

2017

The Cost of Unconventional Gas Extraction: A Hedonic Analysis

Michael S. Delgado

Todd Guilfoos

Andrew Boslett

Follow this and additional works at: https://digitalcommons.uri.edu/enre_facpubs

**The University of Rhode Island Faculty have made this article openly available.
Please let us know how Open Access to this research benefits you.**

This is a pre-publication author manuscript of the final, published article.

Terms of Use

This article is made available under the terms and conditions applicable towards Open Access Policy Articles, as set forth in our [Terms of Use](#).

The Cost of Unconventional Gas Extraction: A Hedonic Analysis

Michael S. Delgado*

Department of Agricultural Economics
Purdue University

Todd Guilfoos[†]

Department of Environmental and Natural Resource Economics
University of Rhode Island

Andrew Boslett[‡]

Department of Environmental and Natural Resource Economics
University of Rhode Island

July 1, 2016

Abstract

We focus on identification and estimation of potentially negative environmental impacts of unconventional natural gas extraction on property values in the United States and advance previous research by contributing new data and new identification strategies for isolating these potential impacts. Our study area consists of two counties in Pennsylvania that are home to large amounts of unconventional natural gas extraction but are otherwise isolated from other resource extraction industries or large urban areas. We deploy parametric, semiparametric, and matching hedonic regression models that include recent quasi-experimental methods and, in contrast to previous research and much popular intuition, we fail to find robust significance that negative environmental externalities of natural gas extraction are manifested in nearby property values. While there may be plausible risks associated with unconventional natural gas extraction, we do not find consistent evidence to suggest that these risks significantly affect nearby property values.

JEL Classification: Q30; Q40; Q51.

Keywords: Unconventional Gas Extraction; Hedonic Analysis; Marcellus Shale; Environmental Costs.

*Michael S. Delgado, Department of Agricultural Economics, Purdue University, West Lafayette, IN 47907-2056. Phone: 765-494-4211, Fax: 765-494-9176, E-mail: delgado2@purdue.edu.

[†]Correspondence to: Todd Guilfoos, Department of Environmental and Natural Resource Economics, University of Rhode Island, Kingston, RI 02881. Phone: 401-874-4398, Email: guilfoos@uri.edu.

[‡]Andrew Boslett, Department of Environmental and Natural Resource Economics, University of Rhode Island, Kingston, RI 02881. Email: aboslet1@my.uri.edu.

Michaela Jellicoe provided excellent research assistance. We are grateful to the Bradford and Lyscom County assessor offices for providing detailed property sale records. We thank Tim Beatty, Raymond J. G. M. Florax, Benjamin Gramig, Martin Heintzleman, Daniel Henderson, Farzad Taheripour, Sophia Villas-Boas, and participants at the 2014 World Congress of Environmental and Resource Economists, 2014 4th Annual Asian Meetings of the International Association of Energy Economics, the 2015 American Enterprise Institute Emerging Scholars Workshop, the 2015 Meetings of the European Association of Environmental and Resource Economics, the University of Alabama, and the Purdue University SHaPE brownbag for helpful feedback. This research was supported in part by computational resources provided by Information Technology at Purdue - Rosen Center for Advanced Computing, Purdue University, West Lafayette, Indiana. This paper was previously circulated under the title 'The Cost of Hydraulic Fracturing: A Hedonic Analysis.'

1 Introduction

Economists have recently begun to show interest in measuring the environmental costs associated with unconventional natural gas production in the United States (e.g., Gopalakrishnan & Klaiber 2014, Muehlenbachs, Spiller & Timmins 2012, Muehlenbachs, Spiller & Timmins 2015, Boslett, Guilfoos & Lang 2015), no doubt in response to the rapid expansion of natural gas exploration and extraction in the Marcellus Shale formation in the northeastern part of the country.¹ While the potential environmental costs of unconventional gas extraction are widely known and documented – e.g., groundwater contamination (Jackson, Vengosh, Darrah, Warner, Down, Poreda, Osborn, Zhao & Karr 2013), air pollution (Litovitz, Curtright, Abramzon, Burger & Samaras 2013), and forest fragmentation (Drohan, Brittingham, Bishop & Yoder 2012) – only recently have economists begun to provide estimates of these potential costs. In light of the scope of potential economic benefits and costs, as well as popularity in public debate, econometric estimates of these potential costs have the potential to substantially influence public policy.

There are few econometric studies that estimate the potential costs of unconventional natural gas production in the Marcellus Shale; these studies use the hedonic method to recover these costs. Gopalakrishnan & Klaiber (2014) find some evidence that shale gas extraction leads to a decline in home prices, ranging from about 1 to 7 percent. This decline in home prices is larger for homes that rely on well water. In their sensitivity analysis, Gopalakrishnan & Klaiber (2014) find mixed evidence that the significance of the decline in home values identified in their preferred specification is robust. Muehlenbachs et al. (2012) focus more narrowly on the external costs of groundwater contamination associated with unconventional gas extraction, and find that unconventional gas extraction wells cause a decrease of up to 24 percent of home values for properties that rely on groundwater resources, as opposed to municipal water sources. Muehlenbachs et al. (2015) focus on housing transactions in both Pennsylvania and New York, and find significant evidence of several possible impacts of shale gas extraction activities (including, for example, different effects on well water or municipal water homes). Finally, Boslett et al. (2015) find a significant decline in housing values in New York State in response to New York’s shale development moratorium, and do not find a significant negative effect of shale development on Pennsylvania housing values.

Gopalakrishnan & Klaiber (2014) rely on data measured in the beginning years of the shale exploration and extraction boom (boom from 2008-2012) in order to minimize the potential for

¹Generally, the term ‘unconventional’ is used to refer to ‘horizontal’ hydraulic fracturing, in contrast to ‘conventional’ or ‘vertical’ hydraulic fracturing. Throughout the paper, we use the term unconventional natural gas extraction. We define these terms in detail in Section 2.

positive externalities of shale exploration or extraction to manifest in rising home values. This strategy may not be able to identify any longer run effects of shale gas exploration if the initial manifestation of externalities was either short lived or lower than the longer run effects. Muehlenbachs et al. (2012, 2015) provide an alternative identification strategy by using a triple difference approach (combined with matching in the later case) to account for unobservable factors that may be correlated with shale extraction activity. Finally, both Muehlenbachs et al. (2012) and Gopalakrishnan & Klaiber (2014) focus on Washington County. Parts of western Pennsylvania, which includes Washington County, have a relatively large amount of unconventional gas production; it also has a relatively long history of coal mining, conventional gas production, and other forms of resource extraction. It is not clear how this history may affect any econometric study focused on unconventional natural gas externalities. Furthermore, Washington County is relatively close to Pittsburgh; while it is possible to control for any confounding effects via some distance to Pittsburgh measure in a regression, it is not necessarily clear how to parameterize the model so as to adequately control for all effects of Pittsburgh on the estimates of the negative effects of unconventional gas extraction. There may be complex and nonlinear interactions between Pittsburgh and the surrounding areas that are difficult to account for statistically.

Our paper builds on existing studies by focusing broadly on the overall environmental costs of unconventional gas extraction in the spirit of Gopalakrishnan & Klaiber (2014), while adopting robust econometric identification strategies that are, in some ways, similar to the models deployed by Muehlenbachs et al. (2012, 2015). We focus separately on two different counties in northeastern Pennsylvania – Lycoming and Bradford counties – each with different strengths for econometric identification. Both Lycoming and Bradford counties are home to relatively large amounts of shale gas extraction; Bradford County is home to the largest number of unconventional wells in Pennsylvania. However, (i) neither Lycoming nor Bradford County have been impacted much by any other type of natural resource extraction, including conventional natural gas extraction, and (ii) neither are home to any large cities. We contend that, in contrast to many other counties in Pennsylvania, both Lycoming and Bradford counties are ideal counties in which to study the potential effects of unconventional natural gas extraction. Figure 1 shows the location of natural resource industries in Pennsylvania, highlighting the location of both Lycoming and Bradford counties (Washington County is located in the far southwest corner of Pennsylvania) and illustrating that these two counties are largely free from other resource extraction activities.

We make three distinct contributions in this study. First, we apply different techniques that rely on different assumptions for identification to the same data set to find if results are significant

across our two counties in Pennsylvania and across different methods to determine the robustness of any statistically significant externalities. We focus on an area that has little of the legacy effects of coal mining and conventional gas extraction or urban development, while also containing relatively large amounts of unconventional gas extraction. This provides depth to the literature of the effects of shale gas development and alleviates any concern that complex interactions between alternative resource extraction activities or urbanity confound our estimates. It is also possible that instead of econometric bias or confounding identification issues, our study area acts as a check on the external validity of results found in other studies.

Second, we employ two methods that reduce our dependence on linear estimates of effects. One is matching, which provides a flexible approach to imputing counterfactuals that tends to be more reliable in cases in which the linear regression structure is not well specified (Imbens & Rubin 2015). We also estimate nonlinear effects of unconventional well proximity using a semiparametric estimation strategy that does not rely on a specific functional form of nonlinearity within the relationship between housing prices and shale wells. The virtue of this model is that it nests many common forms of nonlinearity, and does not require a priori specification of the model structure. These nonlinearities appear to be present and significant in the distance to shale gas development in Bradford County.

Third, we use the border of the shale as an alternative means of identifying negative externalities of shale gas development. Lycoming County is unique in that there is a natural geological boundary in the Marcellus Shale that bisects the county.² We exploit the natural geological boundary of the Marcellus Shale formation as a new strategy for controlling for any positive effects from shale extraction that may inflate home values (e.g., leases and royalty payments), and help isolate and identify the total environmental costs of unconventional gas production. Homes that are not above the Marcellus Shale, but are located near unconventional gas wells on the geological boundary, would be subject to the negative effects of unconventional gas production but not any positive effects. Hence, this econometric setup provides an alternative – and potentially more robust – opportunity to identify the negative environmental effects of unconventional gas extraction.

Our econometric estimates show that there does not exist a robustly significant negative effect of unconventional gas wells on property values in either county, or across the different identification strategies. We do find that nonlinearities can be important to the estimate of distance to unconventional gas wells, but that these effects tend to be localized over space. These results highlight

²Several other counties in Pennsylvania are bisected by this natural geological boundary, but none of these other counties are home to a large enough number of unconventional gas wells to allow for reliable econometric estimation. See Panel A in Figure 1.

the lack of robustness of results of proximity to unconventional gas extraction practices and the difficulty in generalizing results across significantly different regions.

In Section 2, we provide a detailed overview of unconventional gas extraction, the Marcellus Shale formation and its deposits, and the potential external costs of unconventional gas extraction. We detail our econometric strategy in Section 3, and describe the data in Section 4. We discuss our results and robustness checks in Section 5, and Section 6 provides our conclusions and a policy discussion of our results. A supplemental appendix with detailed results from many robustness checks is available from the authors.

2 Overview of Unconventional Gas Extraction in the United States

2.1 Unconventional Natural Gas Extraction

Conventional natural gas wells have been in production in the United States since the early 1800s. A conventional natural gas well operates by drilling a single, vertical pipeline down into the geological shale formation to recover the gas. This technique is effective if the gas is able to move relatively freely beneath the surface, so that only a single well is required to recover large amounts of natural gas in a particular area. However, much of the natural gas in the United States is trapped inside low-permeability shale formations, rendering extraction using conventional techniques costly. This means that a conventional gas well extracts shale gas at a relatively high cost associated with setting up and drilling each well. Since the gas cannot freely move beneath the surface due to the geological characteristics of the shale formation, the quantity of gas that can be extracted from a single well is relatively small. Recent advances in unconventional gas extraction technologies - specifically, the combination of hydraulic fracturing and horizontal drilling - have facilitated a boom in natural gas exploration and production by providing a means of extracting large quantities of gas from a single well.

Hydraulic fracturing is a flow-stimulation technique that works by exploiting the natural formation of pockets within the shale where the gas is trapped. Highly pressurized, water-based fluids are pumped down into the well to expand naturally occurring micro-fractures in the rock formation. Expansion of these fractures allows the gas to flow freely within the rock for easy extraction. While the specific ingredients in the fracturing fluid varies across firms and drilling sites (Fitzgerald 2013), a typical fluid mixture contains water; sand to hold open the fractures in the rock; and multiple chemical additives to reduce friction, prevent the formation of mineral and iron compounds, and reduce the growth of bacteria.

The most recent advance in the natural gas drilling process is horizontal drilling, in which the well-bore turns horizontal once it penetrates the shale formation in which the gas is trapped. This technique allows operators to extract natural gas from a larger surface area. One major limitation of vertical drilling techniques is that the pressurized fracturing fluid, essential to expanding the fractures within the rock, has a limited exposure of about 50 feet inside the rock formation. Only so much fluid will fit at the bottom of the well-bore where the gas is trapped. Allowing the well-bore to turn horizontal once the well-bore has reached and penetrated the shale formation greatly increases the exposure of the fracturing fluids to the gas reservoir. Such horizontal well-bores can extend over a mile from the well-pad.

