The Impact of the Fracking Boom on Rents in Pennsylvania

Lucija Muehlenbachs
Elisheba Spiller
Andrew Steck
Christopher Timmins*†

February 15, 2015

Abstract

Booming shale gas development has been blamed for dramatic rises in housing rental costs, with important distributional consequences. We consider the effect of rapid shale development on Pennsylvania census tract rents, using New York tracts (located on the Marcellus shale formation but under a fracking moratorium) as a control group. We develop a new estimator that combines matching and quantile regression techniques. Results suggest that the impact of shale gas development on rents varies dramatically over the distribution of unobservables and that large effects are confined to only the upper tail of the distribution.

JEL Classifications – Q51, Q33, R00

Keywords – Hydraulic Fracturing, Fracking, Shale Gas, Rental Market, Matching, Quantile Regression

*Corresponding author

†Muehlenbachs: University of Calgary and Resources for the Future, *muehlenbachs@rff.org*. Spiller: Environmental Defense Fund, New York, *espiller@edf.org*. Steck: Department of Economics, Duke University, *andrew.steck@duke.edu*. Timmins: Department of Economics, Duke University, *christopher.timmins@duke.edu*. We thank the Bureau of Topographic and Geologic Survey in the PA Department of Conservation and Natural Resources for data on well completions. We also thank Shakeeb Khan, Jessica Chu, and seminar participants at Duke University. Any errors are our own.

I Introduction

The recent technological developments of directional (horizontal) drilling and hydraulic fracturing (fracking) have allowed previously inaccessible shale oil and gas deposits to radically alter the United States energy markets. Combined, the techniques have allowed resource producers to profitably extract hydrocarbon resources from shale rock formations. As a consequence, drilling activity in the U.S. has increased dramatically over the last seven years. Most new drilling has been concentrated on a few shale formations: the Marcellus shale in Pennsylvania, the Bakken shale in North Dakota, and the Eagle Ford and Barnett shales in Texas. The rapid increase in drilling activity in previously remote areas has exhibited similarities to classic natural resource booms. Local economies have undergone rapid changes, not all of them positive.

The aim of this paper is to empirically examine the effect that the fracking boom has had on housing rental markets in Pennsylvania. Pennsylvania is home to much of the Marcellus shale formation, which stretches from southeast Ohio across West Virginia and Pennsylvania into upstate New York. It is the largest estimated source of recoverable natural gas in the United States. Much of the Marcellus' drilling activity has been concentrated in Pennsylvania, where fracking and horizontal drilling arrived on the scene in late 2005. Growth of wells in the area has been tremendous in the years since: Pennsylvania saw four new wells spudded in 2005 compared to 1,964 in 2011. Drilling of new wells has grown at a compounded annual rate of more than 250% through 2012, despite concurrent headwinds of a wider economic recession and a steep decline in natural gas prices.[1]

Rents are an interesting economic variable in their own right, but also for the formative role they play in the pricing of real estate. In an asset-pricing framework, the value of a house is comprised of the discounted stream of future rents an owner will receive and the expected future price. An analysis of changes in rents therefore complements studies of housing prices by providing a more complete picture of the real estate market. If, for example, housing prices are found to decline while rents increase, one might infer that the expected sales price of the house has fallen drastically.

[1]Authors' calculations. Data from PADEP and DCNR, detailed in Section III.

Rent increases may also be detrimental from a social welfare point of view. If marginal renters become homeless and stretch the safety net resources of the local municipality, landlords will benefit at the expense of the tax base. In a similar fashion, rent increases are likely to have distributional consequences: renters tend to have less wealth than homeowners and are less able to cope with the shock of increased cost of shelter. Additionally, expensive rental housing will require gas firms to pay their workers more, increasing the cost of extracting gas. Finally, large shifts in rents have the potential to cause economic damage over the long term. Residential investment might increase as a result of the boom, only to leave investors holding a loss as the well-workers move on and the surge in demand subsides. Relatedly, higher volatility in rents will reduce residential investment below the socially optimal level if the investors are risk-averse.

Anecdotes surrounding rent increases abound in the media. For instance, Detrow (2013) reports that rents in Towanda, PA rose from $300 per month in 2008 to more than $1,000 per month in 2013, a nominal increase of 233%! These higher rents were accompanied by additional crime, heavy truck traffic, and increased demand for local retail services. The issue has even prompted local municipalities to commission public reports. Boyd et al. (2013) discuss the impacts of the drilling boom on housing markets in five Ohio counties: among their findings are higher rents, pricing marginal renters out of the market and into substandard alternatives. Williamson and Kolb (2011), using interviews from six Pennsylvania counties, find widely heterogeneous responses in rental markets ranging from no change to drastic shortages and price increases upwards of 200% in the span of a few years. They observe that a contemporaneous wave of foreclosures has contributed to an increase in rental demand and that supply has responded slowly.[2] The study mentions only a single company-provided housing facility in Pennsylvania, a solution that is much more common in many western states.[3]

In contrast to these negative consequences from shale gas development, industry sources tend to tout economic benefits. Industry-sponsored studies stress higher employment and income, along with generous estimates of spillover effects including

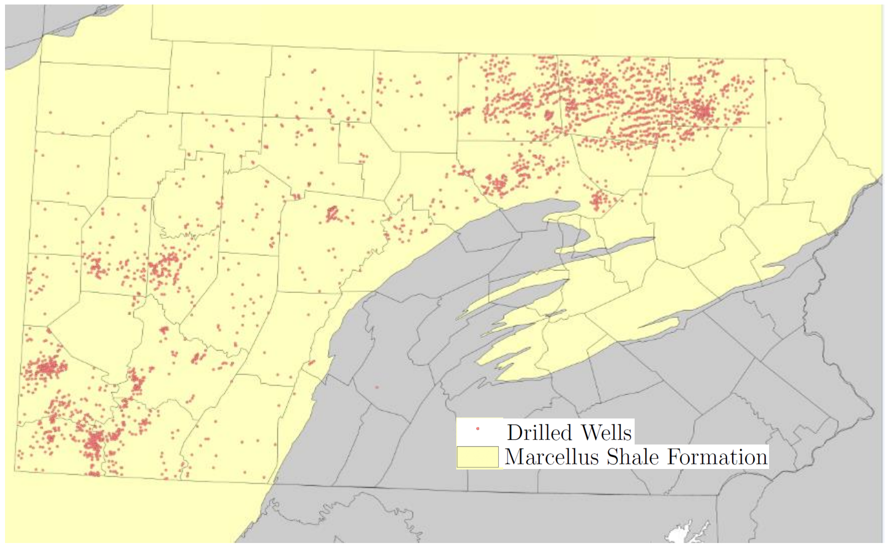
[2]Although some plans had been discussed and implemented, builders face a variety of regulatory, financial and infrastructural hurdles. Their longer investment horizon also requires that the demand shifts are not transitory (Williamson and Kolb, 2011).

[3]This alternative may be preferred by employers due to lower costs, but is not necessarily desirable from the community's standpoint due to the potential for other social issues.

increases in demand for local goods and services (see Kinnaman (2011) for this discussion). Any positive economic effects are channeled primarily through increased employment, and this is likely only in the short run. The process of drilling a new well is much more labor intensive than operating an existing well. Brundage et al. (2011) estimate that the drilling of one new well provides the equivalent of 13.1 full time jobs to 420 people across 150 occupations for six months to a year. However, the labor requirements of operating a drilled well creates an estimated 0.19 full time-equivalent positions. Therefore impacts on rental prices are likely only short-run effects felt while wells are being drilled.

The challenges of municipal planning in the face of the shale boom are exacerbated by the possibility of a subsequent bust. The geology of the Marcellus shale has meant that much of Pennsylvania's activity has been concentrated in the north/northeast region of the state (see Figure 1) – a sparsely populated area before the boom. Zremski (2014) observes a similar phenomenon in nearby Tioga County PA, where the unemployment rate began to increase from its low of 5.8% in October 2011 as drilling activity slowed.

