Local Employment Impacts of Competing Energy Sources: the Case of Shale Gas Production and Wind Generation

Peter Hartley, Department of Economics, Rice University
Kenneth B. Medlock III, James A. Baker III Institute for Public Policy, Rice University
Ted Temzelides, Department of Economics, Rice University
Xinya Zhang, Department of Economics, Rice University *

Abstract

Nowadays, both the wind power and shale gas industries have been developing quickly and receiving significant attention in the media. Since wind and natural gas are competing sources of electricity generation, in order to guide policy, it would be useful to have an idea of how many jobs are created by these two competing resources.

In this paper, we use a panel econometric model to estimate the historical job-creating performance of wind versus that of shale oil and gas. The model is estimated using monthly county level data in Texas from 2001 to 2011. We collect data on the historical employment, number of directional/hydraulic fracturing wells drilled, and new wind capacity built in each county each month and then study the relationship between them. Both distributed lag and spatial panel models have been used based on different data dependence assumptions.

The results show that shale development and well drilling activities have brought strong employment and wage growth to Texas, while the impact of wind industry development on employment and wages statewide has been either not statistically significant or quite small.
1 Introduction

Nowadays, there is a lot of discussion in the media about job creation in the renewable energy industry. At the same time, commentators have talked up about the potential for renewable energy to provide greater energy independence and security, have notable environmental benefits due to reduced CO\textsubscript{2} emissions, and act as a driver for significant economic growth by fostering continual innovation.

Since energy produced through renewable sources is still more expensive than that produced through fossil fuels, state and local governments are spending tens of millions of dollars in subsidies to fund the renewable industry. More than half of all states have put in place Renewable Portfolio Standards\textsuperscript{1} to promote generation from renewable sources. Federal production tax credits and grants also contributed to increases in renewable capacity and generation between 2001 and 2011.

The renewable energy sector has developed quickly in the past 12 years. In particular, as seen in Figure 1, wind was the fastest growing source of non-hydroelectric renewable resource generation, as many operators of wind turbines have benefited from tax credit programs. Other sources of non-hydroelectric renewable electricity generation have included biomass, geothermal, and wood, but these have remained relatively stable since 2000.

In principle, renewable energy has the potential to create many jobs. Furthermore, many of these jobs are guaranteed to stay domestics, as they involve construction and installation of physical plant and facilities. Additionally, domestic wind turbine and component manufacturing capacity has increased. Eight of the ten wind turbine manufacturers with the largest share of the U.S. market in 2011 had one or more manufacturing facilities in the United States at the end of 2011. By contrast, in 2004 there was only one active utility-scale wind turbine manufacturer assembling nacelles in the United States (GE)	extsuperscript{2}.

The American Wind Energy Association (AWEA) estimates that the entire wind energy sector directly and indirectly employed 75,000 full-time workers in the United States at the end of 2011.

\begin{figure}[h!]
\centering
\includegraphics[width=\textwidth]{Figure_1.png}
\caption{Non hydro-power renewable energy generation, 1990-2011 \textit{Data source: EIA}}
\end{figure}

\textsuperscript{1} Renewable portfolio standards (RPS), also referred to as renewable electricity standards (RES), are policies designed to increase generation of electricity from renewable resources. These policies require or encourage electricity producers within a given jurisdiction to supply a certain minimum share of their electricity from designated renewable resources.

\textsuperscript{2} 2011 Wind Technology Market Report by U.S. Department of Energy
At the same time, the recent identification of the vast extent of shale gas and oil reserves and the development of cost-effective horizontal drilling and hydraulic fracturing techniques has caused U.S. production of shale oil and gas to boom. The Energy Information Administration’s 2012 Annual Energy Outlook (EIA 2012) projects that the share of shale gas as a part of total U.S. natural gas production will increase from 4 percent in 2005 to 34 percent by 2015 and 49 percent by 2025. As shown in Figure 2, shale gas is the largest contributor to natural gas production growth; there is relatively little change in production levels from tight formations, coalbed methane deposits, and offshore fields.

The development of shale gas resources has created an investment boom in the oil and gas industry and led to economic revitalization in places like North Dakota, Alberta, West Pennsylvania, Texas, and Louisiana to name a few. During 2007-2011, employment in the oil and gas extraction sector grew at an annual rate of 7.49 percent and 33.5 percent in total. By comparison, during the same period, total employment declined 3.3 percent below the starting value (Figure 3a). Meanwhile, states rich in shale gas have experienced a large increase in employment while the nationwide employment growth rate remains negative (Figure 3b). Furthermore, the substantially expanded U.S. natural gas supply at stable, relative low prices is stimulating downstream investment in natural gas using equipment by numerous manufacturing sectors\(^3\), as well as electricity generators, and the transportation sector. This activity is creating jobs and increasing wage income.

Both the wind power sector and shale gas sector have been developing quickly and receiving significant attention in the media. Since wind and natural gas are competing sources of electricity generation, in order to guide policy, it would be useful to have an idea of how many jobs are created by these two competing resources. While the aggregate net effect on employment from exploiting different sources of energy production is clearly an important question, it cannot be readily answered in the context of traditional macroeconomic models. The reason is that, as these models assume market clearing, they cannot easily account for variations in unemployment rates and thus are not well suited to study the consequences of alternative government policies for aggregate employment levels. Regardless of how one models the operation of labor markets, however, the

---

\(^3\) Especially manufacturing sectors that are sensitive to energy costs, such as basic chemicals, plastics & rubber, pharmaceuticals, aluminum, pesticides, paints, and fertilizers.
impact of any change can be gauged by examining the labor intensity of the different activities.