The application of horizontal drilling, regularly used in offshore drilling, has also facilitated the invention of multi-well pads, in which multiple horizontal well-bores are drilled from a single surface platform known as a well-pad. Drilling multiple well-bores from a single pad reduces the cost of extraction, provides more thorough extraction coverage from a single area, and substantially reduces the environmental impact associated with constructing and operating multiple well-pads (Ladlee & Jacquet 2012). A multi-well pad does not require as much surface equipment or as many well-bores penetrating the ground to extract the gas over the 1-mile radius area, relative to a multitude of conventional wells. Hence, a multi-well horizontal hydraulic fracturing well-pad is capable of harvesting much of the natural gas deposited within a one mile radius of the well-pad, while limiting the internal costs to the gas company, and potentially some of the external environmental costs of drilling.

Over the course of its life cycle, one shale gas well in the Marcellus region can require 4 to 6 million gallons of water (Jiang, Hendrickson & VanBriesen 2014). Along with the transportation of equipment, personnel, and other inputs in the fracking process, these requirements necessitate significant traffic, which could lead to more recognition of shale development happening in the region. After the well is pressurized with water, sand, and additives mixture, which can take from 3 to 10 days, the waste water is removed and disposed of, with the site being remediated thereafter.

2.2 Marcellus Shale

There are numerous shale gas reservoirs located within the United States, the largest of which is the Marcellus Shale located in the northeastern United States, covering 95,000 square miles and spanning parts of West Virginia, Ohio, Pennsylvania, Maryland, and New York. It is estimated that the Marcellus Shale contains 187 Trillion cubic feet (Tcf) of technologically recoverable reserves (United States Energy Information Administration 2012), and can produce enough natural gas to

supply the United States for at least 20 years (National Energy Technology Laboratory 2012). The Marcellus Shale formation lies approximately 4,000 to 8,500 feet below the surface; the depth to the base of treatable water resources is about 850 feet, after which there are thousands of feet of impermeable rock separating the natural gas deposits from groundwater supplies and the surface.

The quantity of natural gas deposits potentially recoverable from the Marcellus Shale formation has led to a rapid expansion of exploration and drilling, particularly in Pennsylvania. Property owners may receive lease payments of at least \$5,000 per acre, and royalty payments of up to 20 percent (Gopalakrishnan & Klaiber 2014). The contracts stipulating the payments are private information, making it difficult to estimate expected financial benefits for any one property. According to the Pennsylvania Department of Environmental Protection, over three thousand new conventional and unconventional gas well permits are issued each year, and nearly the same amount of conventional and unconventional gas wells are drilled. Such a rapid expansion and rush for resource extraction has made the Marcellus Shale formation the subject of intense popular debate, and the focus of federal, state, and local environmental and economic policy.

2.3 Financial Benefits of Shale Gas Development

Mineral right ownership depends significantly upon past history of coal, oil, and gas development. Specifically, the subsurface rights can be split from surface rights in what is called a split estate. The counties that we focus on, Lycoming and Bradford County, do not have a large history of these resource operations, and mineral rights are more likely retained by the owner of the surface rights (Kelsey, Metcalf & Salcedo 2012, Pender, Weber & Brown 2014). In Washington County, where there is over one hundred years of history in coal, oil, and gas development, the mineral rights are more likely to be separated. The actual records of mineral rights ownership is difficult to obtain because they only exist on a deed-by-deed basis (Kelsey et al. 2012), and have not been included in any of the hedonic studies to date.

These mineral rights are important because they are the source of significant royalty and lease payments which make up the largest source of financial benefits to homeowners from shale gas development. Initially, homeowners can expect a signing bonus when agreeing to a lease with a gas company (Weidner 2013), and this payment can vary from \$50 to \$6,000 based on a number of factors such as geological factors and market conditions. In addition to lease payments there are royalty payments which are tied to the productivity of wells, and therefore can vary dramatically. These payments can be substantial: in Bradford County approximately \$160 million have been paid in royalties by 2012 (Loewenstein 2012). These royalty payments, depending on the negotiated

rates, could provide a windfall for homeowners, though there is great uncertainty over the size and regularity of payments as some wells are never put into production, and others may be less productive than originally anticipated. Another factor influencing the size of payments is the number of well bores drilled and put into production on a well-pad using horizontal drilling. The perceived benefits are based on the subjective probability that a property will receive a lease and royalties, which is also influenced by local development or size of acreage of the property. If factors such as local development influence the subjective probability of benefits of shale gas development it may be difficult to separate the negative externalities derived from local development from the expectations of financial benefits.

2.4 Environmental Risks from Unconventional Gas Extraction

At the heart of the policy debate regarding the unconventional extraction of natural gas deposits is the external environmental cost of drilling. In particular, unconventional gas extraction poses several potential environmental costs that are not present in conventional gas extraction.

Although shale formations lie several thousand feet below groundwater tables, the most substantial concern addressed in both the media and academic literature revolves around groundwater and surface water protection. During the drilling process the well-bores must first penetrate the groundwater table and then safely return the fracturing flow-back fluid (containing a mix of toxic compounds used or generated during the fracturing process) and harvested natural gas to the surface without contaminating the groundwater supply. In order to protect against groundwater contamination the well-bore is constructed with multiple layers of protective casing to prevent any leaks. However, contamination might occur through faulty or improperly installed well casings. Studies have found that groundwater (e.g., Jackson et al. 2013) and surface water (e.g., Olmstead, Muehlenbachs, Shih, Chu & Krupnick 2013) pollution levels are associated with shale gas development in the Marcellus Shale. These studies, along with information highlighting potential conduits connecting deep shale basins with shallow aquifers (Warner, Jackson, Darrah, Osborn, Down, Zhao, White & Vengosh 2012), provide sufficient grounds to suspect that the risk of water contamination may be manifested in home values.

Another potentially substantial environmental cost comes from management and disposal of the fracturing flow-back fluids (e.g., Gregory, Vidic & Dzombak 2011, Rahm, Bates, Bertoia, Galford, Yoxtheimer & Riha 2013). Even if well casings are properly functioning and secure, flow-back wastewater must be properly retrieved, stored, and disposed. Several disposal options are available: injection into used underground wells, onsite treatment after which the water can be reused for

unconventional gas extraction, or disposal at an offsite municipal treatment plant. Olmstead et al. (2013), for instance, find that average chloride levels are relatively higher downstream of wastewater treatment plants where water used in the fracturing process is being processed. Clearly, each of these options bears potential risks of leakage, which may cause significant damage to environmental or human health. The adverse risks through public water sources could be seen as minimal since the water is treated and the public generally has a high level of confidence in water that is municipally treated. Recent work (Wrenn, Klaiber & Jaenicke 2015) suggests that there is an increase in consumption of bottled water across the shale region, which may be a function of the perceived risks to water supplies.

In addition to the risks of water contamination, there are other external environmental costs associated with unconventional gas extraction that may have significant negative externalities. For instance, unconventional gas extraction requires massive quantities of water, which decreases groundwater or surface water levels, and may impact regional agricultural productivity, animal health, or recreational activity. Transportation activities may cause damage to local roads and highways that are unprepared for large scale traffic associated with the extraction process (e.g., Brasier, Filteau, McLaughlin, Jacquet, Stedman, Kelsey & Goetz 2011). Natural gas extraction produces harmful gases that likely escape into the atmosphere, and forest fragmentation due to construction of extraction and transportation-based infrastructure that may impact both habitat quality and local landscape aesthetics (Drohan et al. 2012).

It is clear that, despite the macroeconomic benefits of natural gas exploration and extraction, there are potentially substantial costs that policymakers must consider when deciding how to regulate the extraction of natural gas. While these potential environmental costs are well-known and well-documented, the magnitude of these costs are not. Only recently have economists attempted to quantify these costs to better inform policymakers of all the costs associated with unconventional natural gas extraction.

3 Empirical Model and Estimation Methods

Our interest is in estimating the impact of shale extraction activity on neighboring property values. Our hypothesis is:

Hypothesis: Living near an unconventional natural gas well leads to a decrease in property values, *ceteris paribus*.

To do so, we focus on the average treatment effect (ATE), and for some models the average effect of treatment on treated properties (ATT). Our data does not include repeat sales transactions, so we focus on parametric and semiparametric regression and matching hedonic methods, some of which are used by Gopalakrishnan & Klaiber (2014) and Muehlenbachs et al. (2012, 2015).

3.1 Potential Outcomes and Causal Effects

A causal effect is the difference between two potential outcomes for a single observational unit, in which the potential outcomes are characterized by treatment or non-treatment (Rubin 1974). Let $i = 1, 2, \dots, n$ denote a set of properties for which we observe transactions, and define T_i to be a binary indicator for treatment status such that $T_i = 1$ indicates that property i is treated. In the context of shale extraction activity, define treatment to mean that there is at least one single shale extraction well within $r = \{1, 2, 3\}$ miles of a property. Later, we also define treatment to be the continuous distance to nearest well, or the count of shale wells within r miles. For the purpose of exposition, here we focus on binary treatment. Let Y_{i1} be the value for property i when i is treated, and let Y_{i0} be the value for the same property when the property is not treated.

The treatment effect for property i is $\tau_i = Y_{i1} - Y_{i0}$. The average treatment effect (ATE) is $\tau_{ate} = E[Y_{i1} - Y_{i0}]$ or the average difference in potential outcomes across the population, and the ATT is $\tau_{att} = E[Y_{i1} - Y_{i0} | T_i = 1]$ which is the average difference in potential outcomes for the subset of treated properties. The *fundamental problem of causal inference* is that both potential outcomes are never observed for the same property (Holland 1986), so τ_i cannot be observed. Rather, the observed outcome is $Y_i = Y_{i0}(1 - T_i) + Y_{i1}T_i$. Identification of τ_i requires imputation of the counterfactual. As usual, different approaches are potentially applicable, and each has its relative merits. Two general reviews of these issues include Imbens (2004) and Imbens & Wooldridge (2009), and recent textbook treatment can be found in Imbens & Rubin (2015).

In the context of our cross-sectional sample, identification of the treatment parameters requires conditional unconfoundedness and overlap in the covariate distribution. The unconfoundedness assumption requires that there are no unobserved covariates that influence both treatment and the outcome. This is the exogeneity condition in a regression. The overlap assumption requires there to be both treated and untreated (control) units over the support of the conditioning variables. Together, these assumptions allow us to form a control group through which to obtain the counterfactual outcomes and estimate the treatment effect parameters. We now discuss the methods employed in this paper and related identification issues.

3.2 The Regression Approach

3.2.1 Linking the Hedonic Regression to the Potential Outcomes Model

Under certain assumptions, the standard parametric, linear in parameters hedonic model identifies τ_{ate} . Suppose that τ_i is constant so that $\tau = Y_{i1} - Y_{i0}, \forall i$, and the conditional mean of Y_{i0} takes a parametric and linear in parameters form: $Y_{i0} = X_i\beta + \epsilon_i$. Then, maintaining the unconfoundedness assumption, $Y_i = Y_{i0}(1 - T_i) + Y_{i1}T_i$ implies

$$Y_i = X_i\beta + \tau T_i + \epsilon_i \quad (1)$$

and ordinary least squares can be used to estimate τ .

The standard empirical hedonic model is a regression of the price of the house, P_i , on a vector of housing characteristics, X_i , for some conditional mean regression function $m(\cdot)$ (Rosen 1974, Taylor 2003, Palmquist 2005, Parmeter & Pope 2013). In general, the functional form of the hedonic price function is unknown, and typically believed to be nonlinear (Parmeter & Pope 2013); however, in practice, $m(\cdot)$ is often assumed to be parametric and linear in parameters:

$$\begin{aligned} P_i &= m(X_i) + \epsilon_i \\ &= m(X_i; \beta) + \epsilon_i \\ &= X_i\beta + \epsilon_i. \end{aligned} \quad (2)$$

The parameter β is the slope of the hedonic price function for a particular attribute, and is the marginal price of the house for attribute $x \in X$. Hence, an estimate of β recovers the marginal willingness to pay for each attribute. Adding a shale activity treatment indicator to (2) yields the treatment effect regression model (1). The assumption in the hedonic framework is that the assumed structure of $m(X_i)$ is sufficient for ensuring that there is no correlation between unobservable factors that influence housing values and the observed characteristics, X_i (unconfoundedness/exogeneity). If this restriction is not satisfied, estimates of the treatment effect will be biased.³

A common hedonic specification is the semi-log model:

$$\ln P_i = X_i\beta + \gamma T_i + \epsilon_i \quad (3)$$

³In addition, the hedonic framework requires that the houses come from a single housing market, and that the hedonic price function is stable over time or space (Parmeter & Pope 2013).

so that β represents a constant percentage change in the price of a house from a change in attribute $x \in X$ (e.g., Heintzleman & Tuttle 2012). In either case, estimation of the hedonic price function is a straightforward problem to be solved using least squares.

One Achilles' heel of this setup is the potential for omitted factors that affect property values to be correlated with X_i . Yet, to a large extent, these concerns may be assuaged given augmentation of (3) to include interaction effects within X_i – for instance, interaction of the treatment variable(s) with well-water dependency (Muehlenbachs et al. 2012, Gopalakrishnan & Klaiber 2014, Muehlenbachs et al. 2015) or land acreage. In fact, these extensions allow the treatment effect to be heterogeneous, such as differential across well water or public water properties. Or, one can include temporal and spatial fixed effects to nonparametrically control for unobservable and potentially confounding factors. While the possibility of uncontrolled for unobservable factors may always be present, research has shown that careful parameterization of likely sources of heterogeneity and inclusion of a variety of fixed effects can dramatically improve the performance of the benchmark model (Kuminoff, Parmeter & Pope 2012).

