to more permanent positions these earnings effects return to zero. The increase in earnings suggests tightness in this portion of the labor market. Over time, workers will adjust and alleviate this tightness. Thus the earnings effect disappears.

For both net job flows and earnings the estimates from the standard approach are similar to those of the PIL model at most quantiles. The effect on the 75th percentile and the 90th percentile for net job flows are two noteworthy exceptions. For both of these, the sets of results are similar in the first few periods but diverge as the number of lagged quarters increases. While the PIL allows the researcher to see how net job flows vary after a new well is drilled, the standard approach allows only a weighted average of these lagged effects, the weights of which are determined by the timing of each new well spudding in relation to the end of the time-horizon considered by the researcher.

5.3 Manufacturing—The Traded-Goods Sector

The manufacturing industry is the best representation of a traded-goods sector. Demand for these traded-goods will be relatively unaffected by local demand shocks associated with fracking activity. If local economies are developing Dutch disease, this is the industry where we expect to see it. We test for two necessary conditions for Dutch disease. The first condition is that employment in the non-traded industry contracts. The second condition is that production costs rise. In this case we will use labor costs or earnings to proxy for production costs.

Simply looking at the mean effect suggests no evidence of Dutch disease. The top left panel of Figure 5.4 shows that throughout the entire time horizon the effect of a new well on manufacturing employment is very close to zero and insignificant. The top left panel of Figure 5.5 also shows that there is no significant effect on earnings. However, the quantile results tell a rather interesting story that is consistent with Dutch disease, at least in a part of the distribution. Looking at the middle of the distribution—the 25th percentile, the median and the 75th percentile—suggests no effect on manufacturing employment. However, in the tails we find large and significant effects. The top distribution (the 90th percentile) shows a temporary decrease in manufacturing employment immediately after fracking. This decrease is followed by a large and persistent increase in employment. In this quantile we estimate an increase in manufacturing employment of 3.9 jobs per well two years after drilling. Looking at the lower tail of the distribution we find potential evidence of Dutch disease. At the 10th percentile the effect of a well on manufacturing employment concurrent with spudding is zero. It is intuitive that the types of tasks needed to drill and spud a new well are different from the tasks performed by manufacturing firms. Since manufacturing probably doesn't compete for workers directly with drilling we are not surprised by a negligible contemporaneous effect. However, moving further out in the time horizon (allowing for more complete adjustment of labor supply and demand curves) a large and negative effect appears. Two years after a new spudding we estimate a loss of 5.7 manufacturing jobs—satisfying the first necessary condition for Dutch disease.

The second necessary condition for Dutch disease was that average earnings in manufacturing should have an increase associated with fracking activity. This is indeed what we find at the 75th and 90th percentile, as pictured in Figure 5.5—small but significant positive earnings effects for manufacturing workers. It is important to note that the distribution designated by the quantile approach is based on the model residual. Since the employment and earnings effects are estimated separately we cannot make any claims that they occur in the same or different counties. The only claim we make these findings are consistent with the Dutch-disease story.

Figure 5.4

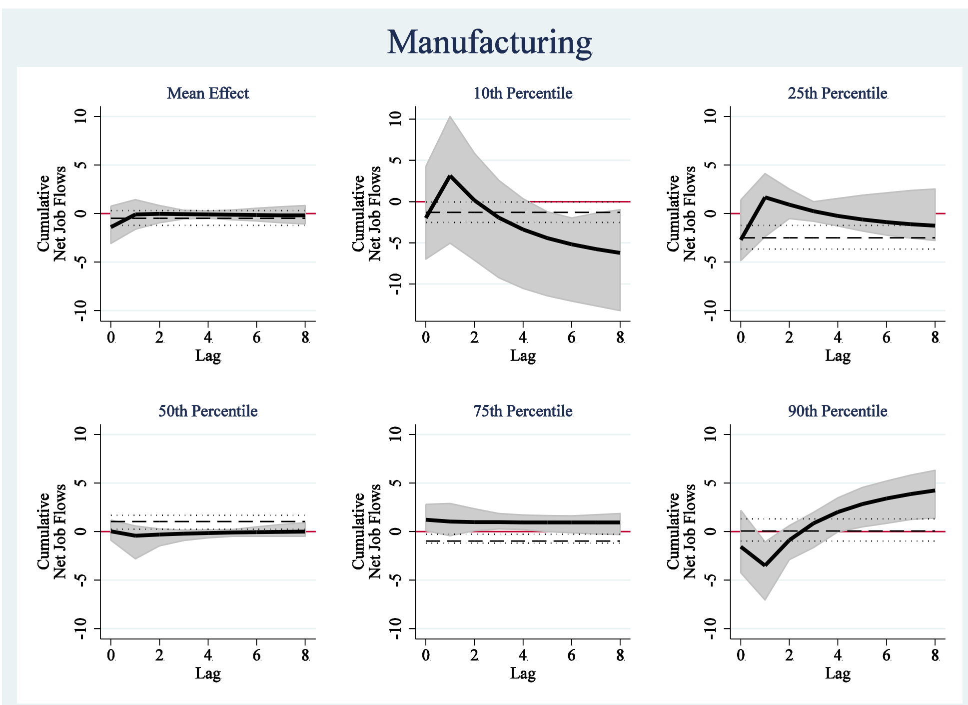
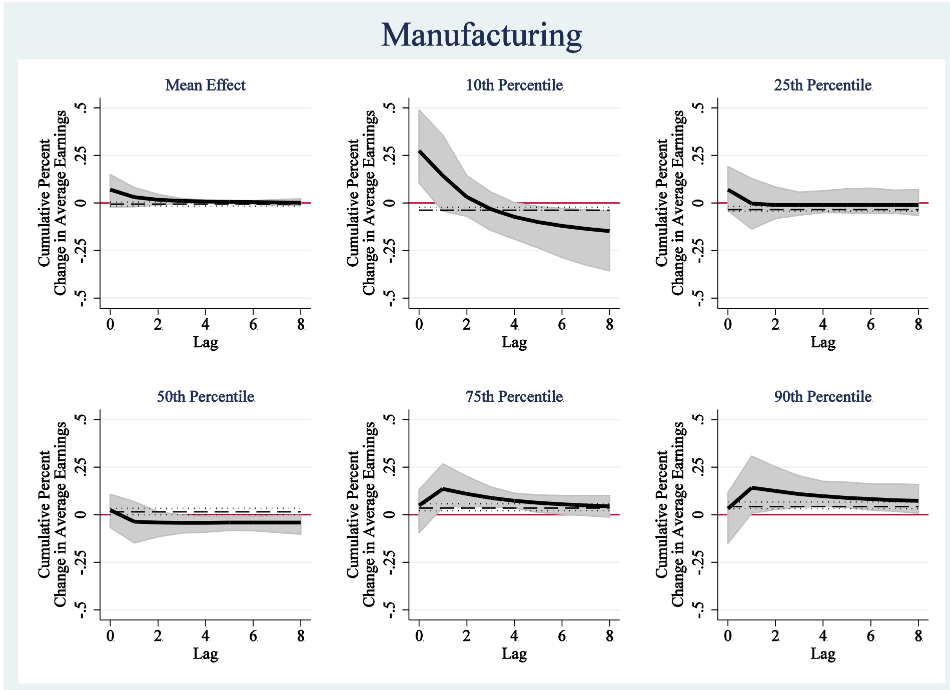


Figure 5.5



Note: For both figures 5.2 and 5.3 the solid black line indicates the cumulative net job flow estimates found using the PIL model (12) and the gray area represents the corresponding 95% bootstrapped confidence interval. The dashed line represents the estimate from the simple first-difference model and the dotted lines are the corresponding 95% confidence interval.

5.4 Total Employment and Other Industries

Due to the large number of results, we leave the remaining industry figures in the Appendix and highlight some remarkable findings from them here. The mean effect for total employment suggests that each additional well is associated with 4.2 additional and persisting jobs. We gain little from the quantile results for total employment as they are statistically insignificant across the board. In addition, we estimate only a very small average earnings effect at the mean that becomes significant only after a one-year lag. Similar to the net job flows results, the quantile results suggest mostly small and insignificant estimates for total average earnings. Once again, gains to employment with tiny earnings effects are suggestive of labor market slack.

The effect of an additional well on employment in the Transportation industry is positive along much of the distribution, but small. Each new well accounts for 0.5 new jobs. This is some upward pressure on earnings in this industry. The Accommodation and Food Service industry shows little response to a new well, for the most part. At the 90th percentile, however, there is a large and delayed negative effect on net job flows. Two years after spudding, the Accommodation and Food Service industry in this quantile will have shed 5.5 jobs per well. One possible story for this large contraction is it is being driven by the loss in total employment seen at the 10th percentile (this loss is sizable but insignificant). Estimates suggest only minimal earnings effects to any part of this distribution.