FIGURE 1
Map of Drilled Wells and the Marcellus Formation in Pennsylvania



Prior academic research has examined many of these purported economic effects of the drilling booms. The subjects of investigation include housing values (Muehlenbachs et al., 2014; Gopalakrishnan and Klaiber, 2014), employment effects (Weber, 2012; DeLeire et al., 2014; Marchand, 2012), labor market spillovers (Considine et al., 2011), and the (net) effects on local governmental finance (Raimi and Newell, 2014). Despite numerous anecdotal accounts of prohibitive increases in rents, no careful empirical investigation has yet been presented on the issue.

We use census tract level data from Pennsylvania and New York to identify the effects of drilling on rental prices. The selected New York tracts serve as a particularly appealing control group for the Pennsylvania tracts, as New York's bordering counties are also on the Marcellus shale and are very similar culturally and socio-economically to the counties in northern Pennsylvania. However, during the period of study New York had enacted a moratorium on fracking which kept the boom on the Pennsylvania side of the border (the ban has since been passed as a state law). This regulatory discontinuity neatly splits the tracts under consideration into treatment and control groups.[4] The exogenous variation in drilling further allows for a causal interpretation of the results.

Our results illustrate that reports of large spikes in housing rents due to fracking by popular media sources are not generalizable. The estimated effect on rents of drilling proximate wells is statistically significant and on the order of a $100 per month increase - a result that is robust across multiple specifications. Quantile regression techniques produce some evidence of significantly heterogeneous effects. Even at the highest quantiles however, estimated effects of $250+ per month fall short of the sensational reports outlined above, suggesting that the reports are outliers. The results are relevant for local policy makers seeking to understand and minimize the disruptive impact of drilling. The expected impact of drilling proximate wells falls significantly short of the most alarming anecdotes, and is even lower at lower conditional rent quantiles (interpreted generally as "loose" rental markets).

The paper proceeds in the following manner. Section II surveys the relevant prior literature on resource booms and fracking. Section III describes the data that are used in the study. Section IV describes the empirical strategy and presents results. Section V concludes and discusses potential topics for future research.

II Related Literature

This paper builds primarily upon prior empirical analyses of natural gas and similar resource booms. Many landmark studies have sought to quantify income or employment spillovers or to understand the effect of a resource boom on local economic growth. Recent papers have examined real estate transactions and employed hedonic techniques to quantify the (dis)amenity of drilling. This paper makes two contributions to the literature. The first is an empirical analysis focused exclusively on rental markets. The second contribution is econometric: we combine quantile regression and matching techniques in an extension of prior applications to panel data. The results allow us to identify significant heterogeneity in the effects of fracking on rents.

The economics of local natural resources booms have been studied since Corden and

[4] Our analysis depends on the stable unit treatment value assumption (SUTVA), i.e. that cross-border spillovers do not exist. The dataset used in estimation will reflect this assumption, with the closest New York tracts being discarded.

Neary (1982). A large literature that includes Carrington (1996), Black et al. (2005), Weber (2012), and Marchand (2012) has considered the effect of booms on local employment and incomes. All find evidence of modest spillovers, many of which are transitory.[5] Carrington (1996) finds that the interelasticity of labor across sectors is much lower than the elasticity of labor supply in the energy sector. Similarly, Black et al. (2005) find small positive spillover effects on sectors that produce locally traded goods (e.g. services), but note that differences in poverty rates disappear after the coal bust. The modest spillovers found by Marchand (2012) and Weber (2012) call into question the assumptions of the industry-sponsored studies.[6]

Three recent studies have looked at the effect of the current fracking boom on the housing market in Pennsylvania. Using a hedonic framework, Muehlenbachs, Spiller and Timmins (2014) find zero or slightly positive impacts of wells on property prices at a close geographical range. The effect becomes large and negative when the home's water is sourced directly from the ground. Gopalakrishnan and Klaiber (2014) find smaller negative results, using data from only Washington County, Pennsylvania. The negative effects contradict the prediction of Corden and Neary (1982)'s model, and suggest that residents' housing prices are capitalizing disamenities that arise from fracking. Both papers exploit geo-coded real estate transactions data that enable the authors to calculate the distance of individual houses from wells. Taken together, the results suggest that the housing market's response to drilling is not straightforward and merits further investigation.

James and James (2014) reinforce this conclusion. A micro-level hedonic analysis for Weld County, Colorado finds significant decreases in home value associated with proximity to wells. Using county-level data for the entire U.S., they find that median house prices decreased more slowly in shale counties than non-shale counties. However using the same data they also find that median rental rates and income increased more quickly in the same counties. The direct consideration of rental rate effects is unique in the budding literature, though it is ancillary in the context of the study on house

[5] See also DeLeire et al. (2014) which studies the same question in the context of this paper, and develops a parallel quantile-matching technique.

[6] Indeed, Kinnaman (2011) examines the assumptions underlying Considine et al. (2010, 2011) and CBER (2008), and notes the studies' links to industry. Among other things, the two Considine et al. papers assume that all royalty windfalls are spent in-state in the year received and that 95% of industry expenditures are spent in PA.

prices. The authors estimate that a county's location on shale increases the rental growth rate by 0.2% to 0.4% annually - a much smaller figure than reported by the press.

This paper makes two contributions to the literature on shale gas focusing specifically on rents. First, our data are available at the census tract level, which represents a significant improvement over James and James' county-level study of rents. Unfortunately, sources of rental data are not as rich as those for real estate transactions, so we cannot match the detail achieved by Muehlenbachs et al. (2014) and Gopalakrishnan and Klaiber (2014). Instead, we use the median of the within-tract distribution of rents. Moreover, whereas James and James (2014) estimate an "intent to treat" estimator using the presence of shale within a county, we estimate effects at the census tract level using broadly defined proximity to shale gas development. Our second contribution is an econometric extension: the application of a fixed-effect quantile regression estimator to the matching framework. This approach allows us to consider the effect of fracking on rents at many points along the conditional distribution and identify heterogeneity in effects that will be relevant to local policymakers.

III Data

In order to investigate the effects of drilled wells on rental markets, data were drawn from two primary sources. Rents and other demographic variables are drawn from the U.S. Census' American Community Survey (ACS), while data on fracking and well drilling come from a proprietary dataset compiled for Muehlenbachs et al. (2014).

III.I Sources

The dataset on drilling and well characteristics is that analyzed in Muehlenbachs et al. (2014). The two primary data sources are the Pennsylvania Department of Environmental Protection (PADEP) and the Department of Conservation and Natural Resources (DCNR) Well Information System (the Pennsylvania Internet Record Imaging System/Wells Information System [PA*IRIS/WIS]). The compiled dataset contains information on a number of potentially important features about the wells: the date the permit was issued (for both drilled and undrilled wells), the date the well was drilled, the corporate operator, whether the well is horizontal or vertical, and some information about the well's subsequent productivity. The richness of this data allows for us to consider multiple ways in which fracking might impact the rental market. Further details on the drilling dataset and its creation can be found in Muehlenbachs et al. (2014).

Given the richness of the well dataset and the local nature of housing markets, we study the effect of drilling at the census tract level. Our source of data on rents at the census tract level comes from the Census' ACS dataset, which is released at the census tract level annually in five year moving averages. As the dependent variable of interest, we use the median rent variable from the 2008-2012 wave of ACS data. We then create a panel dataset using similar data from the 2000 Census, deflating rents by the CPI's housing index (2008-2012 rents are deflated by the five-year average of the index). While *de minimus* drilling was present in Pennsylvania prior to 2000, no wells were drilled in 2000. Thus for the panel dataset, all of the well variables are assumed to equal 0 for the year 2000.