3.2.2 Semiparametric Hedonic Regression

A particularly robust means of hedonic implementation is to avoid parameterizing key parts of the hedonic regression function and instead deploy semi-parametric methods.⁴ As is well-known, semi-parametric techniques are generally free from problems arising from misspecification of the nonparametric part of the regression function (Li & Racine 2007, Henderson & Parmeter 2015). In essence, deployment of semi-parametric techniques frees the econometrician from the burden of correctly parameterizing (part of) the hedonic function. Through this generalization, the potential for omitted heterogeneity to manifest in correlation between X_i and ϵ_i can be dramatically reduced.

Here, we focus on the semi-parametric partial linear structure

$$P_i = m_0(T_i) + X_i\beta + \epsilon_i \tag{4}$$

as our generalized semi-parametric hedonic structure as it provides a straightforward, and econometrically testable, generalization of our benchmark model.⁵ In (4), $m_0(\cdot)$ is some unknown smooth

⁴Nonparametric techniques are also available (Parmeter, Henderson & Kumbhakar 2007), such as direct estimation of $m(\cdot)$ in (2), though sometimes less practical from an empirical standpoint when the dimension of control variables is large. Here, we restrict our discussion to semi-parametric methods. See Li & Racine (2007) and Henderson & Parmeter (2015) for further details.

⁵To provide an example, one possible parametric specification is the cubic design

$$P_i = \gamma_0 + \gamma_1 T_i + \gamma_2 T_i^2 + \gamma_3 T_i^3 + X_i\beta + \epsilon_i \tag{5}$$

function of the treatment variable; in the semi-parametric model, treatment is the distance from the property to the nearest unconventional well. An attractive feature of this design is that while the treatment variable enters nonparametrically into the regression function, the specification is identical to (3) for all the other conditioning variables (and the β coefficients have an identical interpretation). Here, the treatment effect is heterogeneous, and is given by $\partial m_0(T)/\partial t$, or the first order derivative of $m_0(\cdot)$ with respect to some $t \in T$. Since $m_0(T)$ is allowed to be nonlinear, the marginal effect is also allowed to be nonlinear.⁶ For further technical details and discussion, see Li & Racine (2007) and Henderson & Parmeter (2015).

3.2.3 Border Identification

An alternative means of expanding the flexibility of the hedonic regression approach is to exploit boundaries (i.e., discontinuities) for identification. For Lycoming County, we divide our housing market into four separate groups, exploiting the fact that the natural boundary of the Marcellus Shale formation runs cleanly through part of Lycoming County (shown in Figure 1). We anticipate that, if there exists any positive effects of unconventional gas extraction (e.g., value of mineral rights), properties that are located near a gas well but are not over the shale will primarily feel the negative environmental effects. This rests on the assumption that positive effects are primarily realized through leases and royalties, and the positive spillover effects across the shale border are minimal. Some positive spillover could be realized through the job market or increased demand for housing in the regional area.

Specifically, our groups are defined on four possible locations based on (i) whether or not the property is located within r miles of a gas well, and (ii) whether or not the property is located above the Marcellus Shale. Figure 2 provides a graphical depiction of each of the four possibilities,

in which the marginal treatment effect is $\gamma_1 + 2\gamma_2 T_i + 3\gamma_3 T_i^2$. The difference is that the semi-parametric design is more general and is not only limited to the cubic form.

⁶Estimation of (4) is straightforward using two-step kernel regression methods (Robinson 1988, Li & Racine 2007, Henderson & Parmeter 2015). Subtracting $E[P_i|T_i]$ from P_i yields

$$\begin{aligned} P_i - E[P_i|T_i] &= [m_0(T_i) + X_i\beta + \epsilon_i] - [m_0(T_i) + E[X_i|T_i]\beta + E[\epsilon_i|T_i]] \\ &= \{X_i - E[X_i|T_i]\}\beta + \epsilon_i \\ &= X_i^*\beta + \epsilon_i \end{aligned} \tag{6}$$

From (6), estimation proceeds as follows. In the first step, both P_i and X_i are regressed on T_i , which are then used to construct P_i^* and X_i^* . Next, ordinary least squares of P_i^* on X_i^* yields $\hat{\beta}$. The final step uses nonparametric techniques to regress $P_i^{**} \equiv P_i - X_i\hat{\beta}$ on T_i to obtain $\hat{m}_0(T_i)$. While a variety of techniques are applicable, a common approach is to use local-linear least squares

$$\hat{\delta} = (\tilde{T}'K(T)\tilde{T})^{-1}\tilde{T}'K(T)P^{**} \tag{7}$$

in which the i th row of \tilde{T} is $(1, T_i - t)$ for some point $t \in T_i$ and diagonal kernel weight matrix $K(T)$. For the local-linear estimator, $\delta = (m_0(T), \partial m_0(T)/\partial t)'$; that is, the local-linear estimator produces an estimate of both the nonparametric function and its partial derivative in a single step.

with the \oplus symbol denoting potential well locations and the numbers $\{1, 2, 3, 4\}$ denoting relative location of each well group. The groups are defined as follows: Group 1 – houses located within r miles of a well, but are *not* located above the Marcellus Shale; Group 2 – houses located above the Marcellus Shale but are *not* located within r miles of a well; Group 3 – houses that are located above the Marcellus Shale and are located within r miles of a well; and Group 4 – houses that are *not* located above the shale and are *not* near a gas well. Houses in Group 4 form our control group, as we do not expect that natural gas drilling has any positive or negative effects on these houses. That is, for these properties, there is zero potential for drilling leases or royalty payments as these houses are not located above the Marcellus Shale, nor are there likely to be any environmental effects as these houses are not located within close proximity of the wells to feel any negative effects of unconventional drilling. We address some potential caveats below.

Following our outline of these groups, we define the following indicator variables, leaving Group 4 as our reference group:

$$d_1 = \begin{cases} 1 & \text{if located within } r \text{ miles, but not over Shale} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

$$d_2 = \begin{cases} 1 & \text{if located over Shale, but not within } r \text{ miles} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

$$d_3 = \begin{cases} 1 & \text{if located within } r \text{ miles, and over Shale} \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

Adding these three new indicators to our regression model yields

$$\ln P_i = X_i\beta + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \gamma_3 d_{3i} + \epsilon_i. \quad (11)$$

This model allows us to test our hypothesis via the parameter γ_1 , as the properties in Group 1 are likely to feel harmful effects of unconventional drilling, without the luxury of any drilling or mineral rights. In (11), our reference group is Group 4; hence, $\hat{\gamma}_1$ is an estimate of the marginal effect on home values of living near a gas well but not over the Marcellus Shale, relative to homes in the control group that are neither within r miles of a well, nor are over the Marcellus Shale. The interpretation of γ_1 is directly comparable to the estimates and findings of Gopalakrishnan & Klaiber (2014) and Muehlenbachs et al. (2012, 2015).

Our Lycoming County model also allows us to directly identify and estimate the effect of living above the Marcellus Shale, *ceteris paribus*, as well as the net effect of living both over the Marcellus Shale and near an unconventional gas well. These effects are captured through the parameters γ_2 and γ_3 , respectively. Notice that γ_3 bears a similar interpretation to γ in (3), in that it captures the net effect of unconventional gas drilling, relative to homes that are plausibly immune to any effects of natural gas.

3.2.4 A Potential Pitfall of the Regression Approach

As described, the regression approach relies on the unconfoundedness and overlap assumptions to estimate the treatment parameter. We bear in mind the potential sensitivity of the regression approach to a lack of overlap; details of this problem are given in Imbens & Rubin (2015). In the case of lack of overlap, the regression approach relies heavily on the assumed structure for extrapolation to counterfactuals. This issue is not unique to our model or our data; this fact is important to bear in mind when considering our results using this model, as well as the hedonic regressions deployed by Muehlenbachs et al. (2012) and Gopalakrishnan & Klaiber (2014).

Nevertheless, we begin with regression models for the following reasons. First, this model is directly linked to published results by Gopalakrishnan & Klaiber (2014) and unpublished results by Muehlenbachs et al. (2012), which allows us to understand our data in the context of a previously considered regression framework. It is noteworthy that previous studies have found statistically significant and quantifiable effects of shale activity on property prices from the vantage of this model. Second, this model is straightforward to estimate. It is not clear the extent to which this model is affected by potentially confounding factors, and as we show, there are a variety of robustness checks in the context of this model that allow us to assess the extent to which confounding factors may bias our estimates. For example, regressions using 2008 property transactions are likely to be free from any property value/rental market spillover effects from increased labor in the area, because these effects are not likely to be very strong in the first year of shale activity. Third, we explore balance and overlap in the covariates (in the appendix), and while many are not well-balanced it is not clear the extent to which reliance on the regression structure impacts our estimates. Fourth, at least in Bradford County, we are not able to find a credible control group of property transactions that might lend well to a treatment/control design. One might consider properties from New York State or from farther counties in Pennsylvania following Boslett et al. (2015) and Muehlenbachs et al. (2015), but it is also not clear if these other properties only differ along observable dimensions. At least in terms of a starting place, the baseline log-linear regression seems to be reasonable.

3.3 Matching

An alternative approach is to use matching techniques. While regression models include observed covariates in a (often specified) conditional mean to control for potentially confounding effects, matching models remove bias by imputing the missing counterfactual outcomes by using averaged observed outcomes for observations in the opposite (treated or untreated) groups that are statistically similar along observed dimensions.

Letting M denote the number of matched units for each i . The potential outcomes are then

$$\widehat{Y}_{i0} = \begin{cases} Y_i & \text{if } T_i = 0 \\ \frac{1}{M} \sum_j Y_j & \text{if } T_i = 1 \end{cases} \quad \widehat{Y}_{i1} = \begin{cases} \frac{1}{M} \sum_j Y_j & \text{if } T_i = 0 \\ Y_i & \text{if } T_i = 1 \end{cases} \quad (12)$$

in which j indexes the M closest matches. The average treatment effect estimator is $\widehat{\tau}_{ate} = n^{-1} \sum_{i=1}^n [\widehat{Y}_{i1} - \widehat{Y}_{i0}]$, but often to avoid having to impute both sets of counterfactuals given data limitations researchers focus only on the *ATT*.

When deploying matching methods, researchers must decide how many similar control units to match to each treated unit. In practice, one-to-one matching (i.e., selecting one control unit to match to each treated unit) provides the matching estimator with least bias (Imbens 2004). Using multiple matches can reduce the variance of the estimate, but the marginal reduction in the variance from an additional match is diminishing; Imbens & Rubin (2015) state that there is very little to be gained from matching beyond 2 or 3 units (per each treated unit). Muehlenbachs et al. (2015) elect to match each treated property to the 4 nearest control properties.

There are different ways to match treated/untreated properties to obtain counterfactuals. One approach is to match directly on X_i , by minimizing the observed distance between treated and control properties (Imbens & Rubin 2015). This is called multivariate matching, and the distance commonly used to determine a match is the Mahalanobis distance $mdist(X_i, X_j) = (X_i - X_j)'S^{-1}(X_i - X_j)$ in which S is the sample covariance between X_i and X_j . Based on this distance, treated units are paired with the closest (or closest few) untreated units. It is important to keep in mind that matching estimators are biased when the number of continuous covariates used in the match exceeds one; in such case, bias corrected estimators must be used; see Abadie & Imbens (2006, 2011).

An alternative, and common technique is to match on the propensity score. Rosenbaum & Rubin (1983) show that matching on the propensity score asymptotically provides balance to observable factors between treated and untreated properties. Hence, rather than matching on multivariate X_i

directly, matching can be done on the univariate propensity score. Letting $P(X_i) = Pr[T_i = 1|X_i]$ be the propensity score, treated/control observations are matched if they have the same propensity (within some range, since $P(X_i)$ is continuous) of being assigned treatment. Within this match, treatment should be considered to be random. Since the propensity score is univariate, the bias corrected estimator is not needed (Abadie & Imbens 2006). However, a variance correction is needed to account for the fact that the propensity score is itself an estimate (Abadie & Imbens 2016).

When there is concern that treatment is not random, though the selection on observables assumption is assumed to hold, a critical means of assessing model reliability is via assessing whether the matched treated/control samples are statistically similar along observable dimensions. As is evident, matching methods rely critically on being able to ‘balance’ the covariates between the treated/untreated properties, post-match. If balancing fails, then the treatment parameter estimates may still be biased. To provide an assessment of covariate balance, both pre-match and post-match, we deploy several balancing statistics described by Imbens & Rubin (2015).

We use the normalized difference, which provides a measure of dispersion of the means of the two samples that is unit free. The larger the normalized difference for each covariate, the more difficult it will be to deploy adjustment techniques to remove biases. The second measure we use is the log ratio of standard deviations, which measures the difference in dispersions of the two distributions. Finally, the third metric we use is an assessment of the fraction of observations in the tails of the opposing treatment group distribution, which provides a means of assessing overlap in the covariate distributions. See Imbens & Rubin (2015) for details.