6. Discussion

6.1 Comparing the PIL Estimates with the Standard Estimates

Looking across the sets of results for each industry and quantile, a common relationship between the PIL estimates and the standard estimates is apparent. Consider, for example the results for the Mining Industry in Figures 5.2 and 5.3. In some quantiles we find divergence between the estimated employment effects of the PIL model and the standard model. The most striking examples are the marginal effects at the 75th and 90th percentile in Figure 5.2. In both of these cases the marginal effects from the PIL and standard approaches are indistinguishable in the first two quarters. After that, however, in the case of the 75th percentile the PIL model shows employment increasing over the next six quarters while the standard estimate remains constant (by construction). In the case of the 90th percentile, the PIL point estimate is consistently higher than the standard estimate. However, over first few lags the PIL estimates have wide confidence intervals that preclude the conclusion that the PIL and standard estimates are statistically differentiated in the first two quarters.

Looking across the sets of results for each industry and quantile, in many instances we see a similar relationship. The PIL and standard estimates are similar or statistically indistinguishable in the early part of the time-horizon but drift apart over time. This suggests that the standard estimate weighs the short-run effects of each new well more than the long-run effects. This is the case when the wells are concentrated toward the end of the time horizon defining the first-difference. In this case that means the wells are concentrated toward the 2010 portion of the horizon 2000 to 2010 (see Figure 2.1A).

From the policy perspective this is a very important detail. Policymakers considering allowing or encouraging fracking activity as a way to grow the local economy and add new jobs should be concerned with the long-run job gains. Providing temporary work for transient workers is probably not as desirable as creating long-term stable employment. This is an important nuance missed by the standard approach.

6.2 Advantages of Quantile Regressions

Using the quantile approach allows the estimation of the heterogeneous effects of a new well on various parts of the distribution. This is instrumental to our analysis in two ways. First, it is instructive to see that the employment effects of shale development are non-uniform in both the timing and magnitude of net job flows. Manufacturing is salient example of this non-uniformity. In the lower tail we see contraction with fracking activity while in the upper tail we see expansion. Both results are interesting, relevant to policymakers and missed completely in mean regression. It appears that, while Dutch disease occurs, it is not wide-spread. For this reason, without the use of quantiles it would be ruled-out completely. One limitation is that the quantiles are based on the model residual and consequently we cannot say anything about factors driving the heterogeneity or where Dutch disease is likely to strike. Doing so would require modeling decisions about which particular variables to use in interactive terms.

The second contribution of the quantile regression is that it facilitates comparisons between what we estimate using the PIL model and what we estimate using the standard model. For the industries considered, if we were confined to mean effect estimates, we could not conclude that the estimates of the PIL model differed from the standard model for any lag even though the PIL point estimate changes over time and the standard estimates do not. However, allowing for heterogeneous effects uncovers instances where the two differ in ways that confirmed our intuition concerning the difference between the PIL model and the standard model.

6.3 Labor Market Slack

For the most part, the earnings response to fracking activity is small relative to the employment response. This suggests the existence of slack in the labor market. Figure 6.1 plots the mean county, median county, 25th percentile county and the 75th percentile county for a number of indicators of labor market slackness. The top left panel shows that there was an increase in the unemployment rate before most of the fracking activity and this increase persisted through 2013q1. There are two likely sources of this increase in unemployment. First, is the recession. Second, out-of-staters could be moving into Pennsylvania in hopes of finding employment in the fracking boom. The lower left panel of Figure 6.1 plots the size of the labor force. It suggests that increases in the size of the labor force have been small and gradual throughout this period with no obvious acceleration after fracking began.

Figure 6.2 compares the growth in the labor force in Pennsylvania counties situated on the Marcellus Shale with counties off of the shale. The right panel compares the trends and levels of labor force size. It shows small changes in labor for size after 2008 both on and off the shale. This suggests that much of the labor market slackness can be attributed to the recession and slow recovery.

Figure 6.1

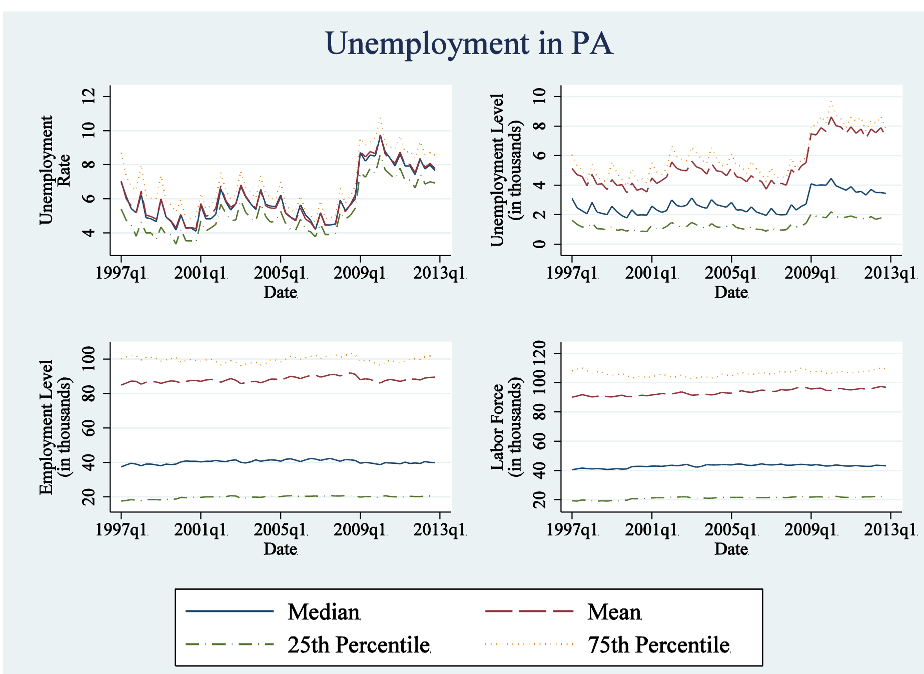
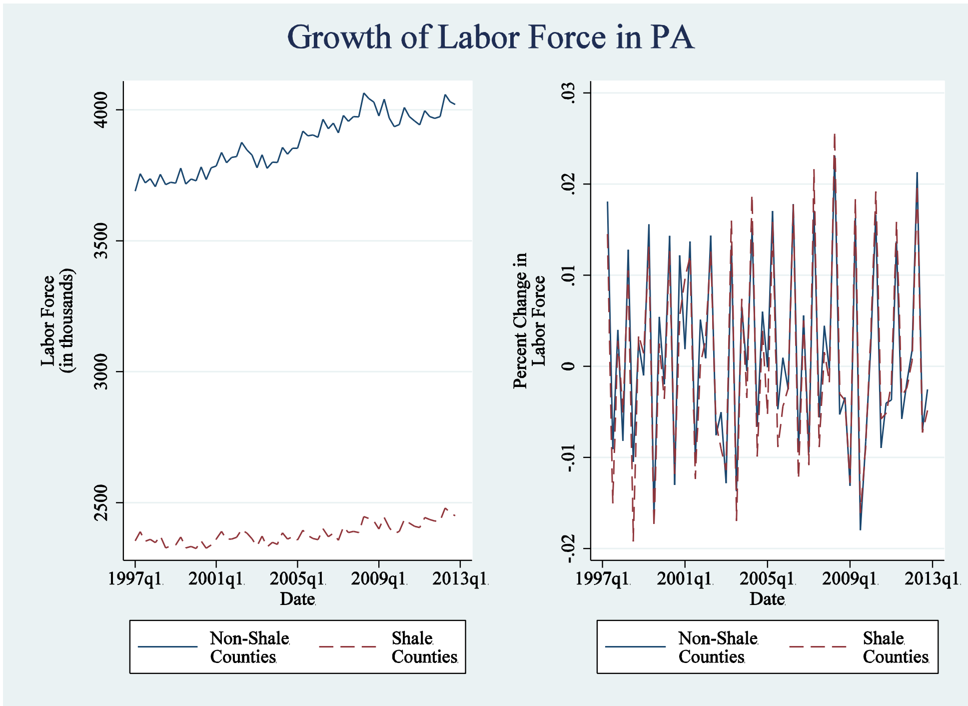