III.II Sample Selection

To match the timing of the ACS data, we consider wells that were drilled in Pennsylvania from the beginning of 2008 to the end of 2012.[7] Because workers for a well might endure a significant commute (Boyd et al., 2013), we associate a well with a census tract using the conservative buffer of 50km (31 miles) around each tract's centroid.

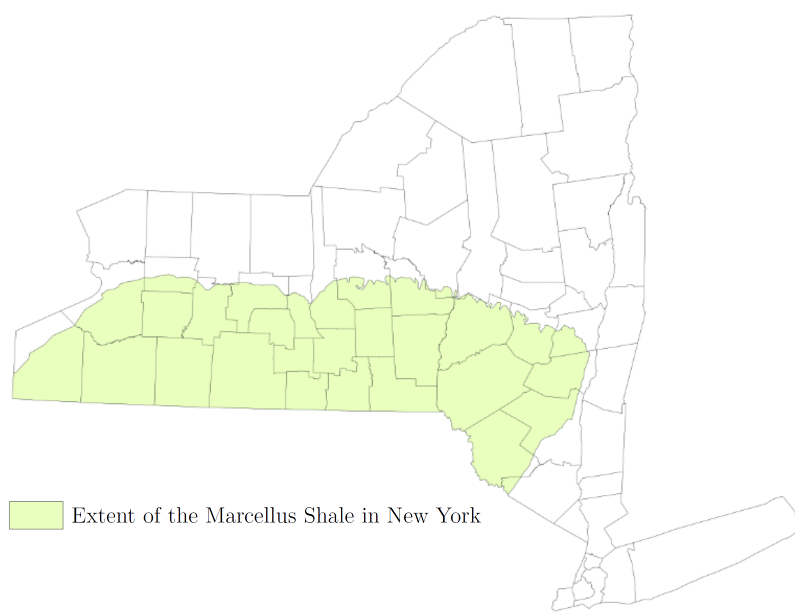
Once we have associated the wells with each census tract, we ignore the timing of drilling and treat wells drilled at any point during the period 2008-2012 homogeneously.[8] Ideally a time series of rental data would allow us to identify both short- and long-term treatment effects of wells, but as we observe only a five-year moving average of rents we are forced to treat a well drilled in any of the five years equally. Because wells drilled might lead to only a short-term spike in rents, this pooling of data may lead to attenuation bias. Our results thus serve as a lower bound of the true effect of drilling activity on concurrent housing rents.

[7] These 5 years include 98% of all Pennsylvania wells drilled from 2000 to 2012.

[8] Measures of the "intensity" of drilling, i.e. the temporal concentration of wells drilled, were created as well. Ultimately these measures did not have much explanatory power and were dropped from the analysis. Results using these measures are available upon request.

The socio-economics of Pennsylvania vary widely. Due to the geology of the shale resources, most drilling takes place in rural areas where market rents are shaped by very different forces than those in the urban centers and suburbs of Philadelphia and Pittsburgh. In order to limit the confounding effects of unobserved heterogeneity, our analysis will be geographically focused: we consider only tracts from northeastern Pennsylvania or the bottom two tiers of New York counties (see Figure 2). The treatment group is composed of northeastern Pennsylvania tracts whose centroid lies within 50km of a spudded well. To prevent contamination between our treatment and control groups, the control group is made up of New York tracts from the bottom two tiers that are at least 50km removed from a spudded well in Pennsylvania (this rules out commutes from across the state border). This exclusion reduces the probability of violating the stable unit treatment value assumption. At the end of the process, we are left with 336 "treated" census tracts in Pennsylvania, and 126 "control" tracts in New York.

FIGURE 2
Marcellus Shale Formation in NY



III.III Summary Statistics

Table 1 presents means and standard deviations of select variables for the treatment and control groups for the years 2000 (columns 1 and 2) and 2008-2012 (columns 4 and 5). It is easy to see that across most observable dimensions, the New York census tracts serve as a reasonable control group to the Pennsylvania tracts where drilling takes place; i.e., the chief difference between the two regions is the moratorium that prevents drilling in New York. The p-values from t-tests for equality of means between the two samples are included in columns 3 and 6, and reveal that the New York tracts are significantly more educated, younger, more urban, and larger. Tests from the baseline period demonstrate that New York tracts were also historically better educated, younger, more urban, and larger than the Pennsylvania tracts.

In a difference-in-differences or fixed effect framework, it is the the differences in the variables that are most relevant. These are shown in the 7th and 8th columns of Table 1, and p-values from t-tests for equality are shown in the final column. From this it can be seen that differences in three relevant covariates are significant at the 5% level: New York census tracts get older, gain more households, and gain a smaller proportion of rental units over time. The bias from aging is unclear, but the more households and fewer rental properties would be associated with higher rents *ceteris paribus*. This suggests that observable differences between changes in our treatment and control sample are likely to bias our estimates of the effect of shale gas development on rents in Pennsylvania downward.

TABLE 1
SUMMARY STATISTICS

	2000 Means			2008-2012 Means			Differences		
	PA	NY	p	PA	NY	p	PA	NY	p
Wells spudded	0	0		260.3	0		260.3	0	
Median Rent ($2010)	465.51	524.20	0.00	678.32	699.40	0.17	212.72	169.72	0.00
Percapita income	17,035.1	17,895.8	0.13	23,894.0	24,259.1	0.62	6,882.6	6,445.7	0.50
Median age	39.29	35.98	0.00	41.59	40.02	0.02	2.315	3.843	0.00
White collar employment	69.78	72.60	0.01	73.08	75.32	0.03	3.347	3.194	0.78
Unemployment	5.881	6.573	0.18	7.929	7.306	0.20	2.017	0.842	0.05
Poverty rate	12.69	14.09	0.25	10.73	9.544	0.17	-2.029	-3.815	0.09
Education - less than HS	18.57	16.04	0.00	12.50	11.34	0.06	-6.020	-5.164	0.14
Education - HS	42.13	33.54	0.00	41.72	33.05	0.00	-0.43	-0.96	0.35
Education - some college	21.26	26.66	0.00	24.67	28.73	0.00	3.38	2.52	0.09
Education - Bachelor degree	18.04	23.76	0.00	21.11	26.88	0.00	3.06	3.60	0.35
Urban population	0.636	0.449	0.00	0.642	0.438	0.00	0.0078	0.0014	0.51
Total population	3,695.8	4,063.5	0.02	3,790.6	4,084.9	0.06	73.55	33.14	0.61
Number of households	1,462.1	1,493.5	0.57	1,505.1	1,590.6	0.18	34.36	86.13	0.01
% of rental units	29.66	30.63	0.62	31.71	30.14	0.43	1.98	-0.30	0.00
Observations	336	126		336	126		336	126	

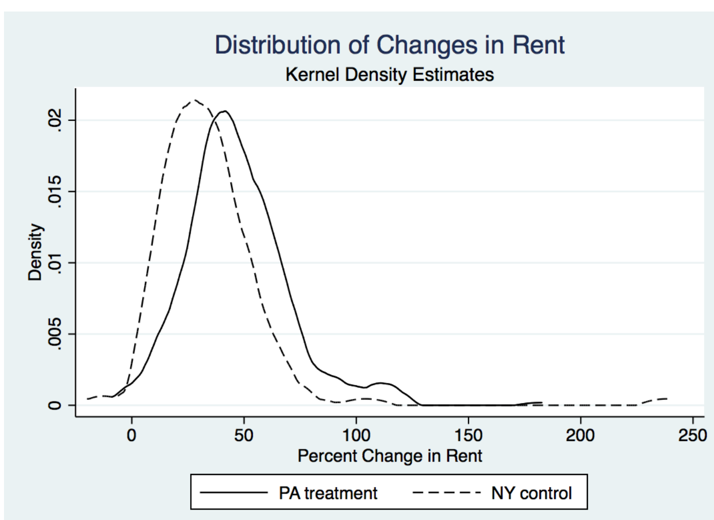
The year 2000 data can also serve as a baseline to consider how rents have changed in different census tracts since the new drilling boom. Table 2 shows some details about the percentage change in rents between 2000 and 2008-2012 for the census tracts of interest, after deflating 2008-2012 rents by the housing component of the CPI. While there is a sizable gap in the unconditional means, the gaps appear much larger at the higher percentiles. This feature is also observable in Figure 3. A shift between the two distributions is quite noticeable, and appears to be bigger at the highest quantiles. These preliminary findings motivate the use of quantile regression techniques in unmasking the heterogeneous effects of drilling.