Generally speaking, two types of studies focus on the employment impacts in the energy industry. One is an input-output (I/O) model, the other is based on survey responses from employers, and uses simple descriptive and analytical techniques. In this study, we collected data on the historical job creation per unit of energy services produced for each energy source and used this data and a simple econometric model to estimate the historical job-creating performance of wind versus that of shale gas. It is a bottom-up approach, like the approach based on surveys. However, the econometric techniques used allow us to compare the employment impacts of these two different sectors in a more systematic and consistent way.

The next section is the literature review, and section 3 is the data description. The general econometric model is introduced in section 4, while estimation methods and results are reported in section 5. Section 6 is the conclusion.

2 Literature Review

There has been a large increase in reports on shale jobs and wind jobs in the past several years. Most previous analyses have been completed by non-government organizations, consulting firms, or universities but there have been few peer-reviewed journal publications.

Generally speaking, there are two types of studies that focus on the employment impacts in the energy industry. One is an input-output (I/O) model, which is intended to model the entire economy as an interaction of goods and services between various industrial sectors and consumers. The other is based on survey responses from employers, and uses simple descriptive and analytical techniques.

---

4 See Section 2 for detailed discussion.
For oil & gas industry, most reports are I/O model studies. Two widely-used I/O models are the IMPLAN model (See IHS (2012), UTSA (2012), Considine et al. (2009), and American-Chemistry-Council (2011)) and the RIMS II model by U.S. Bureau of Economic Analysis (BEA) (See Scott&Associates (2009)). All the studies we have investigated suggest shale oil and gas boom has a large impact on jobs, income and economic growth.

A study on Eagle Ford Shale (UTSA 2012) estimates the total economic output impact of shale activity on local 14-county region in 2011 was just under $20 billion dollars and supported 38,000 full-time jobs. If the studied region was extended to 20-county region, 47,097 full-time jobs were supported instead.

A nationwide shale industry report (IHS 2012) has found that in 2010, the shale gas industry supported 600,000 jobs, and this will grow to nearly 870,000 in 2015 and to over 1.6 million by 2035.

Two reports on Marcellus shale by Pennsylvania State University (Considine et al. 2009) and West Virginia University (Higginbotham et al. 2009) show that the oil and gas industry in Pennsylvania generates $3.8 billion in value added, and over 48,000 jobs in 2009; while in west Virginia, the economic impact of the oil and natural gas industry in 2009 is $3.1 billion in total value added and approximately 24,400 jobs created.

The Jobs and Economic Development Impact (JEDI) model developed by the National Renewable Energy Laboratory (NREL) is a series of spreadsheet-based I/O models that estimate the economic impacts of constructing and operating power plants, fuel production facilities, and other projects at the local (usually state) level. Slattery et al. (2009) employs the JEDI Wind Energy Model to examine economic impacts of the large-scale wind farm construction and tested the model validation using data from NextEra’s Capricorn Ridge and Horse Hollow facilities. They find that the JEDI model overestimates local share of jobs in construction phase in smaller, rural county, and underestimates number of jobs (more than 50%) in large, urban county. Obviously, JEDI model sets same local share value to all counties and does not consider urban effects as well. Plus, JEDI model assumes 100% local share for operations and maintenance (O&M) jobs, which may be implausible especially in small rural counties.

I/O models provide the most complete picture of the economy as a whole. They capture employment multiplier effects, as well as the macroeconomic impacts of shifts between sectors. Hence they could account for losses in one sector (e.g. conventional oil industry) created by the growth of another sector (e.g. the wind energy industry). However, collecting data for an I/O model is highly labor intensive, and the calibration process of default multiplier parameters may be biased due to lack of information and subjectivity.

On the other hand, bottom up estimates are based on industry/ utility surveys, the outlook of project de-

---

5The IMPLAN model uses a national input-output dollar flow table called the Social Accounting Matrix (SAM) to model the way a dollar injected into one sector is spent and re-spent in other sectors of the economy, and measure its economic multiplier effects. The RIMS II provides solely I/O multipliers that measure output, employment, and earnings effects of any changes in a regions industrial activity.

velopers and equipment manufacturers, and primary employment data from companies across manufacturing, construction, installation, and O&M. For wind energy, most reports are analytical-based studies, and only calculate direct employment impacts.

A case study on economic effects of Gulf wind project in Texas reports that they would create 250 - 300 jobs during peak construction period (9 months), and 15 - 20 permanent jobs. A report on wind industry from Natural Resources Defense Council (NRDC) measures number of direct jobs that a typical wind farm may create across the entire value chain. They analyzed each of the 14 key value chain activities independently to determine the number of workers involved at each step in the wind farm building. And they found that just one typical wind farm of 250MW would create 1079 jobs over the lifetime of the project.

Similarly, the Renewable Energy Policy Project (REPP) has developed a spreadsheet-based format of the calculator using data based on a survey of current industry practices. It is used to calculate the number of direct jobs from wind, solar photovoltaic, biomass and geothermal sectors as a result of enactment of an RPS. According to the calculator, every 100 MW of wind power installed provides 475 jobs in total (313 manufacturing jobs, 67 installation jobs, and 95 jobs in O&M).