Figure 6.2

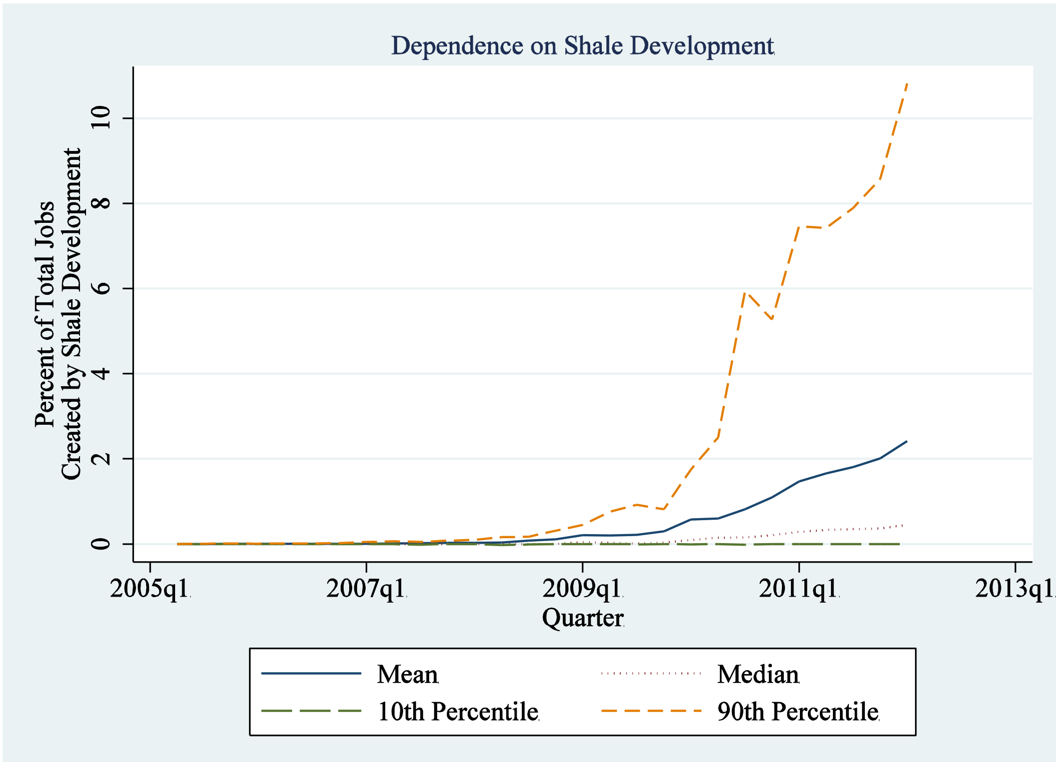


Source: Figures 6.1 and 6.2 from Local Area Unemployment Statistics, BLS.

6.4 Economic Dependence on Shale Resources

Finally, we consider how the portion of jobs in local economies that depend on fracking activity is changing over time. Using the estimated total employment effects per well to calculate the number of jobs in local economies supported by the fracking boom, the figure below describes how the distribution of natural gas-dependent jobs as a fraction of total jobs changes over time. The figure shows a sizeable divergence in the percent of total jobs dependent on shale development. Throughout the period examined, many of the counties show very little dependence. However, for a few counties the shale jobs are quickly becoming very important. Among these are the counties of Bradford, Greene, Sullivan, Susquehanna and Tioga, where by 2012 more than 10 percent of total jobs created were the result of fracking. There is a downside to this growth. While these counties may enjoy the benefits of employment growth, they may be becoming increasingly vulnerable to fluctuations in global natural gas prices.

Figure 6.3

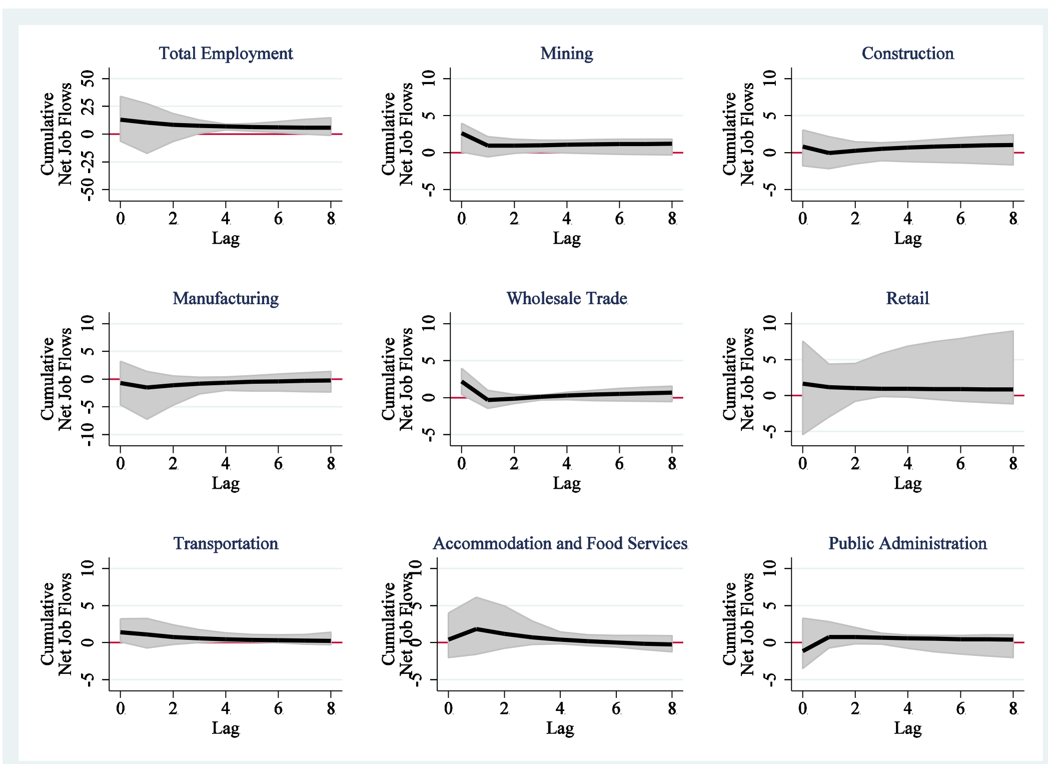


7. Robustness Check

In this section we perform some robustness checks to test the sensitivity to the choice of a donor pool. As discussed above, the synthetic control approach relies on the assumptions that there are no geographic spillovers from fracking counties to donor counties. To see if the results are being driven by potential spillovers of this type we repeat the mean procedure using two alternative donor pools: one omitting counties directly bordering fracking counties and one using New York counties to comprise the donor pool.

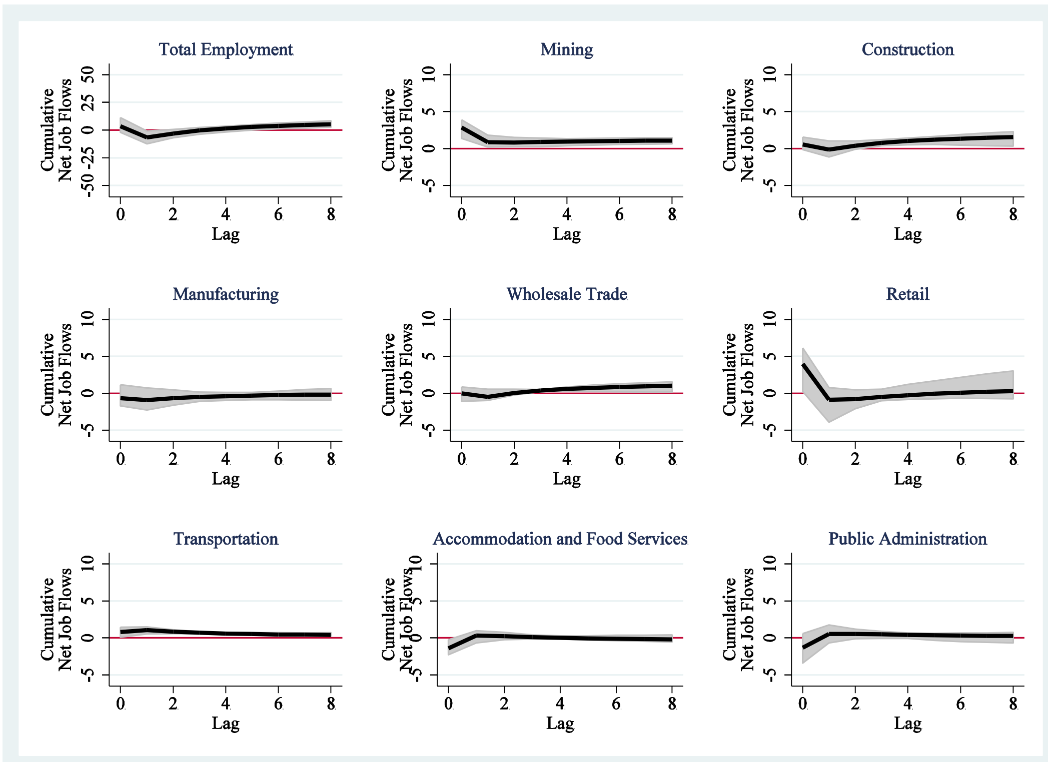
First we restrict the donor pool to include counties within Pennsylvania that do not have any fracking activity and do not share a border with a county that has fracking activity. Any spillover effects from fracking counties into non-fracking counties will likely be from those that are geographically proximal to the fracking activity. Thus by eliminating these donor counties, we can be more confident that the no-spillover assumption is not violated. Figure 7.1 contains the estimated marginal effects on the mean cumulative net job flows. Using the restricted donor pool the mean effect on each of the individual industries is very similar to the results using the unrestricted donor pool. The one possible exception is in total employment. Here the restricted donor pool yields a larger short-run estimate (13 jobs at the time of drilling). However, the wide confidence intervals make this result indistinguishable from our main results. Also, in the long run the restricted and unrestricted results produce similar results.