3.4 Discussion

Here we provide a few comments relating to our models and identification strategies. First, we do not have an *a priori* preferred estimation strategy. Both the parametric hedonic regression and matching models were used by Gopalakrishnan & Klaiber (2014) and Muehlenbachs et al. (2012, 2015); perhaps most importantly, in previous research both models found significant negative impacts of shale activity on nearby home values. Our interest in exploring a variety of models is to obtain the most robust assessment of shale activity in our study area. By understanding situations (data and econometric) in which significant shale activity effects are found, we can better understand the relationship between shale activity and property values.

Second, the timing of treatment in this model is different for each property. A property is “treated” by being sold after a well has been permitted in proximity of the house. Since we do not have repeat sales in our dataset, our study is primarily a cross-sectional look at the effect of

unconventional wells and their effect on property values. Another important aspect of the analysis is the pathways through which externalities are realized. Within close proximity of a well there may be negative externalities, but also the expectations of future benefits through leases and royalties. The literature suggests that groundwater is one pathway of risk from shale gas development through spills and surface water contamination; therefore it is important to include specifications that control for the source of drinking water at homes (e.g., private wells vs. municipal drinking water). It may also be the case that pathways exist through noise and aversion to the sight of well pads and drilling activity, and therefore the externalities may be highly localized over space and over time. We explore these possibilities further in the discussion of results and in the online appendix.

Third, identification in our Lycoming County setup depends on the value of r , which so far has been purposefully left unspecified. If, for example, negative effects of unconventional drilling are highly localized and r is chosen to be too large, then our indicators defined based on distance to a well will not accurately capture these negative effects. Muehlenbachs et al. (2012) indeed find that the negative effects of living near an unconventional gas well are localized, and are effectively zero at a distance of about two miles. Hence, we follow Muehlenbachs et al. (2012) and set $r = 1$ mile, but vary r as an important robustness check of our results.

Fourth, an important assumption embedded in our definition of Group 4 as a control group is that the negative effects of gas drilling do not affect property values substantially far from the wells. If the risk of large-scale groundwater contamination was substantial, one might expect unconventional gas extraction to impact property values for any homes located along the same watershed. Or, alternatively, truck traffic to and from wells might adversely affect home values for homes that are not located near a well but are located along the same roadways or highways. While we recognize these possible issues, we have not found any evidence suggesting that perceived risks from unconventional gas production are large enough to suggest that homes located relatively far from either the Marcellus Shale or any unconventional wells would be affected (e.g., the localized effects found by Muehlenbachs et al. 2012). Furthermore, our baseline model does not rely on Group 4 for identification.

4 Data

4.1 Housing Data

Our housing transaction data come from the county assessor's office from both Bradford and Lycoming counties. In both counties the data on housing transactions includes transactions before

the production of any shale gas. The Bradford County housing data spans 2006-2012, while the Lycoming County housing data spans 2004-2013. The housing attributes that are captured with the sales data include the age of the house at time of sale; the number of stories, bedrooms, and bathrooms; living area in square feet; indicators for basement, garage, central air, and forced air; total property acreage; indicators for town of residence; the date of sale; and precise geographic location.

We use GIS software (ArcGIS Version 10.2) to map the precise location of each property in order to match each property with water sources (in Bradford County), census tract, and relative distances to several measures of shale gas activity. In Bradford County, the location of each property was determined by tax parcel ID number, which was provided by the county assessor's office through a GIS map of the county. In Lycoming County, we use the property address to find the GIS location (latitude and longitude) through the World Geocode Services (ArcGIS Online) in ArcGIS.

Following past research (e.g., Gopalakrishnan & Klaiber 2014), we include an indicator for whether or not the property relies on a private source of water. Municipal water supplies are carefully monitored, so it is likely that homes that rely on private sources of water are exposed to greater risk associated with natural gas extraction (Muehlenbachs et al. 2012). In Bradford County, we identify properties with access to municipal water supplies using GIS to compare county maps of areas on municipal water with the exact location of each property. We assume that homes outside of municipal supplies rely primarily on well water. In Lycoming County, the record of sale noted if the property had private well water.

4.2 Shale Activity and Shale Boundary

The Pennsylvania State Department of Environmental Protection maintains records of all wells that are permitted to drill for natural gas, and of actual drilling activity. The records also provide the latitude and longitude of each well. We use these records to identify and map the location of each well in GIS. These wells are then connected to the housing sales data in two ways. First, we define the nearest well to each property as the nearest well with a permit date which is earlier than the sale date of the home. We record the distance to this nearest well in GIS for each property, to provide a continuous measure of the relationship between shale activity and property values. Second, we count the number of wells within a one mile, two mile, and three mile radius of each property, counting only the wells with permit dates earlier than the sale date of the home. This measure provides a discrete measure of the relationship between shale activity and property

values. We consider both continuous and discrete shale activity measures because we may find a different effect on property values using the discrete well counts relative to the continuous measure of distance to the nearest well. For a greater number of wells in close proximity to a property, all shale-related activities may increase, such as truck traffic, noise, or environmental degradation.

The well locations in both of these counties are shown in Figure 3 panels (a) and (c). The wells in Bradford County are somewhat uniformly scattered across the county, while the wells in Lycoming County generally follow the natural geological formation of the Marcellus Shale. It is clear, however, that in Lycoming County none of the shale wells are located exactly on the natural geological boundary of the Marcellus Shale. It may not be efficient to construct a multi-well pad right on the geological boundary as that would restrict horizontal drilling only to one direction; also the shale thickness becomes thinner at the edges. Therefore, when constructing our treatment and control groups of properties in Lycoming County, there are different ways to define the boundary of the Marcellus Shale. Our first definition is a strict geological boundary. Our alternative measure of the Marcellus Shale boundary is the ‘effective’ boundary that we draw on the map in Figure 3 with a dotted line. This effective boundary represents the apparent boundary that exists because gas producers do not want to drill particularly close to the geological boundary of the shale. It is not a priori clear which boundary – the geological boundary or the effective boundary – is more important for identifying any negative effects of natural gas production. Hence, we consider both definitions in our analysis.

4.3 Data Summary Statistics

Summary statistics for both counties are presented in Table 1. For Bradford County, we have a total of 3316 observations, and for Lycoming County we have 4140 total observations. We see that the average sale price in Bradford County is about \$119,000, whereas the property values are considerably higher on average in Lycoming County (\$138,000). These property transactions primarily occur from 2007-2012 as we look at a primarily cross sectional analysis of proximity to unconventional wells. Further, the property values in Lycoming County that are located over the shale are, on average, higher than those not located over the shale.

The density of shale gas wells in Bradford County is clearly higher than in Lycoming County, as the number of wells within each radius is substantially higher in Bradford County relative to Lycoming County. This is also apparent in Figure 3 in which one can see the location of wells throughout Bradford and Lycoming Counties. The distance to nearest well in Bradford County is, on average, substantially lower than in Lycoming County at just over 4 miles to nearest well.

To provide a more complete picture, we plot the kernel density of the distance to nearest well in Figure 4 for both counties; it is clear that the distribution of the distance to nearest well in Bradford County is shifted to the left relative to Lycoming County, with the median and mode of the data being 2.310 miles and 1 mile, respectively. The median and mode for Lycoming County are about 5.725 miles and 5 miles, respectively. These statistics are expected since the density of unconventional gas production in Pennsylvania is highest in Bradford County.

The homes in Bradford County are relatively older, but we do not see much difference in the average size measures of each home across counties. We see that properties in Bradford County are substantially more rural (higher acreage), with a substantially larger share of properties relying on private sources of water (about 50 percent). Properties sold in Bradford County are, on average, about 5.5 miles from the nearest major highway, but only about 0.3 miles from the nearest roadway. Conversely, properties in Lycoming County are only about 2.4 miles from the nearest major highway, and about 0.2 miles from the nearest roadway.

To account for the possible presence of other natural resource activities, we report the average distance of each property to abandoned mines and coal mines. As described in our introduction, many properties in these counties are located relatively far from any other mining operations, in contrast to counties in Western Pennsylvania that have been the focus of previous research. We do not see much difference in distance to the nearest abandoned mine across the two counties, with the average distance being about 20 miles. The average distance to nearest coal mine in Bradford County is about 17 miles, and almost 30 miles in Lycoming County. It is not likely that these types of mining operations have a significant impact on the relationship between property values and shale gas in our area of study. It is important to emphasize that it is not known how the presence of these activities influences the shale environment. It is possible that these activities are simply part of the contextual environment that perhaps precludes external validity of results derived in their context. But, it is also possible that these activities are potentially confounding effects, and studies that fail to properly manage their influence econometrically will report biased parameter estimates. We prefer to err on the side of caution and develop our estimates in the absence of these effects.

In Bradford County, roughly 36 percent of the observations come from the (small) cities Sayre and Towanda, with the rest of the observations being rural. A relatively larger share of the Lycoming County property sales come from small urban areas, with about 63 percent of observations coming from Duboistown, Montoursville, South Williamsport and Williamsport. In Lycoming County, the largest share of urban observations are located in Williamsport.

5 Results

We present results from the different econometric strategies laid out in Section 3. First, we present results of the basic semi-log regressions, and then turn to results from matching models that estimate the *ATT*. Next we show semi-parametric results which address the possibility that the semi-log regression models are overly restrictive in the assumed linearity, after which we present results that exploit the border of the shale in Lycoming County. Finally we discuss many robustness checks that are detailed in the appendix to this paper. We have laid out the results in this way to begin with the simplest model, and then build upon that model in terms of differing complexity to address its potential weaknesses one at a time.

5.1 Semi-Log Regressions

First, we explore the basic semi-log model given in (3). We systematically consider four separate shale activity regressors: miles to nearest well, and the number of wells within 1, 2, and 3 miles of the property. Each regression includes a full set of hedonic housing variables, spatial controls for alternative resource activities and distance to nearest major highway and road, as well as dummies controlling for unobservable census tract, city, and year effects. We report our regression results for Bradford County in Table 2 and for Lycoming County in Table 3.

In the first column in Table 2 that excludes any shale treatment variables, we see that the relationship between our conditioning variables and housing prices are generally in line with our expectations. We find similar estimates in the first column of Table 3 that uses data from Lycoming County. In columns 2 through 5 in Tables 2 and 3, we see that our shale activity variables are not significant in either Bradford or Lycoming Counties. Our estimates indicate that, given our set of conditioning variables and controls for unobservables, shale activity does not adversely affect property values.

5.2 Water Source Interactions

The nature of the potential negative externalities of shale activity suggests that property owners that rely on private water sources (e.g., a well) are at greater risk as private water sources are not as closely monitored as public sources. We expect that if such externalities are significant, the effects will be more severe for properties that rely on private water. To address this issue, we interact our shale treatment variables with the indicator for water source and report these results in Tables 4 and 5 for Bradford and Lycoming Counties.

We report in Table 4 that the distance to nearest well becomes statistically significant after including the interaction term; however the interaction term itself is not significant. The significant coefficient on the distance to nearest well means that property values increase by about 1.3 percent with each additional mile from the nearest shale well. We also find that the coefficient on the number of wells within a 3 mile radius of the property becomes significantly negative, and the interaction with water source is positive and significant. These estimates mean that properties decrease in value by about 0.28 percent with each additional well placed within 3 miles. The marginal effect of an additional well within 3 miles is the sum of these two coefficients for properties that rely on well water, which is -0.0006 and is statistically insignificant with standard error 0.0023. This means that property values for properties that rely on well water are not significantly impacted by the number of wells within 3 miles, while properties on public water are impacted. We interpret these conclusions with caution as our indicators and interactions for the number of wells within 1 and 2 miles are not significant.

For the Lycoming County water source interactions, we find that only the interaction between distance to nearest well and water source is significantly negative (Table 5). All other terms are insignificant. This significant interaction indicates that properties that do not rely on well water are not significantly impacted by shale activity. The marginal effect of distance to nearest shale well for properties that rely on well water is the sum of these two coefficients, -0.0021 , but is also insignificant with standard error 0.0061. Hence, we do not find any significance that water source is an important source of risk of shale activity in Lycoming County.

5.3 Matching Results

5.4 One-to-One Matching

For our matching estimates, we allow only a single control unit for each treated unit; unreported results consider a one-to-four matching strategy. Here, treatment is defined as a binary indicator of shale activity within 1, 2, or 3 miles of the property. We deploy both propensity score matching and multivariate matching; in the propensity score matching models we report standard *ATT* estimates, with standard errors coming from the Abadie & Imbens (2016) correction. Multivariate matching models report bias adjusted estimates (Abadie & Imbens 2011).

5.4.1 Bradford County

Table 6 shows that there is statistical significance of the *ATT* for the propensity score matching model when treatment is defined via 1 or 3 miles, and at 3 miles in the multivariate matching model. The estimates at 3 miles using both matching estimators is very large.

Tables in the appendix show post-match balancing statistics for the Treatment 1, Treatment 2, and Treatment 3 models, respectively. Both the propensity score matching and multivariate matching procedures have balanced the covariates to a great extent when comparing the balancing statistics to pre-matched data. After matching, the normalized differences and log differences in dispersion is close to zero for nearly all covariates. The propensity score is better balanced in the propensity score model than in the multivariate matching model, but both models have greatly reduced the imbalance compared to pre-match.