TABLE 2
PERCENT CHANGE IN RENT: 2000 TO 2008-2012

	Statistic						
	Mean	Sd	P25	P50	P75	P95	P99
PA treatment (n=336)	46.9	23.4	32.9	43.5	58.4	91.2	116.5
NY control (n=126)	34.4	26.1	19.3	31.2	44.1	67.2	104.1
Total (n=462)	43.6	24.7	28.7	40.8	55	85.1	116.5

Source: Census Data

FIGURE 3
KERNEL DENSITY ESTIMATES OF THE DISTRIBUTION OF RENT CHANGES



IV Empirical Strategy and Results

In this section we present our model of rent determination, the accompanying estimation routine and results. We first estimate a mean effect, using the matching difference-in-differences estimator from Heckman et al. (1998). Then we deepen the analysis by adding quantile regression techniques. To do so, we extend the quantile panel estimator developed in Chen and Khan (2008) to the context of matching. Results demonstrate significant variability in effects along the distribution of the unobservable error term.

IV.I Matching

Matching estimation is a useful and flexible technique for evaluating average treatment effects when only nonexperimental data is available. Intuitively, matching works by directly comparing a treated units' outcome to a similar-looking control units' outcome. However, the technique can also be applied in panel data contexts leading to a difference-in-differences type matching estimator. Asymptotic theory for the difference-in-differences matching estimator was developed by Heckman et al. (1997, 1998), and has been employed in various contexts by Galiani et al. (2005) and Gertler et al. (2004) among others.

In general, matching methods feature a distinct advantage: by avoiding the need to specify a functional form for the outcome variable, misspecification issues do not arise. The theoretical framework has been outlined in Rubin (1974) and Holland (1986). This framework posits that each individual *a priori* has two potential outcomes, one if treated (R^T) and one if untreated (R^C). If z is a dummy variable indicating treatment, then the outcome variable observed by the econometrician is precisely $R \equiv zR^T + (1 - z)R^C$. The counterfactual outcome is never observed, but the economic variable of interest is the treatment effect, $\Theta \equiv R^T - R^C$.

The available data readily allows for the recovery of the conditional distributions $F(R^T|X, z = 1)$ and $F(R^C|X, z = 0)$. However, the joint distributions and the distribution of the treatment effect are not as easily estimated. Consequently, developed estimation techniques generally focus on recovering the *mean effect of treatment on the treated*, $E[R^T - R^C|z = 1]$. This value provides the expected treatment effect for any program participant and is usually the parameter of interest in evaluating treatment programs.

For matching to provide valid inference, two key assumptions must be met. The first is that the two potential outcomes are independent of participation status conditional on observed characteristics X.

Assumption IV.1. $(R^C, R^T) \perp\!\!\!\perp z|X$

Assumption IV.1 can be re-stated as either $\Pr(z = 1|R^C, R^T, X) = \Pr(z = 1|X)$ or $E[z|R^C, R^T, X] = E[z|X]$. In words, neither the observed nor the counterfactual outcome reveal any information about the probability of treatment, conditional on X. This assumption was termed "strict ignorability" of the treatment assignment by Rosenbaum and Rubin (1983).

Furthermore, it is generally assumed that for every value of X the probability of treatment is neither 0 nor 1.

Assumption IV.2. $0 < Pr(z = 1|X) < 1 \quad \forall X$

Assumption IV.2 guarantees that appropriate matches can be found for all treated and untreated observations.[9] Even if Assumption IV.2 is not strictly met in the data, matching can still be an appropriate estimation method on the region of common support, $\mathcal{S}_P$. Of course in this case, interpretation is altered as the econometrician has estimated the average treatment effect, *conditionally on the region of overlapping propensity for treatment.*

Matching is further complicated when the dimensionality of X is high. "Small cell" problems might exist for discrete X, and continuous X are plagued by slow rates of convergence due to the curse of dimensionality. Fortunately Rosenbaum and Rubin (1983) demonstrated that when R^C outcomes are independent of z conditional on X, they are also independent of z conditional on $P(X) \equiv \Pr(z = 1|X)$. In other words, matching on the propensity score is a valid approach. Using the propensity

[9] It should be noted that if the only parameter of interest is the mean treatment effect on the treated, then these assumptions are stronger than necessary. Specifically, Assumption IV.1 can be replaced with a condition on only R^C: $E[R^C|X, z = 1] = E[R^C|X, z = 0] = E[R^C|X]$. Likewise, Assumption IV.2 can be relaxed as the econometrician only needs to guarantee that a non-treatment observation can be found for each treated observation. This only requires that $\Pr(z = 1|X) < 1 \quad \forall X$. See Heckman et al. (1998).

score reduces the matching problem to a single dimension, but splits the matching estimation into two stages. The first stage involves estimating the propensity score $P(X)$ with a binary discrete choice model; typically this is a fully parametric logit or probit model, but semiparametric or nonparametric methods might also be used. In the second stage, treated and control observations are matched together based on their estimated propensity scores.

Matching estimators form their matches in a variety of ways: nearest neighbor, caliper, stratification/interval, kernel/local linear, etc. For example, in the latter case observations are matched to a weighted average of several other observations, with weights determined nonparametrically. Furthermore, matching may be performed either with or without replacement, i.e. a single control observation might be used more than once. A significant disadvantage of matching without replacement is that results might be altered by different orderings of the data. For simplicity and due to the fairly small sample size, this paper will use a simple nearest neighbor matching strategy with replacement. In this framework we find single matches so the matching weights are $W(i,j) \in \{0,1\}$ with $\sum_j W(i,j) = 1 \; \forall i$. The match is determined by the closest propensity score from the other group, conditional on both propensity scores falling within the $\mathcal{S}_P$. Defining $\mathcal{T}$ and $\mathcal{C}$ to be the treatment and control groups, the match for observation $i \in \mathcal{T}$ is found by:

$$W(i,h) = 1 \iff h = \arg\min_j |P_i - P_j|, \; j \in \mathcal{C}, \; P_i, P_j \in \mathcal{S}_P.$$

When applying matching to panel data and the difference-in-differences framework, the assumptions above must be modified slightly; for two time periods t and t' the estimator now requires that the expected (realized or potential) non-treated difference over time be the same regardless of treatment.

Assumption IV.3. $E[R^C_{t'} - R^C_t | P, z = 1] = E[R^C_{t'} - R^C_t | P, z = 0]$

The general form of the difference-in-differences matching estimator is given by:

$$\hat{\delta}_{GM} = \frac{1}{n}\sum_{i\in\mathcal{T}\cap\mathcal{S}_P}\left[(R^T_{t'i} - R^C_{ti}) - \sum_{j\in\mathcal{C}\cap\mathcal{S}_P} W(i,j)(R^C_{t'j} - R^C_{tj})\right] + \frac{1}{n}\sum_{i\in\mathcal{C}\cap\mathcal{S}_P}\left[\sum_{j\in\mathcal{T}\cap\mathcal{S}_P} W(i,j)(R^T_{t'j} - R^C_{tj}) - (R^C_{t'i} - R^C_{ti})\right]. \quad (1)$$

In the estimation below, we consider a straightforward nearest neighbor propensity score matching routine. In order to ensure common support, (i.e. satisfy assumption IV.2) we discard treated observations whose propensity is above the 99.5^{th} quantile of the distribution of control observations, and control observations whose propensity score is below the 0.05^{th} percentile of the treated observations propensity score. With the single nearest neighbor matching routine, equation 1 simplifies to:

$$\hat{\delta}_{NNM} = \frac{1}{n}\sum_{i\in\mathcal{T}\cap\mathcal{S}_P}\left[(R^T_{t'i} - R^C_{ti}) - (R^C_{t'j} - R^C_{tj})\right] + \frac{1}{n}\sum_{i\in\mathcal{C}\cap\mathcal{S}_P}\left[(R^T_{t'j} - R^C_{tj}) - (R^C_{t'i} - R^C_{ti})\right], \quad (2)$$

where $j \in \mathcal{C}$ is the appropriately matched control observation for each observation $i \in \mathcal{T}$.