Figure 7.1



Second, we restrict the donor pool to include counties in the state of New York that are on the Marcellus Shale. These counties likely make good comparisons with the Pennsylvania shale counties as they have are similar in economic, demographic and geographic dimensions. The main difference is a state-wide moratorium on fracking in New York. For individual industries, the results reported in Figure 7.2 confirm the baseline results. Point estimates for total employment are, once again, different in the short-run but statistically indistinguishable.

Figure 7.2



8. Conclusion

Recent technological innovations have produced a boom of unconventional, hydraulically fracked natural gas wells across much of Pennsylvania. We find that this boom has increased local employment by a small, but statistically meaningful, amount. We also find that earnings in most industries are unresponsive to fracking activity. One exception to this is the mining, quarrying and oil and natural gas industry, which exhibits a small but statistically significant increase in earnings in the upper quantiles. By employing a PIL model we identify time-varying employment effects that are missed by the more coarse standard approach. Using this model together with quantile regression to allow for heterogeneous effects, we find evidence of Dutch disease as parts of the distribution experience contraction in manufacturing employment and increases in average earnings.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*. 105(490):493-505.

Abadie, A., and W. Imbens (2006). "Notes and Comments On the Failure of the Bootstrap for Matching Estimators." *Econometrica*. 76(6): 1537-1557.

Acemoglu D. and F. Zilibotti (1997). "Was Prometheus Unbound by Chance? Risk, Diversification, and Growth." *The Journal of Political Economy*. 105(4): 709-751.

Allcott, H. and D. Keniston (2013). "Dutch disease or Agglomeration? The Local Economic Effects of natural Resource Booms in Modern America." Working paper.

Arellano, M. (2003). *Panel Data Econometrics*. Oxford University Press.

Bartik, T. (1991). "Who Benefits from State and Local Economic Development Policies?" Upjohn Institute for Employment Research.

Black, D., T. McKinnish and Seth Sanders (2005). "The Economic Impact of the Coal Boom and Bust." *The Economic Journal*. 115(503): 449-476.

Brundage, T., J. Jacquet, T. Kelsey, J. Ladlee, J. Lobdell, J. Lorson, L. Michael, and T. Murphy. "Pennsylvania Statewide Marcellus Shale Workforce Needs Assessment." Marcellus Shale Education & Traning Center (MSETC). June 2011.

Chen S. and S. Khan (2008). "Semiparametric Estimation of Nonstationary Censored Panel Data Models with Time Varying Factor Loads." *Econometric Theory*. 24:1149-1173.

Considine, T., R. Watson, and S. Blumsack. "The Pennsylvania Marcellus Natural Gas Industry: Status, Economic Impacts and Future Potential." The Marcellus Shale Coalition. July 2011.

Corden W., and J. Neary (1982). "Booming Sector and De-Industrialisation in a Small Open Economy." *The Economic Journal*. 92:825-848.

Fetzer, T. (2014). "Fracking Growth." Mimeo.

Gamper-Rabindran, S., S. Khan, and C. Timmins (2010). "The Impact of Piped Water Provision on Infant Mortality in Brazil: A Quantile Panel Data Approach." *Journal of Development Economics*. 92:188-200

Henderson, J. (2003). "Marshall's scale economies." *Journal of Urban Economics*. 53:1-28.

Imbs, J. and R. Wacziarg (2003). "Stages of Diversification." *American Economic Review*. 93(1): 63-86.

Jacobs, J. (1969). The Economy of Cities. Random House. 1969.

Kelsey, T., M. Shields, J. Ladlee, and M. Ward. "Economic Imipacts of Marcellus Shale in Pennsylvania: Employment and Income in 2009." Marcellus Shale Education & Traning Center (MSETC). August 2011.

Koenker, R. and G. Bassett (1978). "Regression Quantiles." *Econometrica*. 46(1):33-50.

Koenker, R. and K. Hallock (2001). "Quantile Regression." *Journal of Economic Perspectives*. 15(4):143-156.

Koren, M. and S. Tenreyro (2007). "Volatility and Development." *The Quarterly Journal of Economics*. 122(1): 243-287.

Maniloff and Mastromonaco (2014). "The Local Economic Impacts of Unconventional Shale Development." Working paper.

Manresa, E. (2013). "Estimating the Structure of Social Interactions Using Panel Data." Unpublished Manuscript. CEMFI, Madrid.

Marchand, J. (2012). "Local Labor Market Impacts of Energy Boom-Bust-Boom in Western Canada." *Journal of Urban Economics*. 71:165-174.

Meinshausen N. and B. Yu. (2009). "Lasso-type recovery of sparse representations for high- dimensional data." *Annals of Statistics*. 37(1): 2246-2270.

Moretti, E. (2010). "Local Multipliers." *American Economic Review: Papers and Proceedings*. 100:373-377.

Muehlenbachs, L., E. Spiller, and C. Timmins (2014). "The Housing Market Impacts of Shale Gas Development." NBER Working Paper 19796. Cambridge, MA.

Murkerjee, S. (2010). "Presentation to the Economic & Workforce Development Workgroup." Prepared for the Governor's Marcellus Shale Advisory Commission. April 2010.

"Pennsylvania Marcellus Shale Workforce Needs Assessment." Marcellus Shale Education & Training Center. Summer 2011.

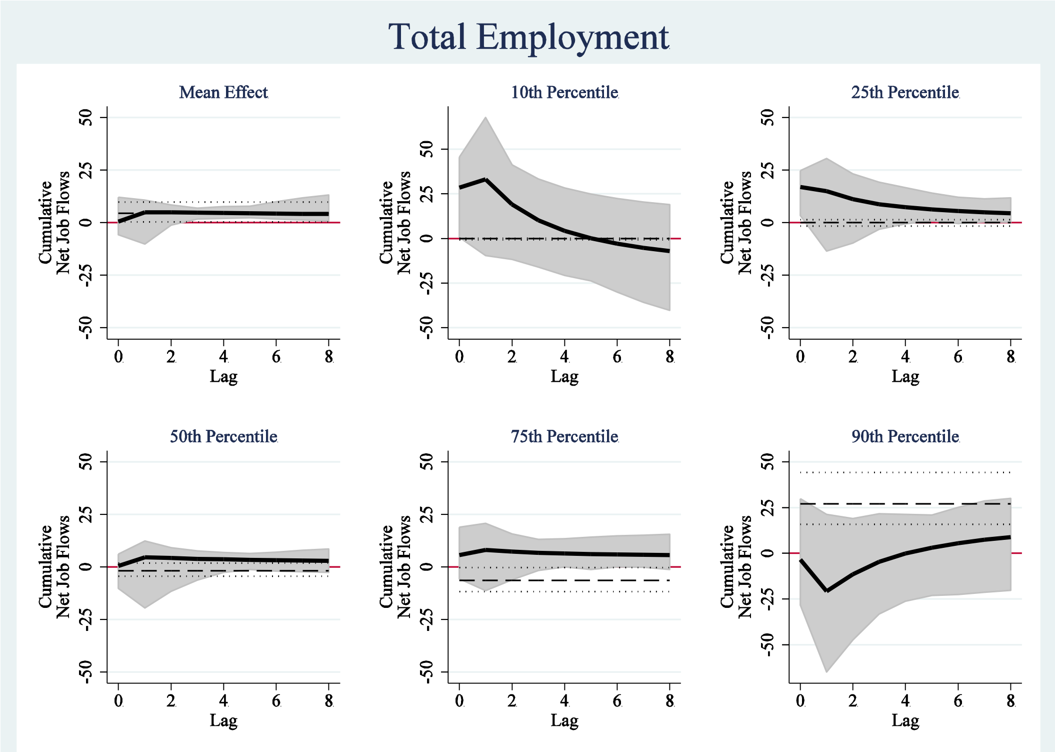
Politis, N. and J. Romano (1994). "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions." *The Annals of Statistics*. 22:2031-2050.