3 Data Description

In this paper, we use data from Texas, because it contains rich shale gas and oil resources while also being the national leader in wind installations and a manufacturing hub for the wind energy industry. According to EIA, Texas accounted for 40 percent of U.S. marketed dry shale gas production in 2011, making it the leading unconventional gas producer among the states. Meanwhile, Texas leads the nation in wind-powered generation capacity and is the first state to reach 10,000 megawatts of wind capacity.

In Texas, there are 254 counties. For each county \( i = 1, ..., 254 \), we have observations in \( T = 132 \) months of 11 years (2001 - 2011), making the panel balanced.

I took total employment in all industries as a dependent variable. I did not use data of specific energy industries because, besides direct job creation, I want to consider the total employment effects, including indirect jobs, such as jobs created in upstream and infrastructure supplying industries, and induced jobs, such as jobs added in sectors supplying consumer items (food, auto, and housing, etc.) and services. Another candidate dependent variable is the average weekly wage, since it may also be impacted by an increase in the demand for workers. We use monthly employment data and quarterly wage data from Quarterly Census of Employment and Wages (QCEW) Database of Bureau of Labor Statistics (BLS). The latter has been

---

7 Slides is available at Gulf Wind: Harnessing the Wind for South Texas
8 More information is available at http://www.repp.org/labor/
9 77 of 254 are urban counties.
10 QCEW employment and wage data are derived from microdata summaries of 9.1 million employer reports of employment and wages submitted by states to the BLS in 2011. These reports are based on place of employment rather than place of residence.
adjusted to a real wage using the implicit price deflator (IPD) of GDP from BEA.\footnote{An implicit price deflator of GDP is the ratio of the current-dollar value of GDP, to its corresponding chained-dollar value, multiplied by 100.}

In order to evaluate the impact of shale and wind industry development on employment and the local economy, we need to devise a method for measuring the activity of the shale and wind industries. The key explanatory variables we use are the number of unconventional wells completed and the new installed wind capacity in each county each month.

Other variables could be used to reflect other aspects of shale activity. These include the number of permits, rig counts, the number of wells spudded, and shale gas production. We choose the number of wells completed because the completion date indicates the end of the construction period of each well. During the construction period, more direct and on-site jobs are created; after the construction, on-site jobs decrease while indirect and induced effects would last. To fully describe the impact of shale on employment, especially the multiplier effects of job creation in the local economy, we allow well drilling activities to affect employment with a lag and study both pre-completion and post-construction effects.

In the shale industry, the entire process from spudding to producing marketed output can take up to 3-4 months, among which horizontal drilling currently takes approximately 18-25 days from start to finish. Then wells are fractured to release the gas before the well is completed. It is then connected to a pipeline, which transports the gas to the market. Among all these steps, hydraulic fracturing is most labor intensive and the last step before the completion. Hence we expect drilling activities to have a peak impact on employment in the pre-completion period in the month of well completion.

The well information is taken from the Drilling Info Database\footnote{Data is available at \url{http://info.drillinginfo.com/}}. We choose wells that are both directional/horizontal drilled and hydraulically fractured\footnote{This filter option is only available for Texas data}, so that we exclude conventional oil/gas wells from our data set. There were 31050 directional/horizontal and fractured wells completed in 174 counties of Texas during 2001 - 2011, including 25467 gas wells, 4963 oil wells, and 620 other types of wells. From Figure 4, we can see that shale gas developed very quickly in the

![Figure 4: Number of completed new wells, Jan, 2001 - Dec, 2011](image)
past 12 years, from 1 well per month in Jan, 2001 to around 500 in 2011. The completion date and location
of each well are used to count the number of wells completed in each county each month.

To measure wind activity in each county we used installed nameplate capacity online per month. Power
generation data is not used because more jobs are created during the construction period than in the O&M
period. The installed capacity and online year of all wind projects in Texas through 2007-2011 can be found at
American Wind Energy Association (AWEA). For the wind projects before 2007, I used EIA electricity data on
plant level output\textsuperscript{14} and a wind industry progress report by Wind Today. To find the online month and county
location of each wind project, I referred to some additional sources, such as project information from projects’
websites and local news of its online year. For those wind farms that cover several neighboring counties,
I divided installed capacity of farms equally between each of the counties. Until 2011, 125 wind projects
had been constructed in 40 counties, with total installed capacity of 10006MW, compared to 6 counties and

4 Econometric Model

We estimated the original regression relationship treating the data via a panel data approach given data set
of 254 counties in Texas covering the years 2001-2011. Panel data has several advantages relative to either
time series or cross section data. For one, it allows us to look at dynamic relationships which we cannot do
with a single cross section. A panel data set also allows us to test for effects in counties with shale and wind
activities and those without, which cannot be done with a time series alone. A major problem with straight
time series analyses is that many exogenous factors change at the same time making it difficult to assign an
outcome to any one particular change. The panel enables us to interpret differences between counties over
time as policies vary in both dimensions.

We did a panel data approach given a data set of 254 counties in Texas during 2001-2011. Comparing to
time series and cross section data, having data over time for the same counties is useful for several reasons.
For one, it allows us to look at dynamic relationships which we cannot do with a single cross section. A panel
data set also allows us to control for unobserved cross section heterogeneity.