These results show inconsistency between the estimates of propensity score matching and multivariate matching models, though there is negative significance of negative 15 percent from the propensity score model.

5.4.2 Lycoming County

Table 7 reports the one-to-one matching *ATT* estimates for Lycoming County. We find a significantly negative effect of shale activity within 1 mile of the property using the multivariate matching estimator. Otherwise, we do not find any statistical significance of the *ATT* using either type of matching.

Post-match balancing metrics are reported in tables in the appendix. As was the case in Bradford County with one-to-one matching, both the propensity score matching and multivariate matching methods substantially improve balance in the sample. For instance, for all covariates, the normalized differences and log differences in dispersion measurements substantially decrease. It is worth noting that while the multivariate matching balance greatly improves post-matching, several covariates (age and highway) are still relatively different in normalized mean, and distance to nearest road has a relatively large dispersion. While we find statistical significance of *ATT* from this model, some covariates remain slightly out of balance.

The result from the multivariate matching suggest a 15 percent reduction in home prices with 1 mile proximity to shale gas development, though the propensity score matching model finds insignificance.

5.4.3 Additional Matching Models and Discussion

Though there appears to be some significance in estimates using matching as an estimator, we find troubling differences between different matching estimates that fail to be robust. We also extend this analysis in the appendix by matching on different covariates, and look at heterogeneity using interactions with our groundwater variable. All provide mixed evidence that does not robustly find evidence for significance of the external effect of being within proximity of a well in either Bradford County or Lycoming County.

5.5 Semiparametric Specification

Recent research has indicated that hedonic price functions may be highly nonlinear and difficult to parameterize (e.g., Parmeter et al. 2007); it is not clear how any econometric functional misspecification will manifest in our results. Consider, for example, several parametric specifications of the distance to nearest well measure for Lycoming County. In a simple linear semi-log setup, we find that distance to nearest well is not significant. The parameters in a semi-log specification that is quartic in distance to nearest well is statistically significant up to the cubic term. A quadratic specification yields insignificance, while a cubic specification finds significance for the cubic polynomial term. In short, the relationship may be nonlinear, and while there are standard parametric methods for specifying nonlinearities, it is not clear which specification returns estimates that accurately portray the relationship between shale activity and housing values. Hence, we turn towards a semi-parametric specification in order to alleviate the burden of sorting through multiple, potentially misspecified, nonlinear specifications.

We report our semi-parametric regression estimates of the gradient of the nonparametric function in Figure 5. Panels (a) and (c) plot the estimated gradient function, and the dotted line above and below are bootstrapped confidence intervals. In panels (b) and (d) we consider differential gradients with respect to water source (this is done via interaction of distance to nearest well and our binary water indicator inside the nonparametric part of the function). For these panels, we do not plot confidence bounds in order to keep the differential effects visually distinguishable. Note that for any point in the support of the gradient function, the interpretation of the function value is identical to that of a homogeneous parameter in a standard parametric specification. That is, at any particular point, the gradient function value indicates the marginal increase in the price of a property given a marginal increase in the distance to nearest well.

From the figure, we can see that the gradient function for Bradford County is nonlinear, posi-

tive, and marginally significant at a 5 percent level at relatively short distances, but insignificant at relatively larger distances. The marginal decrease in property values from shale activity for properties located about 1 or 2 miles from a shale well is about \$5,000, while this marginal effect goes to zero around 4 miles. This result indicates that there may be statistically significant – localized – environmental impacts of shale gas activity. Panel (b) shows that properties that have access to public water have a parallel, but higher gradient compared to well water homes. Confidence bands around these functions overlap, indicating that these gradient differentials may not be significantly different from each other. The private water gradient is never significant, and the public water gradient is only significant at very close range.

A similar plot is given in panels (c) and (d) of Figure 5 for Lycoming County, but there we fail to find any statistical significance in distance to nearest well. At a range of up to 1 mile, we can see a markedly positive gradient function estimate, but it is not significant. We conjecture that the relatively lower density of shale well activity in Lycoming County, compared to Bradford County, renders our estimates less precise (see Figure 4). In addition, we do not find any significance in the water source interaction, and we do not see any significant differences in the estimated gradient functions across water source.

The lessons we draw from the semi-parametric model generally indicate that there may exist a nonlinear relationship between unconventional shale activity and property values. As evidenced by several parametric forms of nonlinearity (i.e., quadratic, cubic, quartic), this relationship may be difficult to parameterize, and for this reason we prefer the semi-parametric approach to alleviate this concern. Recall that, given our identification strategy in Bradford County, our shale activity estimates may be cautiously interpreted as a lower bound. Hence, our semi-parametric model for Bradford County indicates that there may be statistically identifiable external effects of shale activity of at least \$5,000, or about 5 percent of the value of the average property. This estimate is qualitatively similar to previous estimates. However, our findings are different in that there does not appear to be a significant water source differential.

5.6 Exploiting the Shale Boundary

We focus on Lycoming County, and exploit the geological boundary of the Marcellus Shale for identification of the negative effects of unconventional gas extraction. We report these results in Table 8. Recall our definitions of treatment groups as being groups of homes that are, for example, located within r miles of a gas well but are not located above the shale (Group 1). In Table 8, we vary r to cover distances of $\{1, 2, 3\}$ miles. Each of these regressions has a full set of hedonic and

spatial controls, as well as indicators to account for unobservable census tract, city, or year effects.

The top panel in Table 8 shows the coefficient estimates on our treatment group indicators relative to Group 4 as the control group using the natural geological boundary of the Marcellus Shale to determine groups. Notice that for $r = \{1, 2\}$, Group 1 is not defined because there are no properties located within that range. This is not unexpected because, as mentioned previously and shown in Figure 3, most shale wells are not located particularly close to the geological boundary in order to maximize the shale activity from each particular well-pad. We only find statistical significance of Group 3 when using the natural geological boundary of the shale to separate our treatment and control groups. This group refers to those properties that are located over the shale and within r miles of a gas well. Hence, this coefficient measures the net effect of shale gas drilling. Our estimates suggest that this net effect is positive and significant, and may range from about 1 to 9 percent of property values. This finding is inconsistent with net costs found in some studies but could be explained by the expectation of potential royalty payments (Boslett et al. 2015), which may outweigh any possible negative effects of nearby drilling.

To address the potential issue that the natural geological boundary is not the appropriate boundary to consider when constructing these groups, since well pads are not located right along the geological boundary, we also explore treatment group regressions using two different measures of shale boundary. The first is an effective shale boundary, which is constructed using GIS to identify the existing boundary at which shale activity appears to operate (see also the dashed line in Figure 3). The second adds a one mile buffer to the effective shale boundary. These results are reported in the middle and lower panels of Table 8. In each of these models, we are able to identify homes in each treatment group, regardless of the definition of r .

We continue to find insignificance of our treatment group indicators for both panels. However, when considering the effective shale boundary, we find that Group 1 has a positive and significant coefficient (at a 10 percent level) for both $r = \{1, 2\}$. This result indicates that homes located near shale activity, but not over the shale such that the property owner might expect an opportunity to lease shale drilling rights, have relatively higher values than homes located far from all shale activity. This result is not consistent with our prior expectation, but it also does not appear to be robust across regressions. This result, however, is not found in our regressions using the effective shale with a one mile buffer to define treatment groups.

To explore sensitivity of these regressions to interaction between treatment groups and well water status of each property, we re-ran each of these regressions including well water interactions with each treatment group. In general we find that there are not any significant interactive effects,

nor does the inclusion of the interaction term alter the qualitative estimates of the other parameters in the model. It is important to recognize that, despite the uniqueness of Lycoming County in that there is a shale boundary as well as a relatively large amount of shale activity, we do not find a negative and significant impact of shale activity on property values. One possible reason for a null result using this approach is that the border identification strategy will only detect externalities at larger distances than actually may be present. There are very few observations that are off-shale, within 1 mile of permitted well, and rely on private well water. We have more identifying observations in the 2 mile proximity band, which may provide more reliable results. However, if the negative externality only exists at less than 2 miles, then this approach will have difficulty with identification. Alternatively, the fact that we do not find negative externalities at larger distances, which is consistent with other findings in previous research, reinforces the evidence that negative externalities may not be present at these larger distances from proximity of wells.

5.7 Further Robustness Checks

Our results have so far indicated that there is not robust statistical significance of shale activity on nearby property values, with some tentative evidence in Bradford County suggesting that unconventional gas extraction may have a negative impact in close proximity, or with an increase in well intensity. Yet, we do not find evidence that these effects are particularly robust. We also note that regressions using census tract measurements instead of general census tract fixed effects do not yield qualitatively different results. We now present some additional regression results considered for robustness. Results are not reported here, but are available in the Appendix.

We explore the impact of shale well permitting on property sales occurring within 60 days of the issuance of the shale well permit. Our regressions are motivated out of concern that shale gas activity has a localized effect over time. The results indicate that homes that rely on private water are not significantly impacted by additional permitted wells, but properties that are on public water are adversely impacted in Bradford County. Unlike in Bradford County, we do not find any statistical significance of shale activity on home values when focusing exclusively on sales occurring within 60 days of the issuance of a shale well permit. It seems, at least in Lycoming County, that there do not exist significant effects over shorter time horizons.

One robustness check is to remove all city properties from the regressions. Figure 3 clearly shows, particularly for Lycoming County, that many properties are clustered in a small urban area. While these cities are small and we have controlled linearly for any potential urban or city effects in our regressions, our results may still be biased if there exists any nonlinear city or urban

effects. Furthermore, urban properties may be better insulated from shale activity, and including many of these properties in a mean regression may wash away any potential environmental effects felt specifically by rural properties. Despite the potential for city effects to influence our primary results, removing all city observations does not qualitatively alter our regression results.

We also remove all properties in Lycoming County that are not over the shale or effective shale, and re-estimate our semi-log specification. We consider this regression to make a similar comparison to Bradford County and other studies by eliminating the shale boundary from the analysis. Our results from these regressions do not indicate any qualitative change from our previous regressions.

We also supply basic matching estimates for both Bradford and Lycoming County and vary the control units that are matched on. We also provide balancing data and describe improvements in balancing pre-/post-match for each set of results in the appendix and for the matching results presented in the main text of this paper. In general, while matching improves the balancing between treated and non treated sample, we do not find robust evidence that proximity to shale gas wells is statistically robust.

Another strategy that may be successful in dealing with unobservables and omitted variable bias in estimation is a difference-in-difference strategy. We use properties within one mile of a well, differencing before and after the permit of the well, and a control group that is between 2 and 3 miles from a well, differencing before and after the permit of the well. We find that using this quasi difference-in-differences strategy, and exploring other distances to define the control group, does not provide robust estimates of negative effects of properties being within proximity to shale gas development. A similar null result is obtained using a triple differencing approach that uses private well water as the third variable.

Our final sets of regressions include defining shale activity simply as incidence of well pads - that is, instead of focusing on unconventional gas extraction wells, we simply count each well-pad as a single incidence of shale activity. Second, we eliminate all conventional gas extraction wells from our dataset, allowing us to focus exclusively on horizontal wells. Using only horizontal wells does not uncover any new findings. Our well-pad models for Bradford County do not return any significance of any shale activity parameters; however we find that the number of well pads within 3 miles of a property is positive and statistically significant in Lycoming County, but becomes insignificant when considering interactions with water source.

6 Conclusion

The results from this study are in contrast to previous findings that identified statistically significant effects of shale activity on property values (Muehlenbachs et al. 2012, Muehlenbachs et al. 2015, Gopalakrishnan & Klaiber 2014). Using a variety of parametric, semiparametric, and matching hedonic regression models, and utilizing the natural and effective borders of the shale as identification tools, we fail to find robust significance of negative environmental externalities of unconventional natural gas extraction. We find some evidence that there may be significant effects that are localized geographically to the placement of the shale wells in Bradford County. Other studies have shown significant and robust negative effects of shale gas development, particularly with properties that are close to shale gas wells and rely on private sources of groundwater. Our evidence of these effects is substantially weaker than that of other studies, and is not robust.

We find significance in a few of our models, such as in the some of the matching models and in Bradford County using semiparametric methods. And in many cases we retrieve estimates with the expected sign, though not always. The troubling aspect of these results is that they do not sustain across our two counties or across similar methods.

We point out that there are several possible explanations for these results. First, as we have discussed, we have carefully selected a study area in which the possibility that alternative mining operations confounds our estimates is minimized. We have further controlled for the distance to alternative mining operations, but we report that regressions excluding these mining variables (not reported) yield qualitatively similar estimates to our main results. Therefore, our estimates are more credibly free from these alternative operations that may effect estimates of the relationship between shale activities and property values. Additionally, our use of the shale border should cleanly separate the benefits from mineral rights that could be complicating other analysis, though this depends on the extent that the externality is localized at greater distances than 1 mile. As we have argued, it is not clear whether these effects confound estimates if inadequately controlled for, and may be a factor driving previous results. Second, the importance of nonlinearities and more complex functional forms that are consistent with hedonic theory (Parmeter et al. 2007, Parmeter & Pope 2013) may be causing differences between the additive linear models in other regions, as distance to wells appears to enter nonlinearly into the hedonic price function. Third, results from the studies in Pennsylvania so far are difficult to generalize across the region at large, which may speak to the external validity of those results to our study counties. Lastly, it is possible that external effects are not fully realized in this study area because of the lack experience in natural resource

extraction and there is learning of relevant external costs and benefits to properties possessed by the buyers and sellers at the time of transactions.