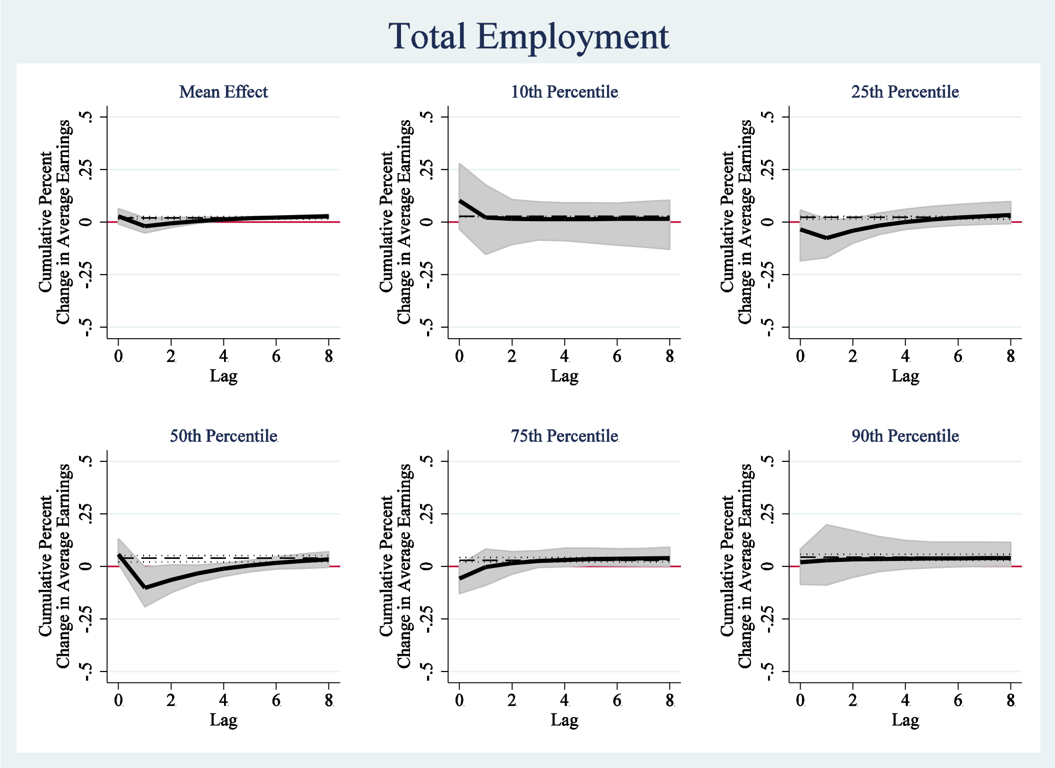
Politis, N., J. Romano, and M. Wolf (1999). Subsampling. New York: Springer.

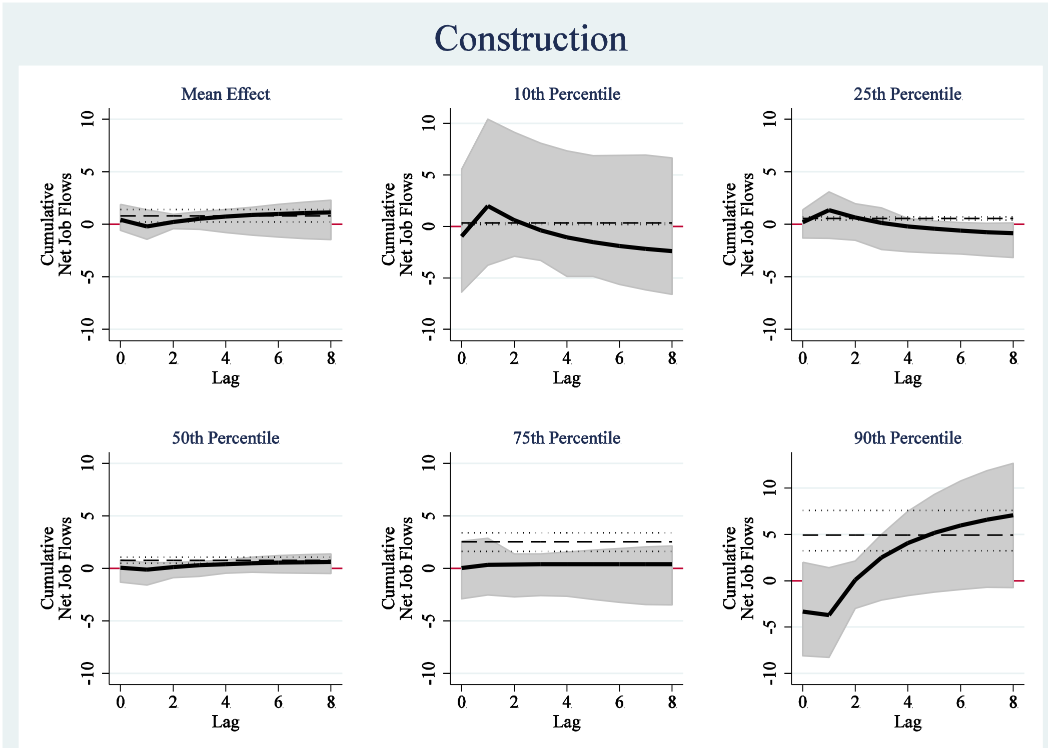
Weber, J. (2013). "A Decade of Natural Gas Development: The Makings of a Resource Curse?" *Resource and Energy Economics*. 37:168-183.

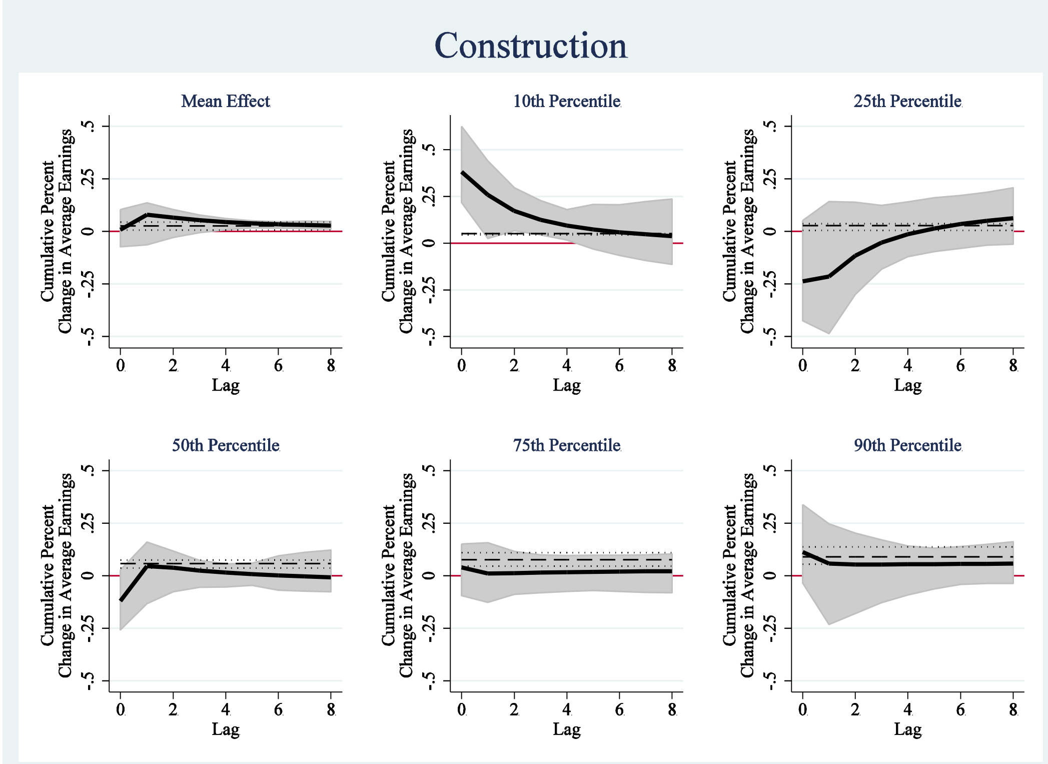
Appendix A

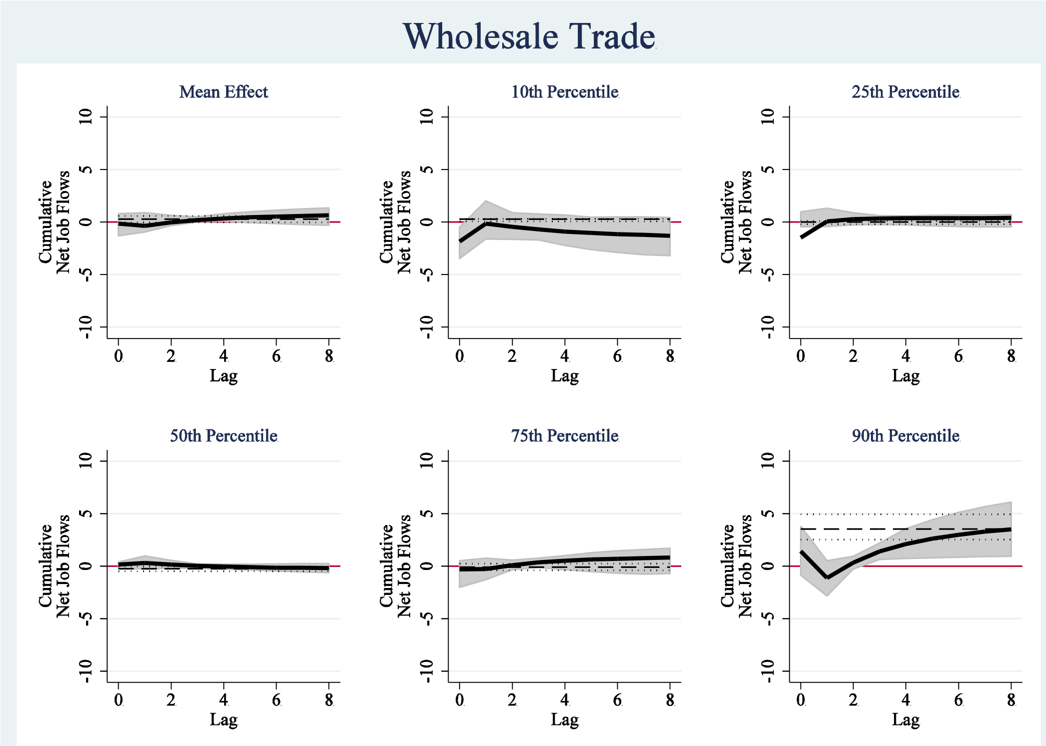
Mean and Quantile Results for Total Employment and Other Industries