4.1 Assumptions

We start with a static linear unobserved effects model

$$y_{it} = x_{it}\beta + \theta_t + c_i + u_{it}, t = 1, 2, ..., T,$$  \hspace{0.5cm} (1)

where $y_{it}$ is a scalar, $x_{it}$ is a $1 \times K$ vector for $t = 1, 2, ..., T$, and $\beta$ is a $K \times 1$ vector. $c_i$ is a time-invariant
unobservable effect, and $\theta_t$ represents a series of time fixed effects.

\textsuperscript{14}Data is available at EIA website.
To make the model more realistic, we allow for arbitrary dependence between the unobserved effects $c_i$ and the observed explanatory variables $x_{it}$. For example, underground geology characteristics would be included in $c_i$ and these would undoubtedly be correlated with the number of wells drilled in county $i$. Also, wind capacity highly depends on the climate and especially the wind resource of the county, which is part of $c_i$ as well.

Another assumption I would like to make is that the explanatory variables are strictly exogenous conditional on the unobserved effect $c_i$. This terminology, introduced by Chamberlain (1982), requires that

$$E(u_{it}|x_i,c_i) = 0, t = 1, 2, ... , T.$$  

That is to say, once $x_{it}$ and $c_i$ are controlled for, $x_{it}$ has no partial effect on $y_{it}$ for $s \neq t$ and $u_{it}$ has zero mean conditional on all explanatory variables in all time periods.

Typically, we feel comfortable with assuming zero contemporaneous correlation, that is, $u_{it}$ is uncorrelated with the number of wells drilled or the wind capacity installed at $t$, but what about correlation between $u_{it}$ and, say, $x_{i,t+1}$? Does future well drilling activity or wind farm construction depend on shocks to the county employment in the past? We don’t think such feedback is very important in our case, since total employment of the county is certainly not the main goal of energy companies. So it seems reasonable to assume that past employment has few, if any, effect on energy companies’ future decision making processes.

Another issue is that the explanatory variables could have lasting effects, so that correlation exists between $u_{it}$ and past $x_{i,t-1}, ..., x_{i,1}$ and sequential exogeneity fails. It is likely to be the case here since we expect well drilling activity and wind activity to have lasting effects on local employment. One way to soak up correlation is to include lags of explanatory variables into the model. Strict exogeneity would still hold if enough lags are included. The other way is to use instrumental variables (IV). However, the IV method is usually not recommended because it is often difficult to find suitable instruments.

Note that the strict exogeneity assumption never holds in unobserved effects models with lagged dependent variables. The reason is that $y_{it}$ is correlated with $u_{it}$ and would show up as part of explanatory variables at $t + 1$ so $E(u_{it}|x_{i,t+1}) \neq 0$. Additional care is required when we include lagged dependent variables as explanatory variables on the right hand side.

### 4.2 Finite Distributed Lag Model

Since we expect drilling and wind activity could have lasting effects on local employment, we should include lags of explanatory variables into the model. A finite distributed lag (FDL) model might be appropriate if the impact of the explanatory variables lasts over a finite number of periods $q$ and then stops. The FDL unobserved effects model expands equation (1) to the form:

$$E_{it} = \sum_{k=0}^{q} \beta_k \text{wells}_{i,t-k} + \sum_{k=0}^{q} \delta_k \text{wcap}_{i,t-k} + c_i + \theta_t + u_{it}$$  

(3)
where $E_{it}$ denotes total employment, $wells_{it}$ denotes number of directional/fractured wells drilled, and $wcap_{it}$ is installed wind capacity for $i = 1, 2, \ldots, 254$ and $t = 1, 2, \ldots, T$. Our interest lies in pattern of coefficients $\{\beta_k, \delta_k\}_{q=0}^T$. $\beta_0$ and $\delta_0$ are the immediate change in $E_i$ due to the one-unit increase in $wells_i$ and $wcap_i$ respectively at time $t$. Similarly, $\beta_k$ and $\delta_k$ are the changes in $E_i$ $k$ periods after the temporary change. At time $t + q$, $E_i$ has reverted back to its initial level: $E_{i,t+q} = E_{i,t-1}$.

We are also interested in the change in $E_i$ due to a permanent increase in any of the explanatory variables. For example, following a permanent increase in $wells_i$, after one period, $E_{i,t+1}$ has increased by $\beta_0 + \beta_1$, and after $k$ periods, $E_{i,t+k}$ has increased by $\beta_0 + \ldots + \beta_k$. There are no further changes in $E_i$ after $q$ periods. This shows the sum of the coefficients on current and lagged $wells_i$ is the long-run change in $E_i$, which is also referred to as the long-run propensity (LRP). For the impact of variable $wcap_i$ on $E_i$, the same story applies.

However, it is rarely the case that we actually know the right lag length or have a strong enough theory to inform us about it. Some other problems may also arise with an FDL model. For example, time series are often short and so the inclusion of the lagged variables may eat up a lot of degrees of freedom. In addition, the fact that the explanatory variables are likely to be highly correlated is likely to lead to severe multicollinearity.