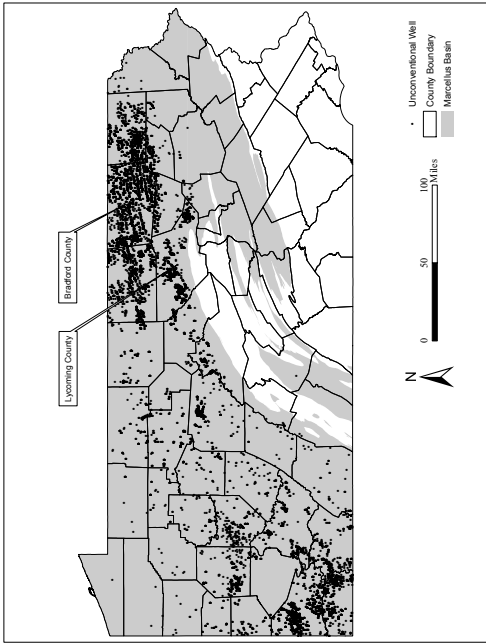
Though our result that shale operations are largely insignificant factors of property values, except perhaps at relatively short distances, seems robust, our work does not necessarily imply that there are no external costs to unconventional natural gas extraction. In particular, there may be long run costs that are not apparent at this point in time, or asymmetric information about risks that are not made public to property buyers. Further, there may be other social costs associated with unconventional gas extraction that may be relevant but not investigated here, such as impacts on crime rates, housing availability (i.e., shortages), or health effects. See Muehlenbachs et al. (2015) for a first analysis of some of these potential effects.

These results are of particular relevance for policymakers as shale gas exploration is rapidly expanding in many areas of the world. Our results have highlighted the importance of study area and potential confounding factors associated with different study areas; model specification and econometric functional form; and identification strategy as important dimensions to consider when evaluating the external costs of unconventional gas extraction. Policymakers should carefully consider the implications of these issues when developing policies surrounding unconventional natural gas extraction in both the United States and abroad.

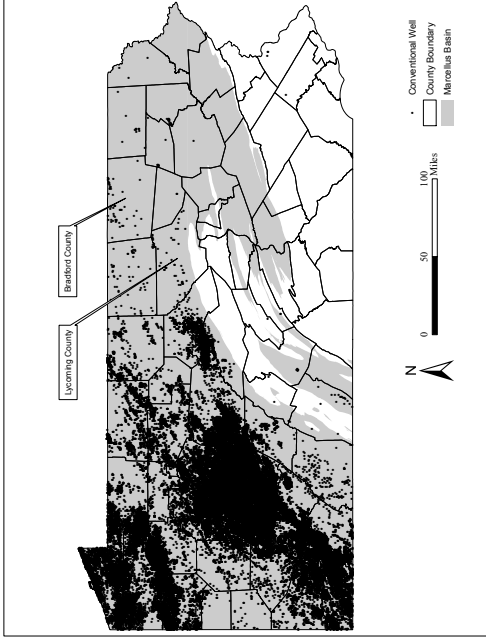
References

- Abadie, A. & Imbens, G. (2006), ‘Large sample properties of matching estimators for average treatment effects’, *Econometrica* **74**, 235–267.
- Abadie, A. & Imbens, G. (2011), ‘Bias-corrected matching estimators for average treatment effects’, *Journal of Business and Economic Statistics* **29**, 1–11.
- Abadie, A. & Imbens, G. (2016), ‘Matching on the estimated propensity score’, *Econometrica* **84**, 781–807.
- Boslett, A., Guilfoos, T. & Lang, C. (2015), ‘Valuation of expectations: a hedonic study of shale gas development and New York’s moratorium’, *Journal of Environmental Economics and Management* **77**, 14–30.
- Brasier, K. J., Filteau, M. R., McLaughlin, D., Jacquet, J., Stedman, R. C., Kelsey, T. W. & Goetz, S. J. (2011), ‘Residents’ perceptions of community and environmental impacts from development of natural gas in the Marcellus Shale: a comparison of Pennsylvania and New York cases’, *Journal of Rural Social Sciences* **26**, 32–61.
- Drohan, P. J., Brittingham, M., Bishop, J. & Yoder, K. (2012), ‘Early trends in landcover change and forest fragmentation due to shale gas development in Pennsylvania: a potential outcome for the north central Appalachians’, *Environmental Management* **49**, 1061–1075.
- Fitzgerald, T. (2013), ‘Frackonomics: some economics of hydraulic fracturing’, *Case Western Law Review* **63**, 1337–1362.
- Gopalakrishnan, S. & Klaiber, H. A. (2014), ‘Is the shale energy boom a bust for nearby residents? Evidence from housing values in Pennsylvania’, *American Journal of Agricultural Economics* **96**, 43–66.
- Gregory, K. B., Vidic, R. D. & Dzombak, D. A. (2011), ‘Water management challenges associated with the production of shale gas by hydraulic fracturing’, *Elements* **7**, 181–186.
- Heintzleman, M. D. & Tuttle, C. M. (2012), ‘Values in the wind: a hedonic analysis of wind power facilities’, *Land Economics* **88**, 571–588.
- Henderson, D. J. & Parmeter, C. F. (2015), *Applied Nonparametric Econometrics*, Cambridge University Press.
- Holland, P. (1986), ‘Statistics and causal inference’, *Journal of the American Statistical Association* **81**, 945–970.
- Imbens, G. W. (2004), ‘Nonparametric estimation of average treatment effects under exogeneity: a review’, *Review of Economics and Statistics* **86**, 4–29.
- Imbens, G. W. & Rubin, D. B. (2015), *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge University Press.
- Imbens, G. W. & Wooldridge, J. M. (2009), ‘Recent developments in the econometrics of program evaluation’, *Journal of Economic Literature* **47**, 5–86.
- Jackson, R. B., Vengosh, A., Darrah, T. H., Warner, N. R., Down, A., Poreda, R. J., Osborn, S. G., Zhao, K. & Karr, J. D. (2013), ‘Increased stray gas abundance in a subset of drinking water wells near Marcellus Shale gas extraction’, *Proceedings of the National Academy of Sciences* **110**, 11250–11255.
- Jiang, M., Hendrickson, C. T. & VanBriesen, J. M. (2014), ‘Life cycle water consumption and wastewater generation impacts of a Marcellus Shale gas well’, *Environmental Science and Technology* **48**(3), 1911–1920.
- Kelsey, T. W., Metcalf, A. & Salcedo, R. (2012), ‘Marcellus Shale: land ownership, local voice, and the distribution of lease and royalty dollars’, *The Center for Economic and Community Development, University Park*.
- Kuminoff, N. V., Parmeter, C. F. & Pope, J. C. (2012), ‘Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?’, *Journal of Environmental Economics and Management* **60**, 145–160.
- Ladlee, J. & Jacquet, J. (2012), ‘Research and policy brief series’, *Analysis* **12**, 1.
- Li, Q. & Racine, J. S. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

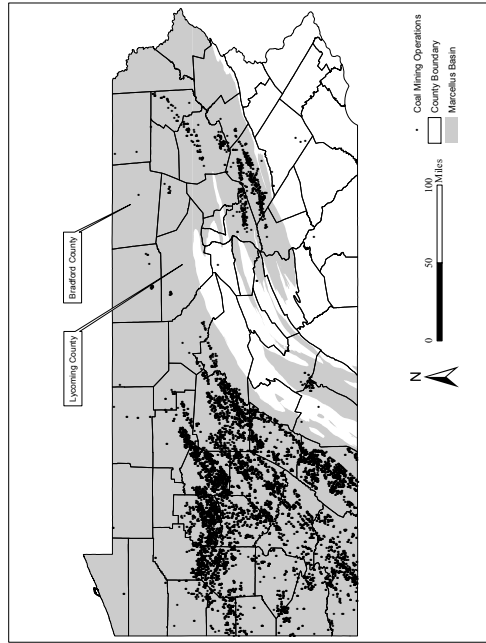
- Litovitz, A., Curtright, A., Abramzon, S., Burger, N. & Samaras, C. (2013), 'Estimation of regional air-quality damages from Marcellus Shale natural gas extraction in Pennsylvania', *Environmental Research Letters* **8**, 14–17.
- Loewenstein, J. (2012), 'Natural gas royalties in Bradford County reach \$160 million', *The Daily Review* .
- Muehlenbachs, L., Spiller, E. & Timmins, C. (2012), 'Shale gas development and property values: differences across drinking water sources', *working paper* .
- Muehlenbachs, L., Spiller, E. & Timmins, C. (2015), 'The housing market impacts of shale gas development', *American Economic Review* **105**, 3633–3659.
- National Energy Technology Laboratory (2012), *Role of Alternative Energy Sources: National Gas Technology Assessment*.
- Olmstead, S. M., Muehlenbachs, L. A., Shih, J., Chu, Z. & Krupnick, A. J. (2013), 'Shale gas development impacts on surface water quality in Pennsylvania', *Proceedings of the National Academy of Sciences* **110**, 4962–4967.
- Palmquist, R. B. (2005), 'Hedonic methods', *Measuring the Demand for Environmental Quality* .
- Parmeter, C. F., Henderson, D. J. & Kumbhakar, S. C. (2007), 'Nonparametric estimation of a hedonic price function', *Journal of Applied Econometrics* **22**, 695–699.
- Parmeter, C. F. & Pope, J. C. (2013), 'Quasi-experiments and hedonic property values methods', *Handbook on Experimental Economics and the Environment* pp. 3–67.
- Pender, J. L., Weber, J. G. & Brown, J. P. (2014), 'Sustainable rural development and wealth creation: Five observations based on emerging energy opportunities', *Economic Development Quarterly* **28**, 73–86.
- Rahm, B. G., Bates, J. T., Bertoia, L. R., Galford, A. E., Yoxthimer, D. A. & Riha, S. J. (2013), 'Wastewater management and Marcellus Shale gas development: trends, drivers, and planning implications', *Journal of Environmental Management* **120**, 105–113.
- Robinson, P. M. (1988), 'Root-n consistent semiparametric regression', *Econometrica* **56**, 931–954.
- Rosen, R. (1974), 'Hedonic prices and implicit markets: product differentiation in pure competition', *Journal of Political Economy* **82**, 34–55.
- Rosenbaum, P. R. & Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**, 41–55.
- Rubin, D. (1974), 'Estimating causal effects of treatments in randomized and non-randomized studies', *Journal of Educational Psychology* **66**, 688–701.
- Taylor, L. (2003), The hedonic method, in P. Champ, T. Brown & K. Boyle, eds, 'A Primer on the Economic Valuation of the Environment', Kluwer, pp. 331–394.
- United States Energy Information Administration (2012), *Annual Energy Outlook*.
- Warner, N. R., Jackson, R. B., Darrah, T. H., Osborn, S. G., Down, A., Zhao, K., White, A. & Vengosh, A. (2012), 'Geochemical evidence for possible natural migration of Marcellus Formation brine to shallow aquifers in Pennsylvania', *Proceedings of the National Academy of Sciences* **109**, 11961–11966.
- Weidner, K. (2013), 'A landowners guide to leasing land in Pennsylvania', *Pennsylvania State University Cooperative Extension* .
- Wrenn, D. H., Klaiber, H. A. & Jaenicke, E. C. (2015), 'Unconventional shale gas development, risk perceptions, and averting behavior: evidence from bottled water purchases', *Risk Perceptions, and Averting Behavior: Evidence from Bottled Water Purchases (March 20, 2015)* .



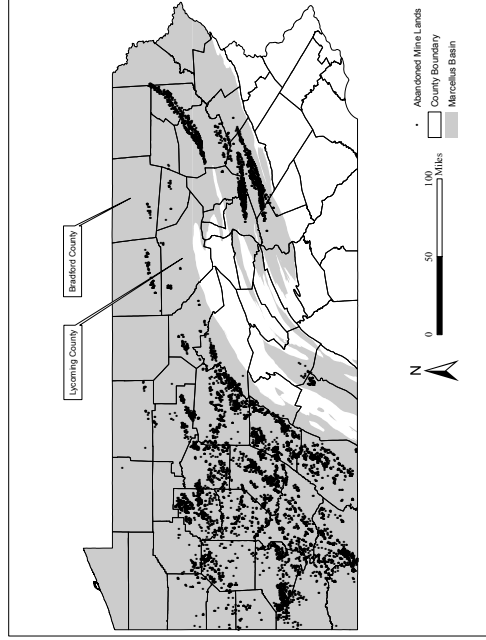
(a) Unconventional well operations



(b) Conventional well operations



(c) Coal mining operations



(d) Abandoned mine lands

Figure 1: Maps of locations of natural resource extraction in Pennsylvania.