IV.I.1 Matching and Bootstrapping

Standard errors are a bit more difficult to calculate when using a matching estimator. Abadie and Imbens (2008) point out that a typical bootstrap is poorly motivated in the matching case and might lead to either the over- or underestimation of standard errors. As an alternative, they suggest a subsampling bootstrap such as that developed in Politis and Romano (1994). The subsampling routine requires minimal assumptions in order to achieve consistency. In particular, in contrast to the traditional bootstrap, it only requires that the statistic being estimated has a limiting distribution and relaxes the assumption that the distribution be smooth.[10] The subsampling routine is implemented by drawing fewer observations than the sample

[10]Matching estimators tend to have highly non-smooth distributions because they are functions of distributions of the data.

size at each iteration, and drawing without replacement. Politis et al. (1999) discuss the routine in detail in a number of contexts, and outline a data-driven approach to selecting the subsample size.

Furthermore, Politis et al. (1999) describe how to calculate bias-reduced estimates. The empirical bias of a parameter estimate $\hat{\gamma}$ is given by:

$$bias(\hat{\gamma}) = \sqrt{\frac{b}{n}} \left(\bar{\hat{\gamma}}^* - \hat{\gamma} \right),$$

where b is the size of the subsample at each iteration, n is the full sample size, $\bar{\hat{\gamma}}^*$ is the mean of the subsampled point estimates, and $\hat{\gamma}$ is the estimate using the full dataset. Below, we present only bias-corrected estimates: $\hat{\gamma}_{BC} \equiv \hat{\gamma} - bias(\hat{\gamma})$.

IV.I.2 Matching Results

Table 3 presents the results from estimating equation (2).[11] The treatment and control groups are formed as outlined in Section III, and so the binary treatment is defined by having at least one well drilled within 50km.

Observations are matched using propensity scores generated by a probit model linear in tract income, age, education, percentage white-collar employment, poverty rate, population, number of households, rental vacancy and the proportion of the housing market made up of rentals. All of these variables for matching are taken only from the baseline data from the year 2000. Thus matching occurs only on the basis of pre-treatment observables and is not confounded by changes that are simultaneous with the introduction of drilling.

The method estimates a statistically significant treatment effect on the treated of an \$89 increase in monthly rents. This \$89 represents a 13.07% increase in rents at the sample mean of \$684.03, or a 13.19% increase at the treatment group mean of \$678.32. One advantage of this method is its non-parametric nature and lack of sensitivity to functional form.

The estimated effect is much smaller than might be expected from reading popular

[11] The discrepancy in sample size from earlier is due to the imposition of the common support restriction.

TABLE 3
Matching DID Results

	(1)
	Log rent
Wells indicator	89.42***
	[28.40]
Observations	458

Standard errors in brackets
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

media sources. This motivates an investigation into heterogeneous effects along the distribution of the error term. The following section outlines an estimation technique that combines matching and quantile regression before presenting results.

IV.II Quantile Regression

The impact of drilled wells on a census tracts' median rent might not be the same at the mean as at other points of the rent distribution. Indeed, that would only be true if drilling represents a "common effect" or "location shift" (Abadie et al., 2002; Heckman et al., 1997). The varied conclusions of the public studies, combined with the disagreement between anecdotes and preliminary statistical evidence indicate that the common effect assumption may be violated.

Following Koenker and Bassett (1978) and Koenker and Hallock (2001), we consider a model of rent determination featuring linear heteroskedasticity:[12]

$$R_i = \theta z_i + X_i'\beta + (\xi_1 z_i + X_i'\xi_2)\,\epsilon_i. \tag{3}$$

We use θ and z_i to denote the coefficient of interest and an indicator for treatment respectively, and assume that ϵ_i is independent from z_i and X_i' and conditionally mean zero. It is easy to see that within this setup, marginal effects can vary by quantile. Letting ρ_ϕ denote the ϕth quantile of ϵ_i, one can see that the ϕth conditional quantile

[12]For now, the fixed effect and time subscripts are omitted for exposition.

of R_i, q_ϕ, is given by:

$$\begin{aligned} q_\phi &= \theta z_i + X_i'\beta + (\xi_1 z_i + X_i'\xi_2)\,\rho_\phi \\ &= (\theta + \xi_1\rho_\phi)z_i + X_i'(\beta + \xi_2\rho_\phi) \\ &\equiv \theta_\phi z_i + X_i'\beta_\phi. \end{aligned}$$

Defining $\mu_\epsilon \equiv E[\epsilon_i](= 0)$, the conditional mean of R_i is accordingly:

$$\begin{aligned} q_\mu &= \theta z_i + X_i'\beta + (\xi_1 z_i + X_i'\xi_2)\,\mu_\epsilon \\ &= (\theta + \xi_1\mu_\epsilon)z_i + X_i'(\beta + \xi_2\mu_\epsilon) \\ &\equiv \theta_\mu z_i + X_i'\beta_\mu \quad \Big(= \theta z_i + X_i'\beta\Big). \end{aligned}$$

The difference between the two measures is:

$$\begin{aligned} q_\phi - q_\mu &= \theta_\phi z_i + X_i'\beta_\phi - \theta_\mu z_i + X_i'\beta_\mu \\ &= \xi_1(\rho_\phi - \mu_\epsilon)z_i + X_i'\Big(\xi_2(\rho_\phi - \mu_\epsilon)\Big), \end{aligned}$$

and so depends on the shape of ϵ_i's distribution. For example, if the distribution is skewed to the left (and ξ_1, ξ_2 are positive), then the marginal effect retrieved from mean estimation will understate the true marginal effect at the median and higher quantiles. Likewise, if ϵ_i's distribution is right-skewed, mean estimates will overstate the true marginal effect at low quantiles.

The use of quantile techniques allows the econometrician to control in some sense for unobserved heterogeneity. Heterogeneous marginal effects are recoverable at different points along the conditional quantile distribution.[13] This property is highly useful in the present context: although the census dataset provides us with a huge number of observable variables, many determinants of the marginal effects of drilling wells on rents remain unobserved. For example, the dataset does not include information about the recent development history of the census tract, the labor, capital and financial capacity for timely new construction, regulatory or infrastructural hurdles to construction, the presence of temporary housing alternatives such as hotels, worker preferences driven by word of mouth, or the decision of a firm involved in the

[13] Furthermore, the distribution of ϵ_i can be identified up to a scale and location normalization.

drilling to locate a "base of operations" nearby. Our estimates will allow us to predict marginal effects at different points of the quantile distribution of these unobservables, conditional on some observable controls.

IV.III Quantile Regression and Matching

The application of quantile techniques to the matching context is not straightforward. A threat to consistency arises from the fact that the quantile and difference operators are not interchangeable. However, the method developed by Chen and Khan (2008) in an application to panel data is easily extended to the matching context.