### 4.3 Autoregressive Distributed Lag Model

We can solve the multicollinearity problem mentioned above by including a lagged dependent variable with fewer lags of explanatory variables, and the model changes to the autoregressive distributed lag (ADL) model. In many ways, the ADL model is similar to the FDL model, except it is now easy to see that the impact of explanatory variables persists over time but at a geometrically declining rate. Denoting the number of lagged dependent variables as $p$, an ADL($p,q$) model with unobserved effects has the form:

$$E_{it} = \sum_{j=1}^{p} \lambda_j E_{i,t-j} + \sum_{k=0}^{q} \beta_k wells_{i,t-k} + \sum_{k=0}^{q} \delta_k wcap_{i,t-k} + c_i + \theta_t + u_{it}$$

where $\{\lambda_j\}_{j=1}^p$ are autoregressive coefficients. If there is a temporary change in wells, $E_{it}$ will initially go up by $\beta_0$ in period 1, then by $\beta_1 + \lambda_1 \beta_0$ in period 2, and then by $\beta_2 + \lambda_1 (\beta_1 + \lambda_1 \beta_0) + \lambda_2 \beta_0$ in period 3 etc. In other words, the effect of having a lagged dependent variable is to make the effect from the previous period persist. Eventually, the effect of the impulse will disappear and we will return to our original equilibrium as long as the process is stationary.

Another advantage of the ADL model is that the inclusion of a lagged dependent variable will often eliminate the serial correlation, particularly if additional lags of the dependent variable are included. Lags of the independent variables may also assist with eliminating serial correlation in the error term. Hence, once we start putting any lagged values of $y_{it}$ into explanatory variables, dynamic completeness is an intended assumption, which clearly implies sequential exogeneity. However, the strict exogeneity assumption is necessarily false as we discussed before. In this case, both the fixed effects (FE) estimator and the first difference (FD) estimator
are inconsistent.

Making decision which model to use and how many lags to include is complicated by the fact that we are unlikely to have enough theory to distinguish between the different models. As a result, Boef & Keele (2008), along with many others, argue that you should start with a general model like the ADL and test down to a more specific model, including the optimal values for \( p \) and \( q \).

### 4.4 Spatial Panel Models

In this section, we discuss cross-sectional dependence (XSD) in panels. This can arise, for example, if spatial diffusion processes are present, relating panel members (in our case counties) in a way that depends on a measure of distance. The Pesaran CD test and CD(\( p \)) test (Pesaran 2004) are used to detect XSD.

In our data set, both tests show the presence of XSD at 0.000 level. This is not surprising, since it seems likely that employment might be correlated across counties. Therefore, we use a spatial panel model to study this spatial interaction effect across counties and try to capture the indirect effect of a county’s energy sector development on employment within other counties. Spatial interaction effects could be due to competition or complementarity between counties, spillovers, externalities, regional correlations in industry structures and many other factors.

Interactions between spatial units are typically modeled in terms of some measure of distance between them, which is described by a spatial weights matrix \( W \). \( W \) is a 254 × 254 non-negative matrix, in which the element \( w_{ij} \) expresses the degree of spatial proximity between the pair of objects \( i \) and \( j \). Following Kapoor et al. (2007), the diagonal elements \( w_{ii} \) are all set to zero to exclude self-neighbors. Furthermore, only neighborhood effects are considered in this paper, that is, \( W \) is a contiguity matrix\(^{15}\):

\[
 w_{ij} = \begin{cases} 
 1, & \text{if } i \text{ and } j \text{ are neighbors} \\
 0, & \text{otherwise} 
\end{cases} \tag{5}
\]

Then the contiguity matrix is transformed into row-standardized form, which assumes the impact on each unit by all other neighboring units are equal. Given a spatial weights matrix \( W \), a family of related spatial econometric models can be expanded from equation (1):

\[
 E_{it} = \rho \sum_{j=1}^{N} w_{ij} E_{jt} + \beta_1 \text{wells}_{it} + \beta_2 \text{wcap}_{it} + u_{it}, \tag{6}
\]

where \( \rho \) is the spatial autoregressive coefficient. Followed Kapoor et al. (2007), the composite error \( u_{it} \) follows a first order spatial autoregressive process of the form:

\[
 u_{it} = \lambda \sum_{j=1}^{N} w_{ij} u_{jt} + \epsilon_{it} \tag{7}
\]

\(^{15}\)\( W \) is also called as adjacency matrix.
and \( \epsilon \) follows an error component structure

\[
\epsilon_{it} = c_i + \nu_{it}
\]

(8)
to further allow \( \epsilon_{it} \) to be correlated over time.

A spatial panel model may also contain a spatially lagged dependent variable (\( \rho \neq 0 \)), known as a spatial autoregressive (SAR) model. Alternatively, there also could be a spatial autoregressive process in the error term (\( \lambda \neq 0 \)), in which case the model is known as a spatial error model (SEM).

The spatial lag model posits that the dependent variable depends on the dependent variable observed in neighboring units and on a set of observed local characteristics. The spatial error model, on the other hand, posits that the dependent variable depends on a set of observed local characteristics and that the error terms are correlated across the space.

5 Estimation Methods and Results

Let us first turn to the general unobserved effect model (1). The pooled OLS estimator can be used to obtain a consistent estimator of \( \beta \) only if the explanatory variables satisfy contemporaneous exogeneity and zero correlation with \( c_i \). However, as we discussed in section 4.1, explanatory variables are necessarily correlated with unobserved individual effects \( c_i \). In addition, F tests of poolability show pooled OLS estimation is inconsistent. Hence, the pooled OLS estimator should not be used.