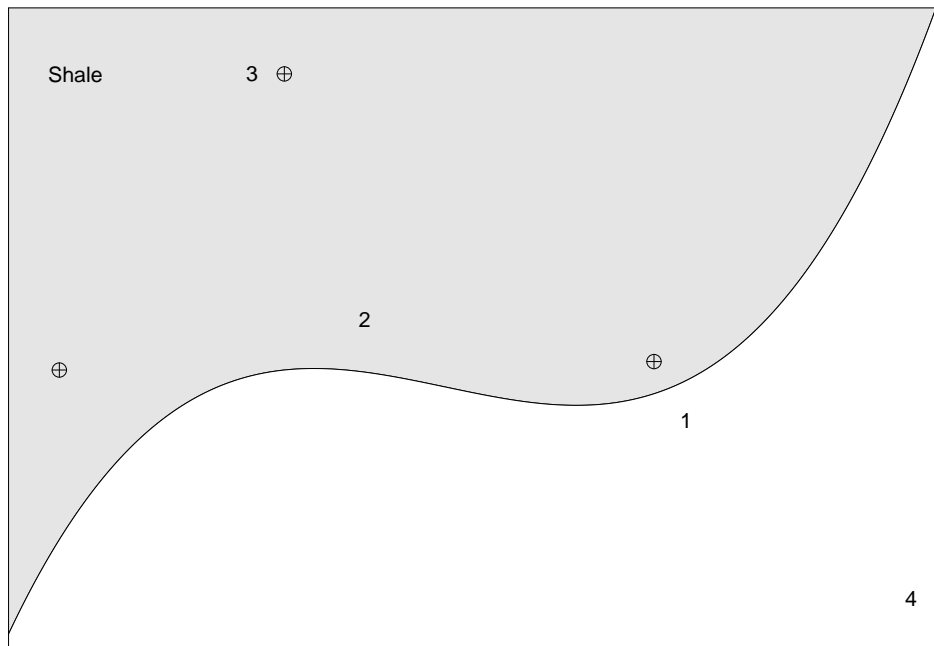
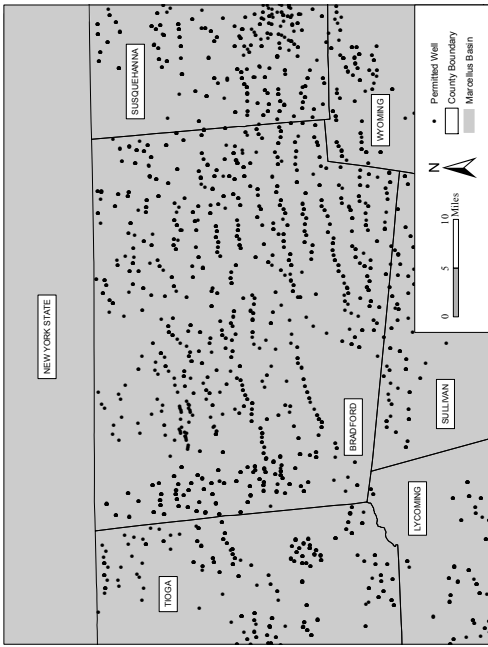
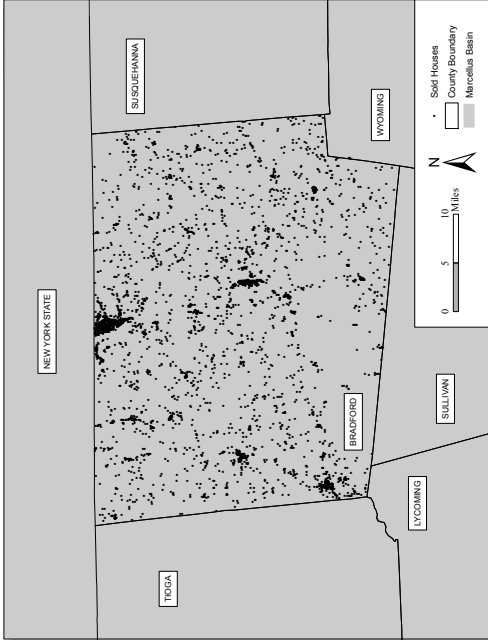


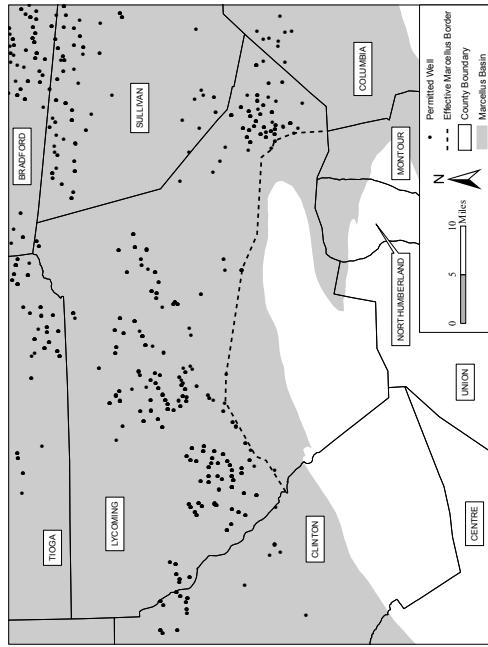
Figure 2: Shale diagram depicting different property exposure groups in Lycoming County. The grey shaded area depicts the shale region, ‘⊕’ symbolizes gas wells, and the location of the numbers in proximity to either the shale or gas wells depicts the approximate location of properties in each group in {1, 2, 3, 4}. For example, Group 1 is off the shale, but near a well; Group 2 is on the shale, but not near a gas well. See text for further details.



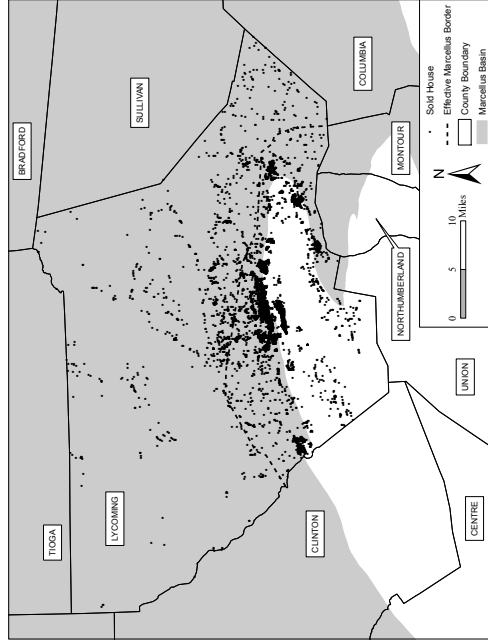
(a) Location of permitted wells: Bradford County



(b) Location of properties sold: Bradford County

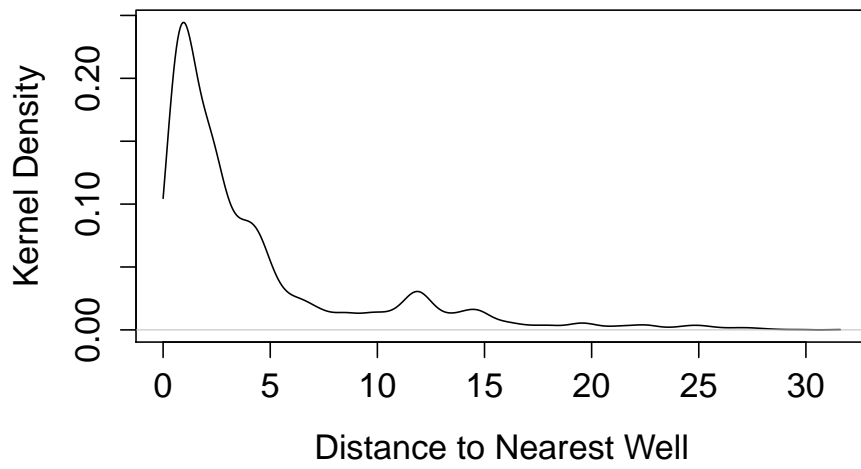


(c) Location of permitted wells: Lycoming County

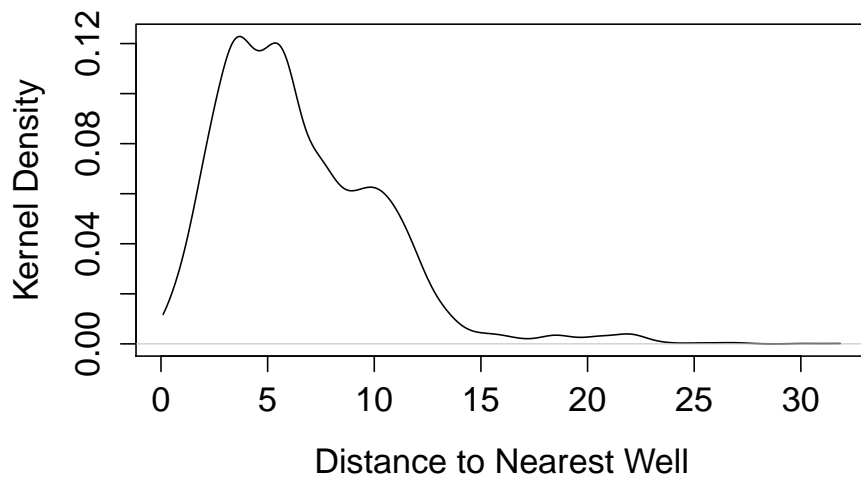


(d) Location of properties sold: Lycoming County

Figure 3: Maps showing the location of permitted wells and properties sold for both Bradford and Lycoming Counties.

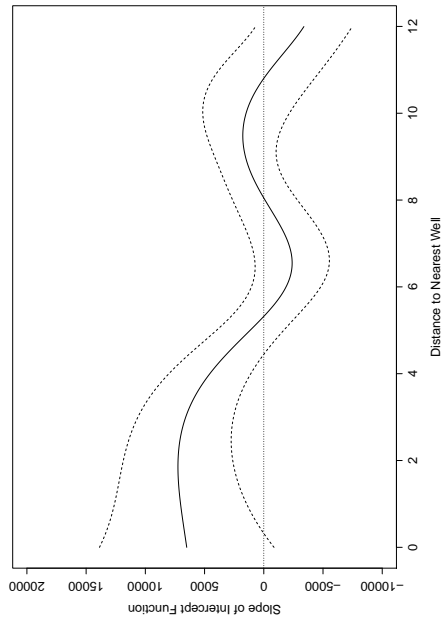


(a) Bradford County

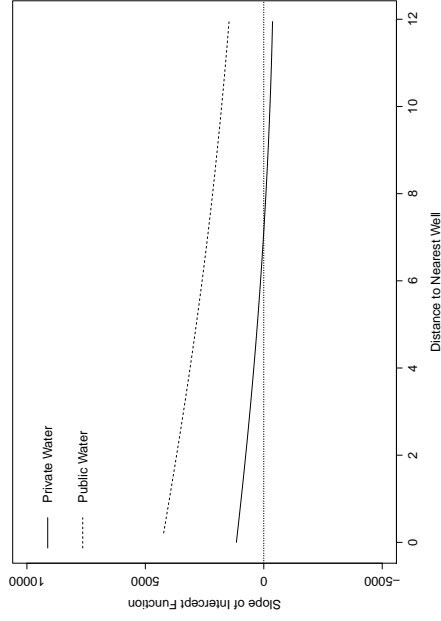


(b) Lycoming County

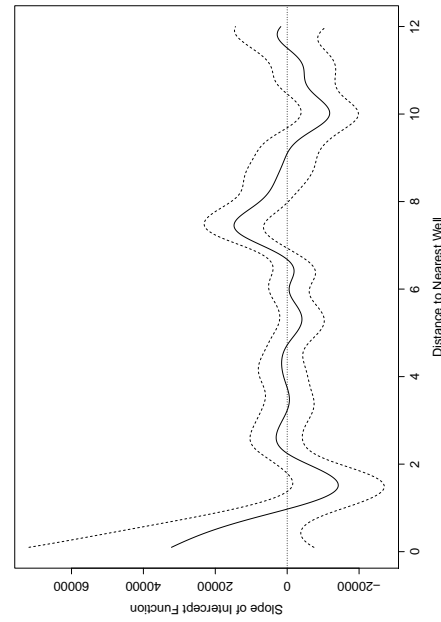
Figure 4: Kernel density plots of the distance to nearest well in miles for Bradford and Lycoming Counties.



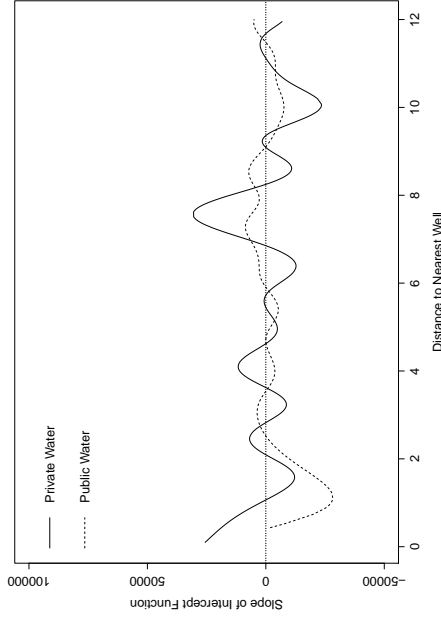
(a) Semiparametric gradient: Bradford County



(b) Semiparametric gradient by water source: Bradford County



(c) Semiparametric gradient: Lycoming County



(d) Semiparametric gradient by water source: Lycoming County

Figure 5: Semiparametric estimates of the marginal effect of distance to nearest well in miles on housing values for Bradford and Lycoming Counties. The dotted lines in panels (a) and (c) are bootstrapped 95 percent confidence intervals around the gradient function estimate. Panels (b) and (d) display the differential gradient function estimates for homes on private versus public water. In those panels, we do not include bootstrapped confidence intervals for clarity of the figures; plots including these confidence intervals indicate no statistical significance of the differential gradient functions and are available from the authors.