Recall our model of interest:

$$R_i = \alpha_i + \theta z_i + X_i'\beta + (\xi_1 z_i + X_i'\xi_2)\,\epsilon_i,$$

and observe that the ϕth conditional quantile has the following form:

$$\begin{aligned} q_\phi &= \alpha_i + \theta z_i + X_i'\beta + (\xi_1 z_i + X_i'\xi_2)\,\rho_\phi && (4)\\ &= \alpha_i + \theta_\phi z_i + X_i'\beta_\phi && (5) \end{aligned}$$

We are interested in estimating the treatment parameter at the ϕth quantile, θ_ϕ. If counterfactual data were observable, this quantity might easily be calculated by differencing the treatment and control median rents at the quantile of interest:

$$\begin{aligned} q_\phi^T - q_\phi^C &= (\alpha_i + \theta_\phi z_i^T + X_i'^T\beta_\phi) - (\alpha_i + \theta_\phi z_i^C + X_i'^C\beta_\phi)\\ &= \theta_\phi(z_i^T - z_i^C) + (X_i'^T - X_i'^C)\beta_\phi\\ &= \theta_\phi. \end{aligned}$$

Of course, this is infeasible because counterfactual data does not exist. We proceed in the spirit of matching, by substituting similar-looking control observations for the nonexistent counterfactual data for treated tracts, and vice versa. Labeling treatment and control groups as $\mathcal{T}$ and $\mathcal{C}$ respectively, we match observations using the propensity score matching function $m(\cdot)\colon \mathcal{K} \mapsto \mathcal{K}',\ \ \mathcal{K},\mathcal{K}' \in \{\mathcal{T},\mathcal{C}\},\ \mathcal{K} \neq \mathcal{K}'$ over a region of common support $\mathcal{S}_P$ on *a priori* observables.

Following the idea of Chen and Khan (2008), we then impose non-parametric structure on the fixed effects:

$$\alpha_i = \alpha_j = f(X_i, X_j) \quad j \in m(i).$$

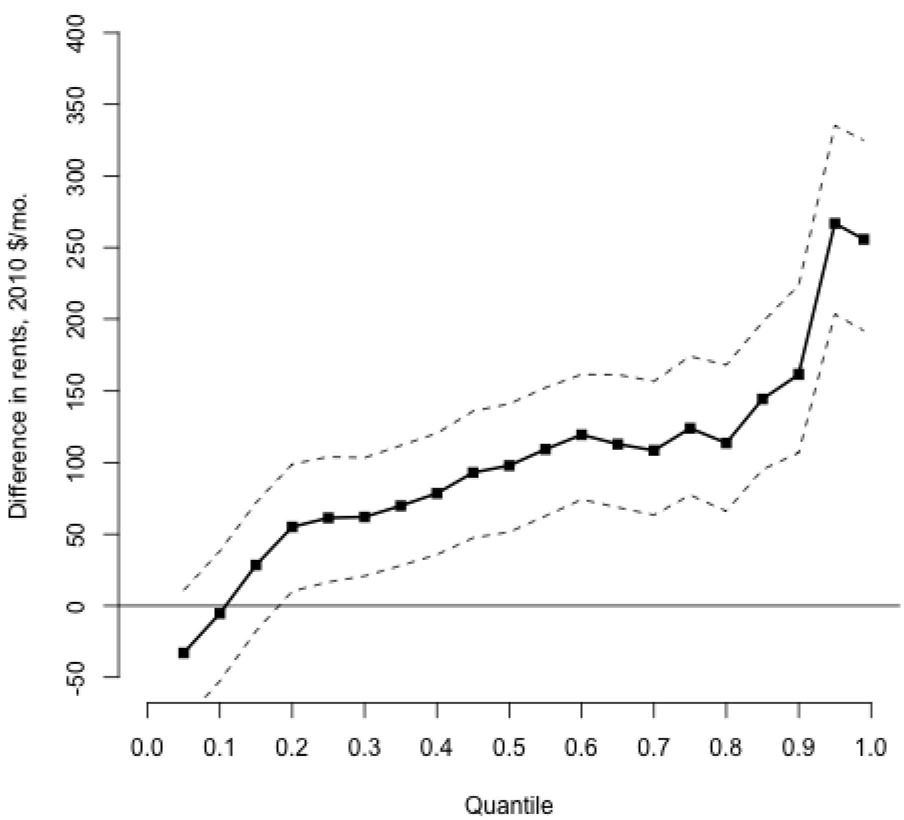
Specifically, we assume that treated census tract i and control census tract j, matched through propensity estimation function $p(\cdot)$ share a common unobserved effect. Further, this effect is an arbitrary function of *both tracts'* observable characteristics, making the method quite general. The structure imposed on the shared effect mirrors that assumed in Chen and Khan (2008) exactly. Paired with the assumption that $\rho^T_\phi = \rho^C_\phi$, a natural estimator of θ_ϕ suggests itself. Assuming for now that $i \in \mathcal{T}$, we have the following:

$$q^T_\phi - q^C_\phi = \theta_\phi z_i + (X'_i - X'_j)\beta_\phi, \quad j \in m(i). \tag{6}$$

Estimation then proceeds with the same two-step procedure outlined in Gamper-Rabindran et al. (2010). In the first stage, we perform traditional quantile regression and calculate fitted values $\hat{q}^T_\phi$ and $\hat{q}^C_\phi$, using data from both the treated observation i and the control observation j. In the second stage, their difference is calculated and regressed on the differenced X variables via OLS. Standard errors are again calculated with the subsampling bootstrap routine with 1,000 iterations.

In practice, we use the same propensity score matching function and binary treatment variable as above. The coefficients of interest are presented below in Figure 4, and the estimated coefficients on other control variables are included in Figures 5 through 7. Estimated effects at various quantiles are depicted (points connected by solid line), as well as the empirical bias-corrected 95% confidence interval (dashed lines). The results suggest significant heterogeneity in the effects of being within 50km of spudded wells on rents. At the lowest quantiles, the effect is a statistical zero. From the 20th to the 80th quantiles, the effect increases from around \$50 to \$125 per month. At the top of the distribution, the effect more than doubles to over \$250 per month. The estimate of \$266.82 at the 95th percentile represents a 39.0% increase at the sample mean. The estimated effect still falls short of the sensational numbers reported in the press, but data limitations allow estimates only at the median rent of a given census tract, observed in a five year moving average.[14]

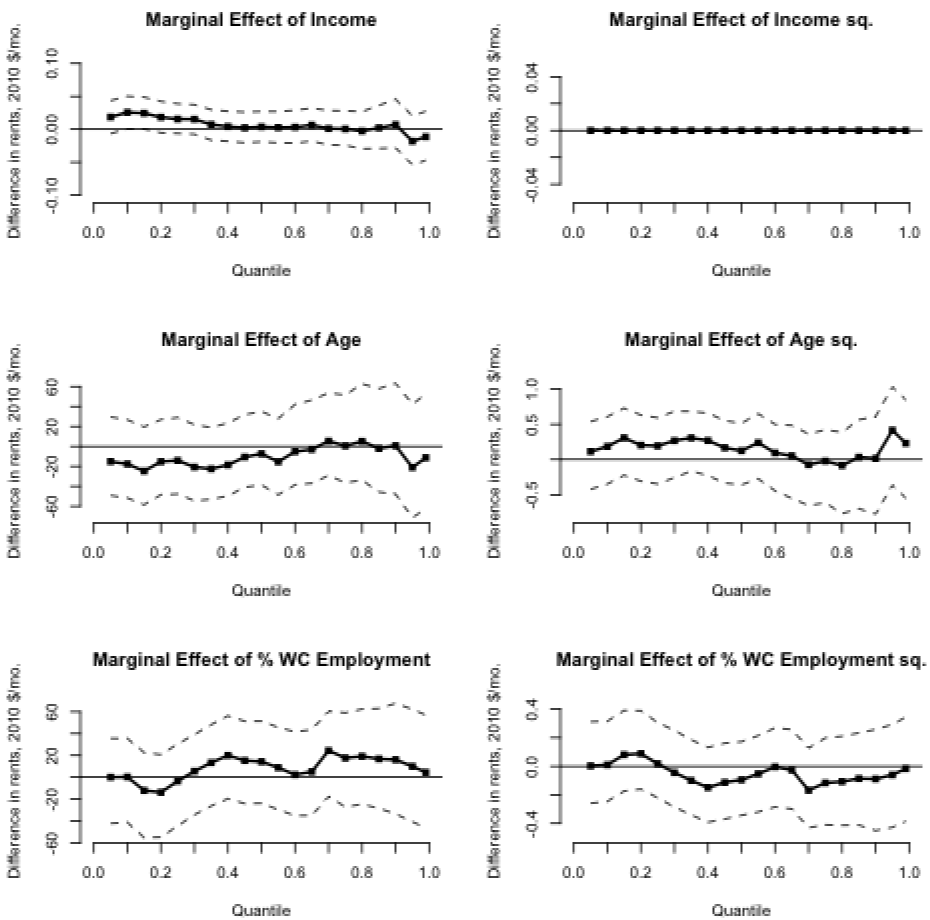
FIGURE 4
Effect of Shale Gas Drilling Activity within 50km at Select Quantiles



Estimated using the approach outlined in Section IV.III. Estimated coefficients at various quantiles are drawn with the solid line. An empirical bias corrected 95% confidence interval is drawn around the estimates in dashed lines.