Since random effects analysis also requires orthogonality between \( c_i \) and observed explanatory variables, as well as strict exogeneity, it is also inconsistent and inappropriate to be used. The result of the Hausman test, namely \( \chi^2 = 601.67 \) with a p-value close to zero, again indicates that the random effects approach is inconsistent.

With a fixed effects (FE) or first difference (FD) approach, the explanatory variables are allowed to be arbitrarily correlated with \( c_i \), but strict exogeneity of them conditional on \( c_i \) is still required. The idea behind the fixed effects approach is to transform the equations by removing the inter-temporal mean and thereby eliminating the unobserved effect \( c_i \). One can then apply pooled OLS to get FE estimators. Similarly, the FD approach transforms the equations by lagging the model one period and subtracting, then applying pooled OLS to get FD estimators.

5.1 Estimation of FDL model

As noted in section 4.1, including lagged dependent variables in the model violates strict exogeneity, implying that the resulting autoregressive FD estimator may suffer from asymptotic bias. Therefore, in this section we drop all lagged dependent variables and use the FDL approach, equation (3). We verified that the strict exogeneity assumption holds as long as enough lags of the explanatory variables are included.
5.1.1 First Difference Estimator

To get the FD estimator, we lagged the model (3) one period and subtracted to obtain:

$$\Delta E_{it} = \sum_{k=0}^{q} \beta_k \Delta \text{wells}_{i,t-k} + \sum_{k=0}^{q} \delta_k \Delta \text{wcap}_{i,t-k} + \theta_0 + \theta_t + \Delta u_{it}, t = 2, 3, ..., T,$$

(9)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>(Std. Err.)</th>
<th>(Robust SE.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wells_{it}</td>
<td>16.31</td>
<td>(5.396)**</td>
<td>[6.06]**</td>
</tr>
<tr>
<td>wells_{i,t-1}</td>
<td>13.17</td>
<td>(6.666)*</td>
<td>[7.081]</td>
</tr>
<tr>
<td>wells_{i,t-2}</td>
<td>0.932</td>
<td>(7.025)</td>
<td>[2.929]</td>
</tr>
<tr>
<td>wells_{i,t-3}</td>
<td>-5.519</td>
<td>(7.127)</td>
<td>[6.006]</td>
</tr>
<tr>
<td>wells_{i,t-4}</td>
<td>12.23</td>
<td>(7.119)</td>
<td>[8.705]</td>
</tr>
<tr>
<td>wells_{i,t-5}</td>
<td>17.89</td>
<td>(6.875)**</td>
<td>[11.13]</td>
</tr>
<tr>
<td>wells_{i,t-6}</td>
<td>22.46</td>
<td>(5.686)***</td>
<td>[12.91]</td>
</tr>
<tr>
<td>wcap_{it}</td>
<td>-0.756</td>
<td>(1.235)</td>
<td>[0.923]</td>
</tr>
<tr>
<td>wcap_{i,t-1}</td>
<td>-0.755</td>
<td>(1.653)</td>
<td>[0.594]</td>
</tr>
<tr>
<td>wcap_{i,t-2}</td>
<td>-0.739</td>
<td>(1.864)</td>
<td>[0.332]*</td>
</tr>
<tr>
<td>wcap_{i,t-3}</td>
<td>-0.212</td>
<td>(1.923)</td>
<td>[0.323]</td>
</tr>
<tr>
<td>wcap_{i,t-4}</td>
<td>0.111</td>
<td>(1.865)</td>
<td>[0.374]</td>
</tr>
<tr>
<td>wcap_{i,t-5}</td>
<td>0.250</td>
<td>(1.654)</td>
<td>[0.432]</td>
</tr>
<tr>
<td>wcap_{i,t-6}</td>
<td>-0.178</td>
<td>(1.236)</td>
<td>[0.266]</td>
</tr>
</tbody>
</table>

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, · $p < 0.1$

Table 1: FD Estimation Results, $q = 6$

To test for the presence of serial correlation in $\Delta u_{it}$, I use Breusch-Godfrey test and Wooldridge’s test for serial correlation in panels. Both tests reject $H_0$ and show serial correlation remains in the idiosyncratic errors. I then increased $p$ one by one and ran the test again until $p = 36$. The results showed that serial correlation remained no matter how many lags of the explanatory variables were included. Existence of serial correlation may imply that the model fails to capture the actual dynamic adjustment process. Nevertheless, the FDL model does have the advantage that it satisfies the requirement of strict exogeneity.

When the serial correlation remains in the error term, we should compute a robust variance matrix for the FD estimator, which accommodates a fully general structure with respect to heteroskedasticity and serial correlation in $\Delta u_{it}$. Following Arellano (1987), this robust variance matrix is consistent and it relies on large $n$ asymptotics with small $T$.

To determine the appropriate lag length $q$, I posited a maintained value that should be larger than optimal.
Here I use 24. Then I did sequential $F$ tests on the last $24 > p$ coefficients, stopping when the test rejects the $H_0$ that the coefficients are jointly zero at 5% level. Using a robust variance matrix to calculate the $F$ statistics, we drop 18 lagged explanatory variables and assign $q = 6$.