Table 1: Sample averages for Bradford and Lycoming Counties. Lycoming County averages are across the entire sample and subgroups corresponding to location over the shale.

Variable	Bradford County	Lycoming County		
		Total Sample	On-Shale	Off-Shale
Outcome				
Sale Price	119022.951	138195.812	139712.478	133089.060
Natural Gas				
Wells in 1 Mile (count)	1.714	0.082	0.106	0.000
Wells in 2 Mile (count)	7.661	0.396	0.513	0.000
Wells in 3 Mile (count)	18.985	1.078	1.350	0.159
Distance to Nearest Well (miles)	4.338	6.502	6.152	7.681
Housing				
Age (years)	71.104	63.773	62.462	68.188
Stories	1.512	1.531	1.536	1.515
Bedrooms	2.984	2.973	2.993	2.906
Bathrooms	1.471	1.529	1.552	1.452
Living Area (sq feet)	16.103	16.597	16.787	15.957
Basement (binary)		0.142	0.144	0.135
Garage (binary)		0.068	0.067	0.073
Central Air (binary)		0.246	0.255	0.215
Forced Air (binary)	0.485			
Acreage	8.157	1.867	1.948	1.595
Private Water (binary)	0.564	0.290	0.307	0.233
Spatial				
Abandoned Mine (miles)	22.282	20.422	19.422	23.789
Coal Mine (miles)	17.141	28.445	29.081	26.516
Major Highway (miles)	5.518	2.423	2.704	1.478
Road (miles)	0.239	0.212	0.211	0.218
Cities (binary)				
Sayre	0.209			
Towanda	0.145			
Duboistown		0.011	0.000	0.047
Montoursville		0.142	0.142	0.143
South Williamsport		0.083	0.000	0.363
Williamsport		0.393	0.476	0.113
Sample				
Sample Size	3316	4140	3192	948

Table 2: Baseline regression results for Bradford County.

Variable	(1)	(2)	(3)	(4)	(5)
Well Distance		0.0057			
		0.0047			
Wells in 1 Mile			0.0005		
			0.0038		
Wells in 2 Miles				-0.0009	
				0.0014	
Wells in 3 Miles					-0.0009
					0.0008
Age	0.0027**	0.0027**	0.0027**	0.0027**	0.0027**
	0.0011	0.0011	0.0011	0.0011	0.0011
Age ²	-0.00002***	-0.00002***	-0.00002***	-0.00002***	-0.00002***
	0.00001	0.00001	0.00001	0.00001	0.00001
Stories	-0.0747*	-0.0762*	-0.0747*	-0.0749*	-0.0754*
	0.0444	0.0444	0.0444	0.0444	0.0444
Bedrooms	-0.0224	-0.0226	-0.0224	-0.0225	-0.0224
	0.0164	0.0164	0.0164	0.0164	0.0164
Bathrooms	0.1723***	0.1731***	0.1723***	0.1725***	0.1726***
	0.0274	0.0274	0.0274	0.0274	0.0274
Area	0.0300***	0.0301***	0.0300***	0.0300***	0.0300***
	0.0032	0.0032	0.0032	0.0032	0.0032
Acre	0.0097***	0.0097***	0.0097***	0.0097***	0.0097***
	0.0008	0.0008	0.0008	0.0008	0.0008
Acres ²	-0.00001***	-0.00001***	-0.00001***	-0.00001***	-0.00001***
	0.0000	0.0000	0.0000	0.0000	0.0000
Forced Air	-0.1061***	-0.1066***	-0.1062***	-0.1061***	-0.1060***
	0.0281	0.0281	0.0281	0.0281	0.0281
Private Water	0.0202	0.0202	0.0197	0.0238	0.0229
	0.0438	0.0438	0.0441	0.0441	0.0439
Abandoned Mine	0.0114**	0.0109**	0.0114**	0.0114**	0.0114**
	0.0050	0.0050	0.0050	0.0050	0.0050
Coal Mine	-0.0099*	-0.0094	-0.0099*	-0.0099*	-0.0099*
	0.0057	0.0057	0.0057	0.0057	0.0057
Major Highway	-0.0157***	-0.0158***	-0.0157***	-0.0157***	-0.0157***
	0.0047	0.0047	0.0047	0.0047	0.0047
Road	-0.0105	-0.0122	-0.0103	-0.0103	-0.0103
	0.0301	0.0301	0.0301	0.0301	0.0301
\bar{R}^2	0.2082	0.2083	0.2080	0.2081	0.2083
Sample Size	3316	3316	3316	3316	3316

Each regression contains a constant term, and uses dummies to control for unobservable census, city, and year effects. Statistical significance is denoted with *, **, *** for significance at the 10, 5, and 1 percent levels, respectively. Standard errors are reported below the coefficients.

Table 3: Baseline regression results for Lycoming County.

Variable	(1)	(2)	(3)	(4)	(5)
Well Distance		0.0018			
		0.0028			
Wells in 1 Mile			-0.0020		
			0.0080		
Wells in 2 Miles				0.0019	
				0.0025	
Wells in 3 Miles					0.0017
					0.0013
Age	0.0009	0.0010	0.0009	0.0009	0.0010
	0.0007	0.0007	0.0007	0.0007	0.0007
Age ²	-0.00002***	-0.00002***	-0.00002***	-0.00002***	-0.00002***
	0.0000	0.0000	0.0000	0.0000	0.0000
Stories	-0.0607***	-0.0610***	-0.0606***	-0.0608***	-0.0608***
	0.0178	0.0178	0.0178	0.0178	0.0178
Bedrooms	0.0010	0.0010	0.0010	0.0010	0.0009
	0.0055	0.0055	0.0055	0.0055	0.0055
Bathrooms	0.1031***	0.1031***	0.1030***	0.1031***	0.1031***
	0.0134	0.0134	0.0134	0.0134	0.0134
Area	0.0258***	0.0258***	0.0258***	0.0258***	0.0258***
	0.0015	0.0015	0.0015	0.0015	0.0015
Basement	0.1165***	0.1165***	0.1165***	0.1167***	0.1168***
	0.0201	0.0201	0.0201	0.0201	0.0201
Garage	0.0168	0.0166	0.0168	0.0169	0.0167
	0.0215	0.0215	0.0215	0.0215	0.0215
Air Conditioner	0.1191***	0.1191***	0.1191***	0.1192***	0.1191***
	0.0168	0.0168	0.0168	0.0168	0.0168
Acre	0.0148***	0.0147***	0.0147***	0.0148***	0.0148***
	0.0020	0.0020	0.0020	0.0020	0.0020
Acres ²	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***
	0.0000	0.0000	0.0000	0.0000	0.0000
Private Water	0.0169	0.0165	0.0167	0.0171	0.0168
	0.0232	0.0233	0.0233	0.0233	0.0232
Abandoned Mine	0.0015	0.0011	0.0011	0.0011	0.0011
	0.0043	0.0044	0.0044	0.0044	0.0044
Coal Mine	-0.0041	-0.0038	-0.0038	-0.0038	-0.0038
	0.0041	0.0041	0.0041	0.0041	0.0041
Major Highway	0.0004	0.0002	0.0002	0.0002	0.0002
	0.0044	0.0044	0.0044	0.0044	0.0044
Road	0.1242***	0.1240***	0.1240***	0.1240***	0.1240***
	0.0279	0.0279	0.0279	0.0279	0.0279
\bar{R}^2	0.4120	0.4119	0.4119	0.4119	0.4121
Sample Size	4140	4140	4140	4140	4140

Each regression contains a constant term, and uses dummies to control for unobservable census, city, and year effects. Statistical significance is denoted with *, **, *** for significance at the 10, 5, and 1 percent levels, respectively. Standard errors are reported below the coefficients.

Table 4: Water interaction regression results for Bradford County.

Variable	(1)	(2)	(3)	(4)
Well Distance	0.0132*			
	0.0071			
Well Distance \times Private Water	-0.0087			
	0.0062			
Wells in 1 Mile		-0.0010		
		0.0089		
Wells in 1 Mile \times Private Water		0.0017		
		0.0094		
Wells in 2 Miles			-0.0041	
			0.0028	
Wells in 2 Miles \times Private Water			0.0036	
			0.0028	
Wells in 3 Miles				-0.0028**
				0.0013
Wells in 3 Miles \times Private Water				0.0022*
				0.0012
\bar{R}^2	0.2086	0.2078	0.2082	0.2089
Sample Size	3316	3316	3316	3316

Each regression contains a constant term, the full set of hedonic controls, and uses dummies to control for unobservable census, city, and year effects. Statistical significance is denoted with *, **, *** for significance at the 10, 5, and 1 percent levels, respectively. Standard errors are reported below the coefficients.

Table 5: Water interaction regression results for Lycoming County.

Variable	(1)	(2)	(3)	(4)
Well Distance	0.0052			
	0.0033			
Well Distance \times Private Water	-0.0073**			
	0.0037			
Wells in 1 Mile		0.0259		
		0.0197		
Wells in 1 Mile \times Private Water		-0.0327		
		0.0212		
Wells in 2 Miles			0.0069	
			0.0063	
Wells in 2 Miles \times Private Water			-0.0058	
			0.0066	
Wells in 3 Miles				0.0033
				0.0036
Wells in 3 Miles \times Private Water				-0.0017
				0.0037
\bar{R}^2	0.4123	0.4121	0.4119	0.4120
Sample Size	4140	4140	4140	4140

Each regression contains a constant term, the full set of hedonic controls, and uses dummies to control for unobservable census, city, and year effects. Statistical significance is denoted with *, **, *** for significance at the 10, 5, and 1 percent levels, respectively. Standard errors are reported below the coefficients.

Table 6: ATT estimates for propensity score and multivariate matching models for Bradford County

Estimate	Treatment 1		Treatment 2		Treatment 3	
	PSM	MM	PSM	MM	PSM	MM
Standard	-0.1515**		-0.0511		-0.4752**	
	0.0782		0.0956		0.2219	
Bias Adjusted		-0.0733		-0.0087		-0.4092***
		0.0720		0.0825		0.0978

Standard errors reported below each ATT estimate. Statistical significance is denoted with *, **, *** for significance at the 10, 5, and 1 percent levels, respectively. Standard errors are reported below the coefficients. The columns Treatment 1, Treatment 2 and Treatment 3 refer to shale activity within 1, 2, or 3 miles of the property.

Table 7: ATT estimates for propensity score and multivariate matching models for Lycoming County

Estimate	Treatment 1		Treatment 2		Treatment 3	
	PSM	MM	PSM	MM	PSM	MM
Standard	-0.0693		0.0221		0.0028	
	0.0842		0.0520		0.0491	
Bias Adjusted		-0.1491*		0.0191		-0.0242
		0.0894		0.0442		0.0413

Standard errors reported below each ATT estimate. Statistical significance is denoted with *, **, *** for significance at the 10, 5, and 1 percent levels, respectively. Standard errors are reported below the coefficients. The columns Treatment 1, Treatment 2 and Treatment 3 refer to shale activity within 1, 2, or 3 miles of the property.

Table 8: Regression results for Lycoming County using treatment groups defined over various definitions of shale location and distance to gas wells.

Variable	Gas Well Proximity in Miles		
	$r = 1$	$r = 2$	$r = 3$
Shale			
TG ₁			0.0129
			0.0658
TG ₂	0.0380	0.0336	0.0353
	0.0594	0.0395	0.0282
TG ₃	0.0075***	0.0937***	0.0574***
	0.0007	0.0007	0.0282
\bar{R}^2	0.4101	0.4105	0.4100
Sample Size	4140	4140	4140
Effective Shale			
TG ₁	0.2096*	0.0839*	0.0133
	0.1171	0.0442	0.0306
TG ₂	-0.0264	-0.0375	-0.0554
	0.0444	0.0462	0.0494
TG ₃	-0.1127	0.0093	-0.0158
	0.0710	0.0549	0.0513
\bar{R}^2	0.4105	0.4104	0.4099
Sample Size	4140	4140	4140
Effective Shale 1			
TG ₁	0.2341	0.0619	-0.0117
	0.4309	0.0548	0.0334
TG ₂	-0.0217	-0.0349	-0.0528
	0.0335	0.0350	0.0370
TG ₃	-0.0497	0.0353	0.0101
	0.0610	0.0444	0.0414
\bar{R}^2	0.4097	0.4103	0.4102
Sample Size	4140	4140	4140

Each regression contains a constant term, the full set of hedonic controls, and uses dummies to control for unobservable census, city, and year effects. Statistical significance is denoted with *, **, *** for significance at the 10, 5, and 1 percent levels, respectively. Standard errors are reported below the coefficients.