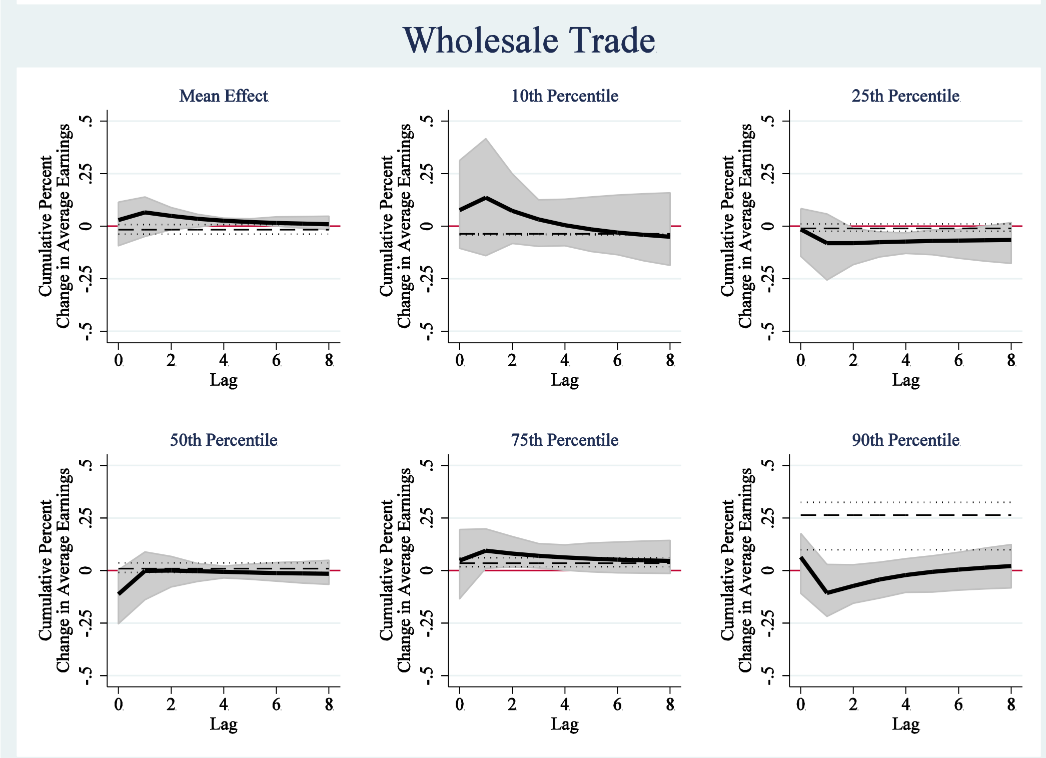


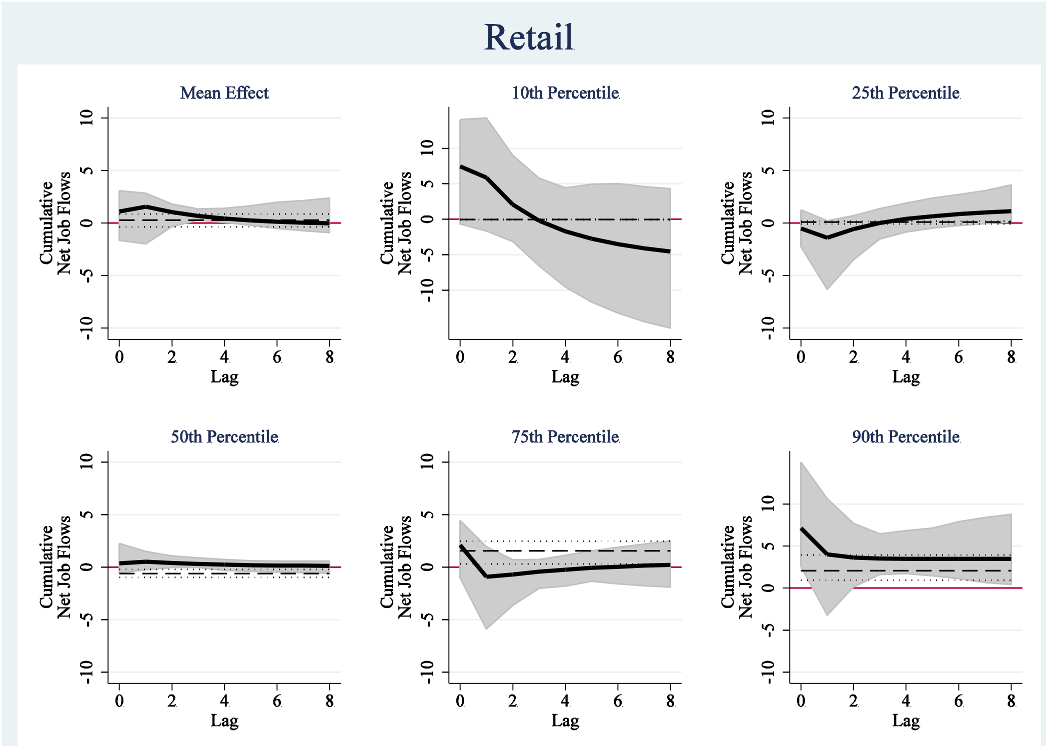



Construction
Mean Effect
10th Percentile
25th Percentile
50th Percentile
75th Percentile
90th Percentile
Cumulative Net Job Flows
Lag

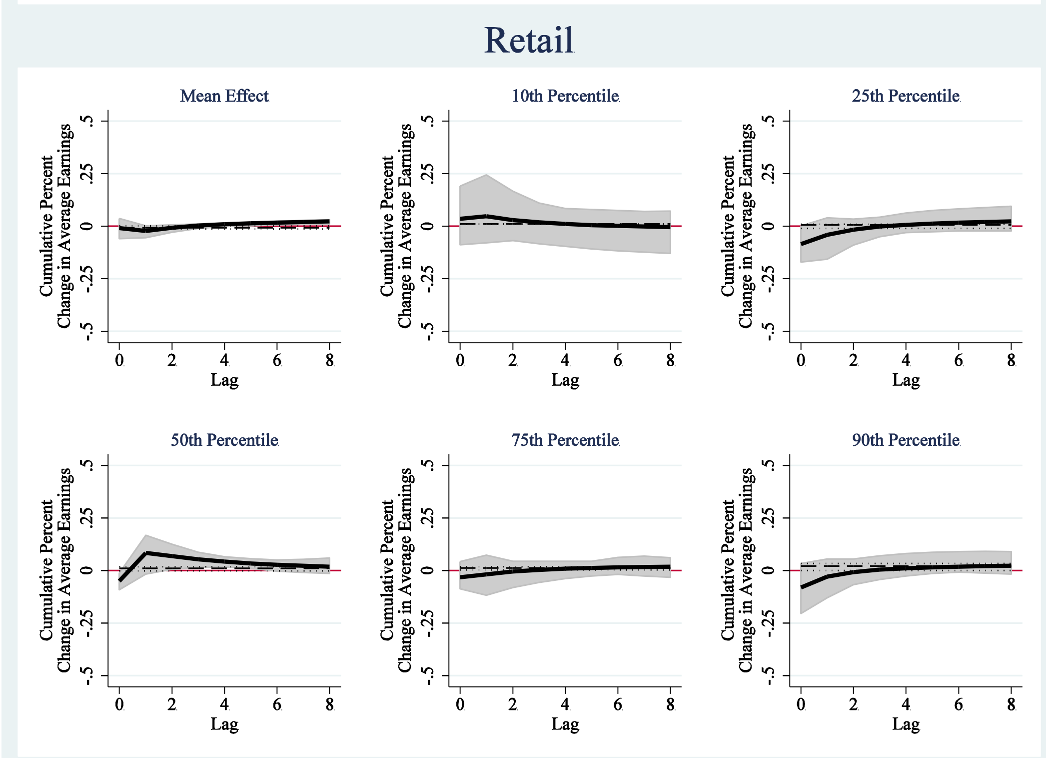

Construction
Mean Effect
10th Percentile
25th Percentile
50th Percentile
75th Percentile
90th Percentile
Cumulative Percent Change in Average Earnings
Lag