The estimation results are reported in Table 1, with both robust standard errors and usual FD standard errors. From the regression results, and using robust standard errors, all coefficients of wind installed capacity to be close to zero and not statistically significant (except the order 2 lag is negative and significant at 5% level). A joint $F$ test on $H_0 : \delta_k = 0$ for $k = 0, 1, \ldots, 6$ shows $F(7, 31734) = 0.78$ with $p-value = 0.6001$, and we cannot reject $H_0$ that the impact of wind activity on employment is not statistically significantly different from zero.

![Short term impact of shale activity on employment](image1)

![Short term impact of wind activity on employment](image2)

Figure 5: FD estimation results with $q = 6$: (a) wells (b) wind capacity

On the other hand, we found $wells_t$, $wells_{t-1}$,... and $wells_{t-6}$ to be jointly significant: the $F$ statistic has a $p-value = 0.0007$. Adding the estimated coefficients of current and lag variables, we have long term multipliers $LRP_{wells} = 77.46$. Assuming that all the jobs created are short-term and only last for 1 month, we divide $LRP_{wells}$ by 12 to get the number of full time equivalent (FTE) jobs 6.42. Given 5482 new directional/fractured wells were drilled in Texas in 2011, about 35000 FTE jobs would have been created.

We graph the estimated short run impact of $wells_k$ and $wcap_k$ as a function of $k$ in Figure 5. The lag distribution summarizes the dynamic effects that a temporary increase in explanatory variables has on the dependent variable. From Figure 5a, we see a generally decline trend on the impact of wells in the first three months, which is expected because workers leave after the well completion. Then the employment increases starting at month 4. It is probably because the emerging of new business opportunities in the neighborhood due to the well drilling activity. We find that the largest effect is at the first and the last lag. Figure 5b shows the impact of new wind capacity added. It may show some useful trending information even though the results is hardly significant. We see the employment effect is negative first and then increase. It peaks about

16This could solve the double counting problem effectively.
four months after the wind farm construction and then declines again. After six months, the impact of well drilling and wind activities fades and the employment falls back to the original level.

Note that we have a really low $R^2 = 0.00084$, which measures the amount of variation in employment that is explained by wells. Since oil and gas related employment is only 2.6% of the total employment in Texas, a low explanatory power of the regression model is to be expected.

### 5.2 Estimation of ADL model

In this section, we include lag dependent variables into the model. Since the ADL model contains lagged dependent variables, as we discussed in section 5.1, the strict exogeneity assumption is violated, and neither FE nor FD estimators are consistent. In this case, the generalized method of moments (GMM) is used.

#### 5.2.1 Generalized method of moments estimator

Following Arellano & Bond (1991), we applied the two-way GMM method to estimate the ADL model, shown in equation (4). We again need to assign appropriate $p$ and $q$ to the model before we estimate it. As before, we start with large enough $p$ and $q$ that they are guaranteed to be larger than their optimal value: $p = q = 24$.

First, when we included one lagged dependent variable $E_{i,t-1}$, Wooldridge’s test for serial correlation reports $\chi^2 = 30.189$ with $p-value = 3.919e - 8$. Strong serial correlation implied that the dynamic data generation processes has not been fully captured. When we included one more lagged dependent variable $E_{i,t-2}$, the test result changed to $\chi^2 = 0.0081$ with $p-value = 0.9285$, and we can conclude that the error $u_{it}$ is then serially uncorrelated. In this case, we say the model is dynamically complete since enough lags have been included so that further lags of dependent and independent variables do not matter for explaining $E_{it}$. Hence we set $p = 2$.

As in previous section I then set $q = 6$ and estimated the two way Arellano-Bond GMM regression. Note that both Wald test and the joint $F$ test cannot reject the coefficients of wind capacity $\delta_0 = \ldots = \delta_6 = 0$ in model, which implies no impact of wind activity on local employment. In the following discussion we therefore focus solely on the wells variable.

Figure 6 graphs the resulting dynamic response of employment to a unit increase in $wells_{it}$ and $wcap_{it}$ in six lags. The employment $E_{it}$ initially go up by $\beta_0$ in period 1, then by $\beta_1 + \lambda_1\beta_0$ in period 2, and then by $\beta_2 + \lambda_1(\beta_1 + \lambda_1\beta_0) + \lambda_2\beta_0$ in period 3, etc. We find the graphs of employment in Figure 6 and Figure 5 have similar trends. This helps us improve conclusion validity.

### 5.3 Estimation of Spatial Panel Models

Recall that in 4.4 we discussed the theory behind the spatial autoregression (SAR) and spatial error model (SEM). In the SAR model, the inclusion of the dependent variable on the right hand side of the above equation
introduces simultaneity bias and the OLS estimator is no longer unbiased and consistent, while in the SEM, the OLS estimator is unbiased, but inefficient. Therefore, maximum likelihood estimation is used to estimate the parameters of both models.