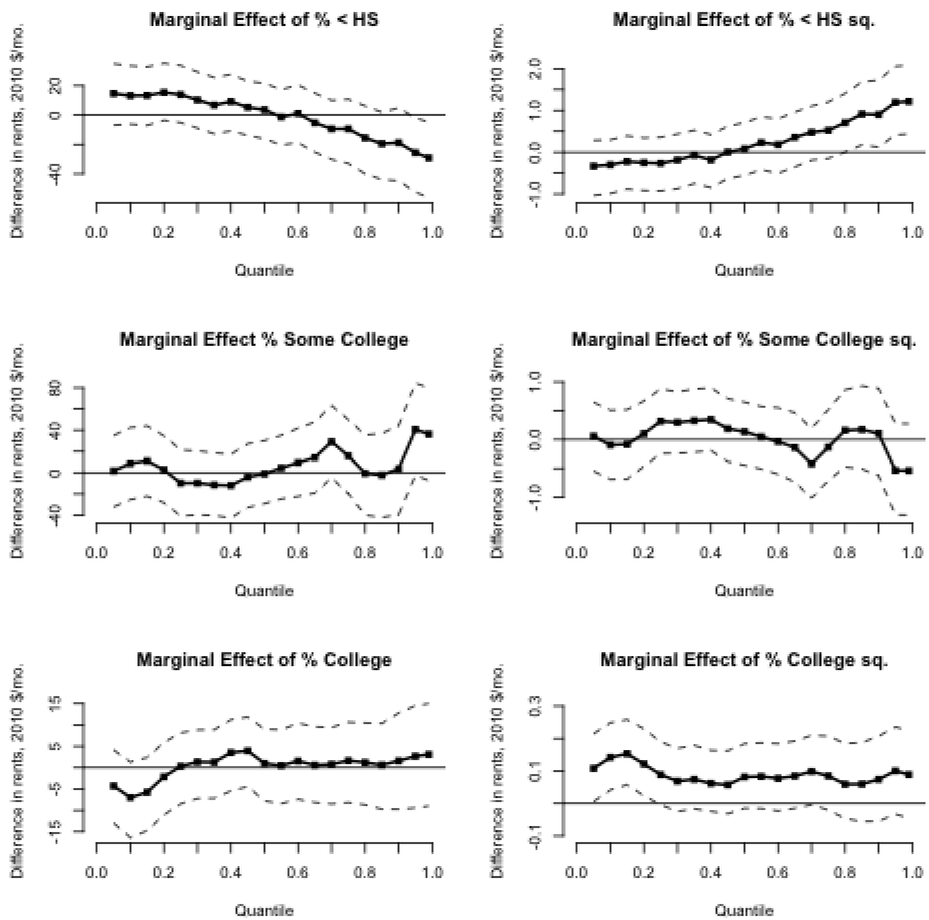
FIGURE 5

Marginal Effect of Control Variables at Select Quantiles



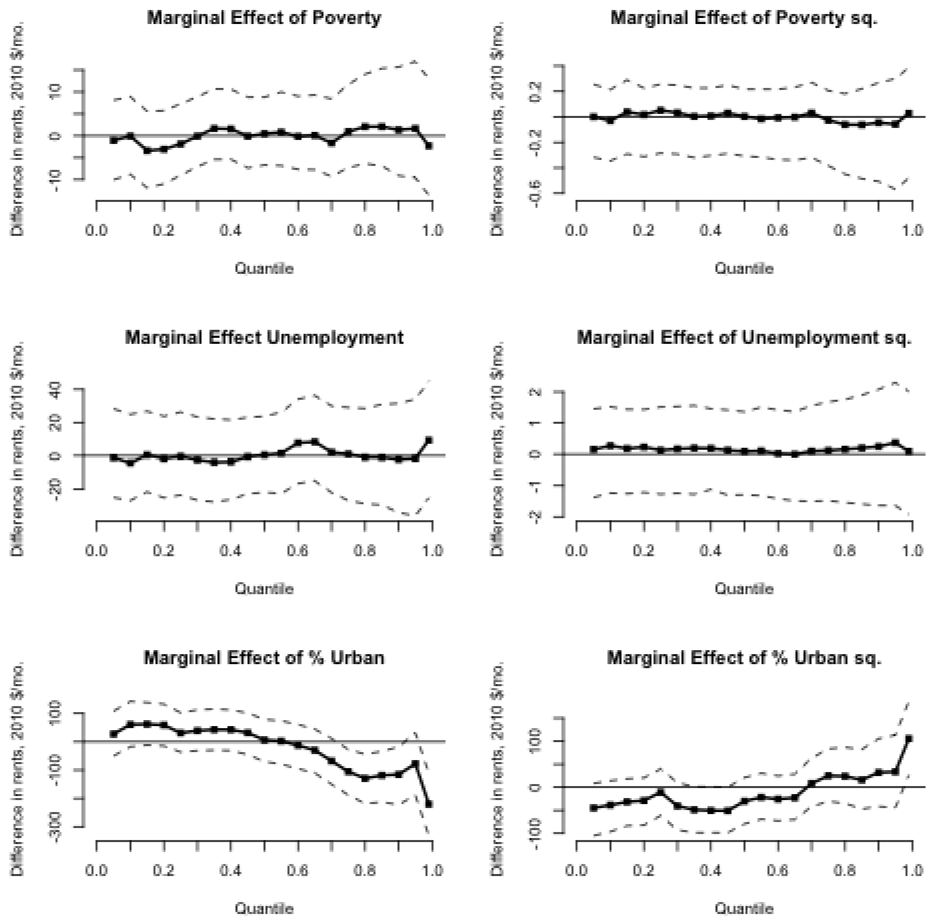
Estimated using the approach outlined in Section IV.III. Estimated coefficients at various quantiles are drawn with the solid line. An empirical bias corrected 95% confidence interval is drawn around the estimates in dashed lines.

FIGURE 6
Marginal Effect of Control Variables at Select Quantiles



Estimated using the approach outlined in Section IV.III. Estimated coefficients at various quantiles are drawn with the solid line. An empirical bias corrected 95% confidence interval is drawn around the estimates in dashed lines.

FIGURE 7
Marginal Effect of Control Variables at Select Quantiles



Estimated using the approach outlined in Section IV.III. Estimated coefficients at various quantiles are drawn with the solid line. An empirical bias corrected 95% confidence interval is drawn around the estimates in dashed lines.

While the use of matching generally reduces the impact of functional form specifications and biases due to misspecification (e.g. see Ho et al. (2007)), this estimator requires functional form assumptions. While the results lose non-parametric flavor, they gain the application of quantile methods in the matching context.

[14]A coarse investigation into how the distribution of rents shifts differentially with the introduction of wells is possible with the band-coded rent data from the Census. No major differences are discernible, supporting the use of median rents as the independent variable.