Wholesale Trade
Mean Effect
10th Percentile
25th Percentile
50th Percentile
75th Percentile
90th Percentile
Cumulative
Net Job Flows
-10
-5
0
5
10
0
2
4
6
8
Lag

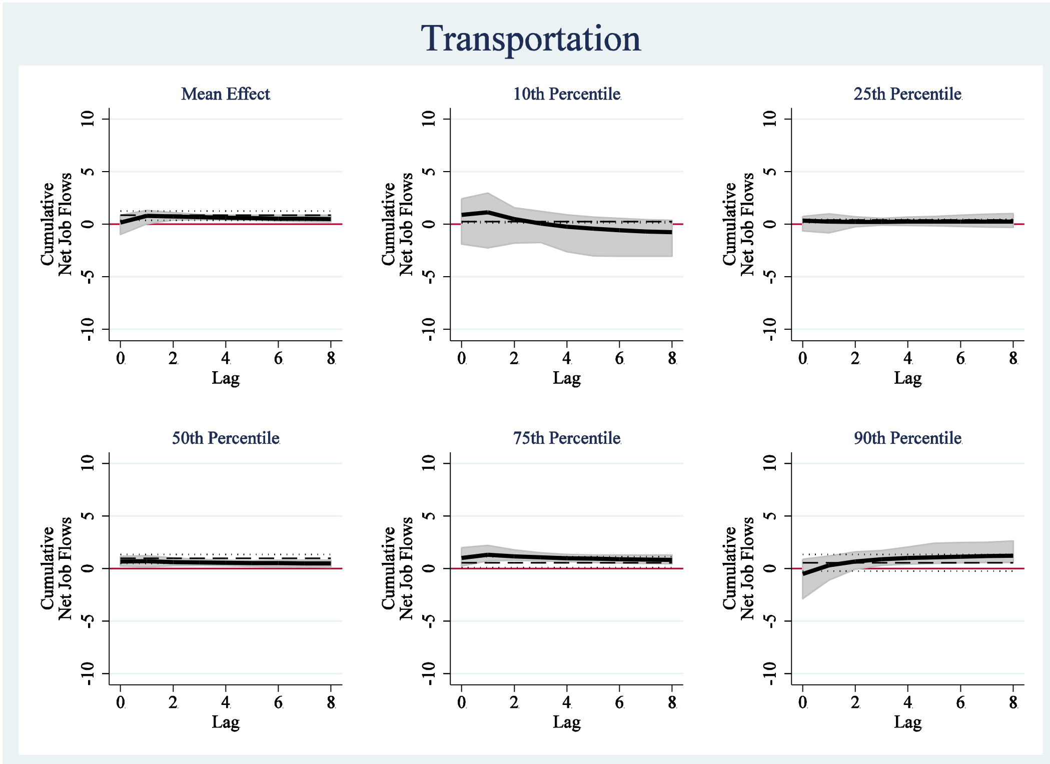

Wholesale Trade
Mean Effect
10th Percentile
25th Percentile
50th Percentile
75th Percentile
90th Percentile
Cumulative Percent
Change in Average Earnings
-.5
-.25
0
.25
.5
0
2
4
6
8
Lag

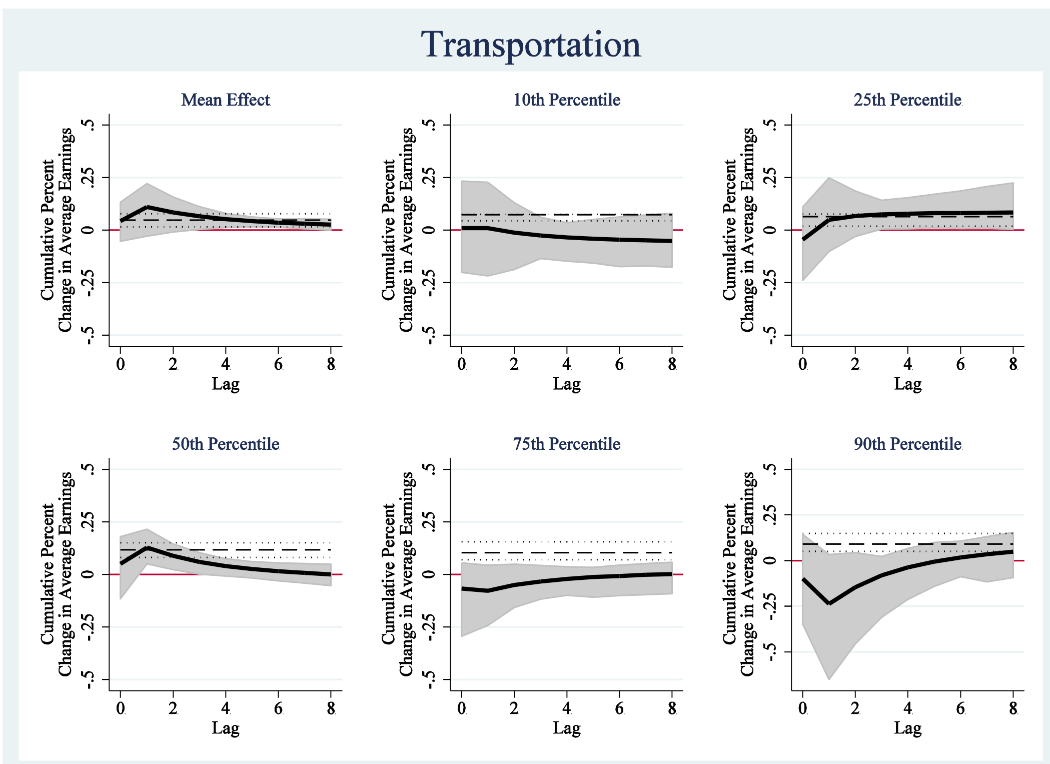


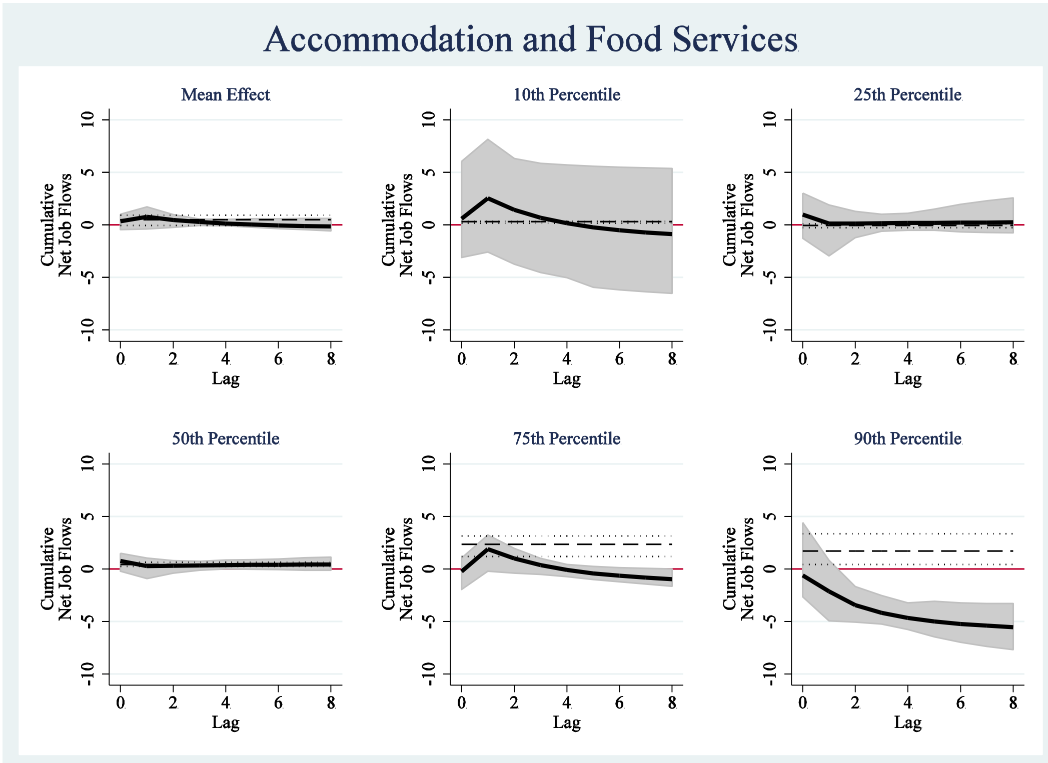
Retail
Mean Effect
10th Percentile
25th Percentile
50th Percentile
75th Percentile
90th Percentile
Cumulative
Net Job Flows
Lag