Both the SAR and SEM models are estimated allowing for two-way fixed effects. The results are reported in Table 2. We find that the spatial interaction coefficients of both models are significant at 0.000 level and very similar: $\rho = 0.1730, \lambda = 0.1734$. Also, both models show large and significant coefficients of wells, and insignificant coefficients of \( wcap \).

| SAR Coefficients: | Estimate | Std. Error | t-value | Pr(>|t|) |
|-------------------|----------|------------|---------|----------|
| $\rho$            | 0.1730   | 0.0081     | 21.43   | $<2e^{-16}^{***}$ |
| wells             | 224.72   | 12.99      | 17.29   | $<2e^{-16}^{***}$ |
| newcap            | 0.05     | 6.366      | 0.0079  | 0.9937   |

| SEM Coefficients: | Estimate | Std. Error | t-value | Pr(>|t|) |
|-------------------|----------|------------|---------|----------|
| $\lambda$         | 0.1734   | 0.0081     | 21.42   | $<2e^{-16}^{***}$ |
| wells             | 235.81   | 13.63      | 17.30   | $<2e^{-16}^{***}$ |
| newcap            | 0.47     | 6.374      | 0.704   | 0.4814   |

| Signif. codes:   | 0 *** 0.001 ** 0.01 * 0.05 .01 1 |

Table 2: Spatial interaction effects on employment

Following LeSage \& Pace (2009), the expectation of the SAR model $y = \rho Wy + X\beta + \epsilon$ is

$$E(y) = (I_N - \rho W)^{-1}X\beta \quad (10)$$

We thus find that employment in county \( i \) depends on developments in neighboring counties as workers in
bordering counties migrate to take advantage of new job opportunities due to shale and wind activity. This provides a motivation for the spatial lag variable $Wy$.

The own- and cross-partial derivatives for the SAR model take the form of an $N \times N$ matrix that can be expressed as:

$$\frac{\partial y}{\partial x'} = (I_N - \rho W)^{-1} I_N \beta_r$$

(11)

These partial derivatives show how drilling/wind activities in county $j$ influence employment in county $i$. For the $r$th explanatory variable, the average of the main diagonal elements of this matrix is labelled as the direct effect, and the average of cumulative off-diagonal elements over all observations corresponds to the indirect effect. The average total effect will be the sum of the two.

This model implies that direct effect of well drilling activity on employment is 225, and it is significant at the 0.000 level. The direct effect measures how wells drilled in a particular county affect employment in that same county. The result shows that about 225 jobs would be created by drilling a well in the same county. Also, the indirect effect estimate of well drilling activity is 46 which makes the total effect grow to 271. If we only consider the direct effect, the results would be underestimated 17%. Compared to FD estimation results, $LRP = 77$, 271 is much higher. This because we did not consider time serial correlation in the model and make the results biased\(^\text{17}\). However, it gives a useful hint that spatial correlation should be considered in our case. Otherwise, the results may be underestimated.

The direct and indirect effects of wind activity are 0.05 and 0.01, respectively, and they are not statistically significant as we found from FD and GMM models. Hence, wind farm installation and construction has not been found to have any impact on employment.

### 5.4 Estimation Results on Wage

We also looked for impacts of shale gas and wind developments on weekly wages instead of employment. Like the employment regression results, the FD approach with robust standard error has been used and sequential F-tests determine that $q = 12$. The results show that the coefficients of 4\(^{th}\) and 9\(^{th}\) lagged wells are significant at 0.05 level; for wind capacity, the coefficients of lag 1 and lag 10 are significant at 0.05 level, while lag 11 and 12 significant at 0.01 level. Figure 7 graphs the resulting dynamic response of wage to a unit increase in wells_{it} and wcap_{it} in 12 lags. We see that the impact of wells rises and falls with a 6 month cycle. The peaked value is about 0.3. The impact of wind capacity installation shows a quite different trend: it increases over time from near zero to 0.13.

The spatial panel regression results for wages are in line with the results without spatial interaction effects. According to SAR model, the estimate of $\beta_1$ is 0.18 and significant at 0.05 level, while the estimate of $\beta_2$ is

\(^{17}\) The model need to be fixed in future research
0.06 and not significant. Additionally, the results show strong spatial correlation: $\rho = \lambda = 0.26$. Note that the spatial correlation effects of wage is even larger than that of employment. 26% increased wage is due to indirect effects while the number in employment case is 17%.

The direct and indirect effect of well drilling activity on wage are 0.18 and 0.06, respectively. And we could say the total effect on wage is 23 cents per well drilled, of which 18 cents is due to drilling activity in the same county, and 6 cents is attributed to drilling activity in the neighbors. The total effect of wind activity is 8 cents per MW, with about 6 cents of direct effect and 2 cents of indirect effect.

6 Conclusions

In this study, we develop a general econometric model to compare job creation in wind power versus that in the shale gas sector. We have discussed the advantages and disadvantages of a number of different models. We then estimated them using county level data in Texas from 2001 to 2011.

Despite different estimation methodologies, the results were quite consistent. Both first difference and GMM methods show that shale development and well drilling activity have brought strong employment to Texas: 77 short-term jobs or 6.4 FTE jobs per well. Given 5482 new directional/fractured wells were drilled in Texas in 2011, about 35000 FTE jobs would have been created. Its impact on wage is not rather distinct. The wage increases 30 cents in month 4 and month 9 after each well completion.

All the estimations show that the impact of wind industry development on employment is not significant from zero. Its impact on wage increases gradually after the construction and peaks about one year later. 13 cents are added to the wage in month 10 to 12.
References


[URL: http://ideas.repec.org/a/bla/restud/v58y1991i2p277-97.html]