V Discussion

The recent increase in drilling activity on the Marcellus shale formation has had major impacts on local economies, particularly in the remote area along the Pennsylvania - New York border. Media reports have documented the challenges brought to local towns by the influx of well workers. Much attention has been given to rental housing, a sector where supply cannot quickly rise to meet a jump in demand.

This paper seeks to investigate the effect of fracking on rents empirically. We exploit a regulatory discontinuity - the moratorium on fracking in New York, in contrast to legal fracking in Pennsylvania - for exogenous variation in fracking activity. Initial results illustrate that the mean effect of fracking on rental housing is economically and statistically significant, although far below the sensational impacts that are reported in the news.

We also extend the panel data quantile regression method of Chen and Khan (2008) to the matching context. This results in a quantile estimator that has a difference-in-differences flavor. The results show significant heterogeneity, indicating that the impact of proximate fracking on rental markets varies widely along the distribution of the unobservable error term. While certain features of the data suggest that our estimates may still be a lower bound on the true effect, we find impacts at the top of the distribution more than three times larger than the mean.

This set of conclusions has implications for policy makers in shale gas areas, especially on the Marcellus shale. Estimated effects are generally significant, but for most census tracts they fall far short of the sensational squeezes reported in the press (which may not be describing changes at the median). Furthermore, the increase in rents in the context of falling housing prices (see e.g. Muehlenbachs, Spiller and Timmins, 2014) raises some interesting economic questions. Would homeowners bothered by fracking be better off renting their house instead of selling it? As economic theory predicts that property values should rise with rents, do hedonic studies such as Muehlenbachs et al. (2014) understate the disamenity associated with drilling?

Future research into the issue might proceed in a variety of ways. A similar approach can be taken with alternate specifications. Higher quality data on rents would allow for the investigation of a number of further questions. For example, annual data would enable the comparison of short- and long-term effects, and might reveal the effects found here to be transitory. They might also combat the attenuation that stems from using 5 year moving average data. More detailed data on the within-tract distribution of rents, as opposed to just the median, might provide a more detailed picture of heterogeneous effects. Finally, it would be interesting to study the response of supply to the surging demand - possible sources include company-built housing, homeowners renting out rooms/houses, and developers building additional rental properties. These alternatives would be expected to alleviate pricing pressure on rents, but might be associated with other negative economic consequences, e.g., the "man camps" which have become common in parts of North Dakota. An additional such disamenity combined with lower rents might serve to depress house prices even further. Complete real-estate market data (e.g. including rent and company housing figures) allowing a comparison of responses in different shale regions of the U.S. might reveal if this is the case.

References

Abadie, Alberto and Guido W Imbens, "On the failure of the bootstrap for matching estimators," *Econometrica*, 2008, *76* (6), 1537–1557.

_ **, Joshua Angrist, and Guido Imbens**, "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," *Econometrica*, 2002, *70* (1), 91–117.

Black, Dan, Terra McKinnish, and Seth Sanders, "The Economic Impact Of The Coal Boom And Bust*," *The Economic Journal*, 2005, *115* (503), 449–476.

Boyd, S, R Gorden, K Leeman, C Poole, S Porter, and R Stewart, "The impact of horizontal shale well development on housing," 2013.

Brundage, Tracy L, Timothy W Kelsey, Janice Lobdell, Larry L Michael, Jeffrey Jacquet, James R Ladlee, Jeffrey F Lorson, and Thomas B Murphy, "Pennsylvania Statewide Marcellus Shale Workforce Needs Assessment," *Williamsport, PA: Marcellus Shale Education and Training Center*, 2011.

Carrington, William J, "The Alaskan labor market during the pipeline era," *Journal of Political Economy*, 1996, *104* (1), 186–218.

CBER, "Projecting the Economic Impact of the Fayetteville Shale Play for 2008-2012," *Center for Business and Economic Research*, 2008.

Chen, Songnian and Shakeeb Khan, "Semiparametric estimation of nonstationary censored panel data models with time varying factor loads," *Econometric Theory*, 2008, *24* (05), 1149–1173.

Considine, Timothy J, Robert W Watson, and Nicholas B Considine, "The economic opportunities of shale energy development," *The Manhattan Institute, June*, 2011.

_ **, Robert Watson, and Seth Blumsack**, "The economic impacts of the Pennsylvania Marcellus shale natural gas play: an update," *The Pennsylvania State University, Department of Energy and Mineral Engineering*, 2010.

Corden, W Max and J Peter Neary, "Booming sector and de-industrialisation in a small open economy," *The economic journal*, 1982, pp. 825–848.

DeLeire, Thomas, Paul Eliason, and Chris Timmins, "Measuring the employment impacts of shale gas development," *Working Paper*, 2014.

Detrow, Scott, "Five years into fracking boom, one pa. town at a turning point," *NPR*, 2013, p. January 14.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky, "Water for life: The impact of the privatization of water services on child mortality," *Journal of political economy*, 2005, *113* (1), 83–120.

Gamper-Rabindran, Shanti, Shakeeb Khan, and Christopher Timmins, "The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach," *Journal of Development Economics*, 2010, *92* (2), 188–200.

Gertler, Paul, David I Levine, and Minnie Ames, "Schooling and parental death," *Review of Economics and Statistics*, 2004, *86* (1), 211–225.

Gopalakrishnan, Sathya and H Allen Klaiber, "Is the Shale Energy Boom a Bust for Nearby Residents? Evidence from Housing Values in Pennsylvania," *American Journal of Agricultural Economics*, 2014, *96* (1), 43–66.

Heckman, James J, Hidehiko Ichimura, and Petra Todd, "Matching as an econometric evaluation estimator," *The Review of Economic Studies*, 1998, *65* (2), 261–294.

_ **, Jeffrey Smith, and Nancy Clements**, "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," *The Review of Economic Studies*, 1997, *64* (4), 487–535.

Ho, Daniel E, Kosuke Imai, Gary King, and Elizabeth A Stuart, "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference," *Political analysis*, 2007, *15* (3), 199–236.

Holland, Paul W, "Statistics and causal inference (with discussion)," *Journal of the American statistical Association*, 1986, *81* (396), 945–970.

James, Alexander and Jasmine James, "A canary near a gas well: Gas booms and housing ma rket busts in Colorado," *Working Paper*, 2014.

Kinnaman, Thomas C, "The economic impact of shale gas extraction: A review of existing studies," *Ecological Economics*, 2011, *70* (7), 1243–1249.

Koenker, Roger and Gilbert Bassett, "Regression quantiles," *Econometrica: journal of the Econometric Society*, 1978, pp. 33–50.

_ **and Kevin Hallock**, "Quantile regression: An introduction," *Journal of Economic Perspectives*, 2001, *15* (4), 43–56.

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

Muehlenbachs, Lucija, Elisheba Spiller, and Christopher Timmins, "The Housing Market Impacts of Shale Gas Development," 2014.

of Labor Statistics, Bureau, "Consumer Price Index - Housing," April 2014.

Politis, Dimitris, Joseph Romano, and Michael Wolf, *Subsampling* 1999.

Politis, Dimitris N and Joseph P Romano, "Large sample confidence regions based on subsamples under minimal assumptions," *The Annals of Statistics*, 1994, pp. 2031–2050.

Raimi, Daniel and Richard G Newell, "Shale Public Finance," 2014.

Rosenbaum, Paul R and Donald B Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 1983, *70* (1), 41–55.

Rubin, Donald B, "Estimating causal effects of treatments in randomized and nonrandomized studies.," *Journal of educational Psychology*, 1974, *66* (5), 688.

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

Williamson, Jonathan and Bonita Kolb, "Marcellus natural gas developments effect on housing in Pennsylvania," *Center for the Study of Community and the Economy*, 2011.

Zremski, Jerry, "A Border Tale of Boom and Bust," *Buffalo News*, 2014, (May 11).