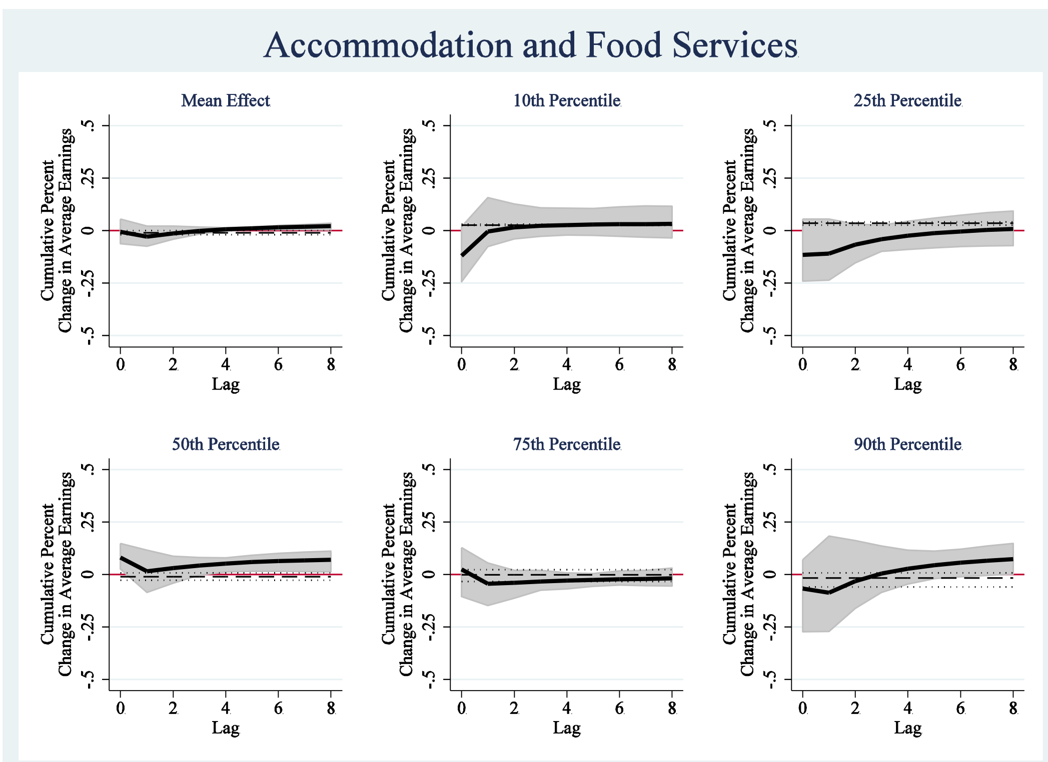


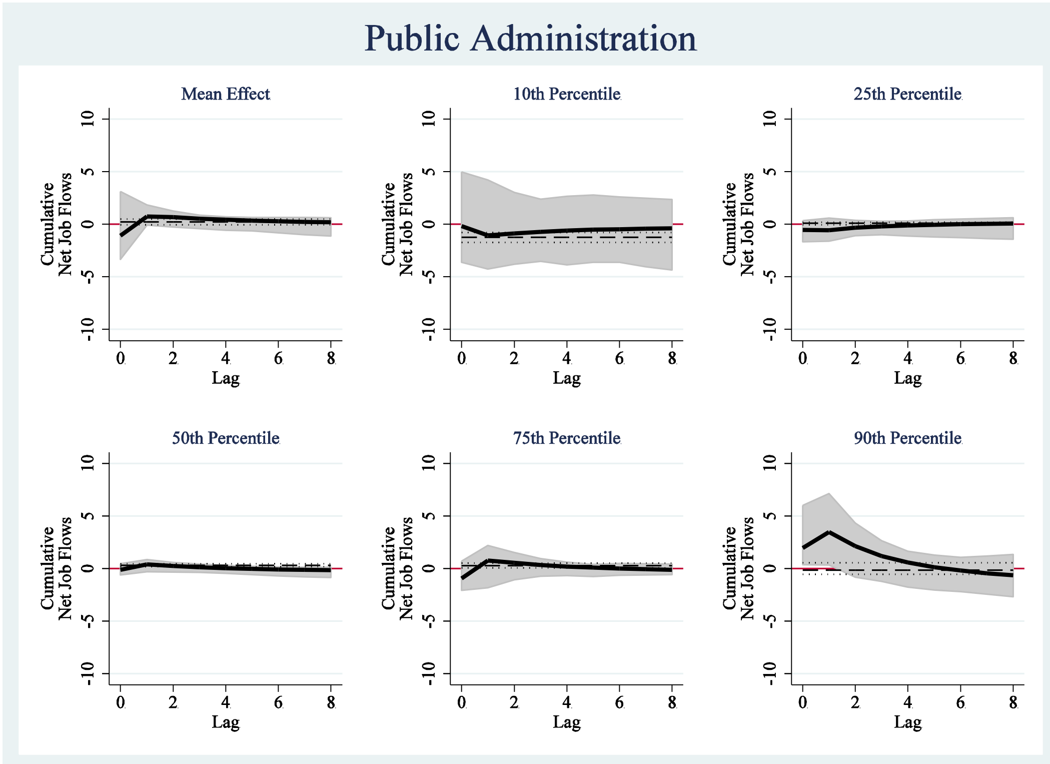
Retail
Mean Effect
10th Percentile
25th Percentile
50th Percentile
75th Percentile
90th Percentile
Cumulative Percent
Change in Average Earnings
Lag

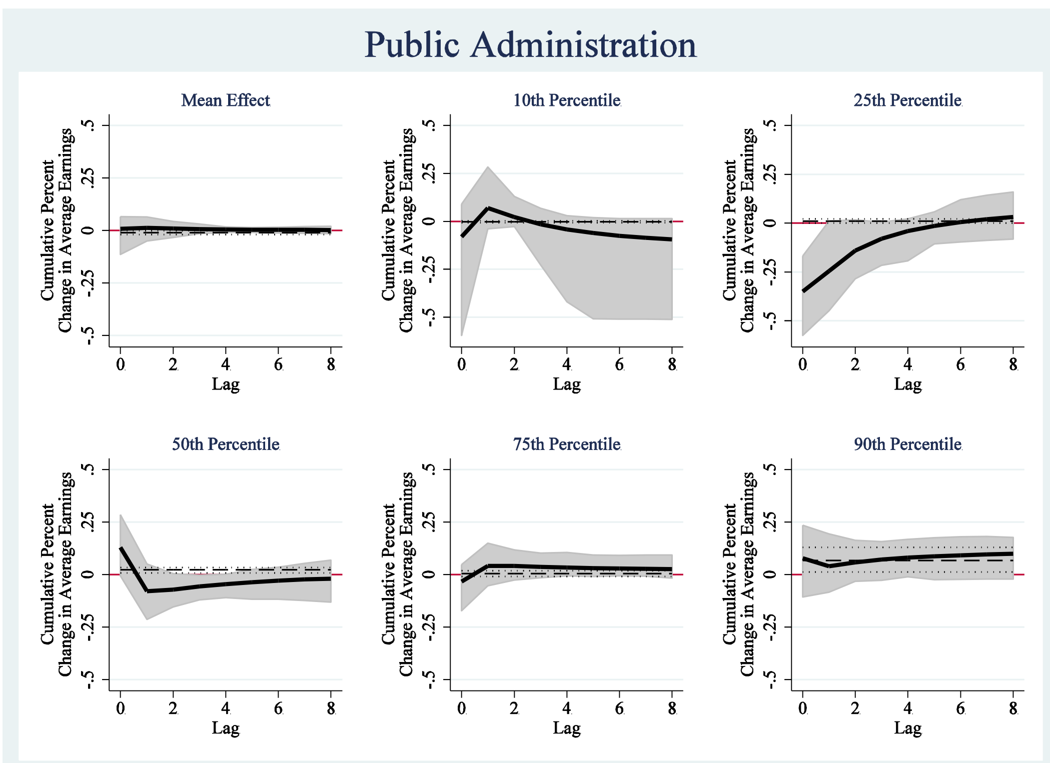

Transportation
Mean Effect
10th Percentile
25th Percentile
50th Percentile
75th Percentile
90th Percentile
Cumulative Net Job Flows
Lag


Transportation
Mean Effect
10th Percentile
25th Percentile
50th Percentile
75th Percentile
90th Percentile
Cumulative Percent Change in Average Earnings
Lag


Accommodation and Food Services
Mean Effect
10th Percentile
25th Percentile
50th Percentile
75th Percentile
90th Percentile
Cumulative Net Job Flows
Lag


Accommodation and Food Services
Mean Effect
10th Percentile
25th Percentile
50th Percentile
75th Percentile
90th Percentile
Cumulative Percent Change in Average Earnings
Lag


Public Administration
Mean Effect
10th Percentile
25th Percentile
50th Percentile
75th Percentile
90th Percentile
Cumulative Net Job Flows
Lag


Public Administration
Mean Effect
10th Percentile
25th Percentile
50th Percentile
75th Percentile
90th Percentile
Cumulative Percent Change in Average Earnings
Lag