THESIS

THE IMPACT OF HYDRAULIC FRACTURING ON HOUSING VALUES IN WELD COUNTY, COLORADO: A HEDONIC ANALYSIS

Submitted by

Ashley Bennett

Department of Agricultural and Resource Economics

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2013

Master's Committee:

Advisor: John B. Loomis

Marco Costanigro
Robin Reich

ABSTRACT

THE IMPACT OF HYDRAULIC FRACTURING ON HOUSING VALUES IN WELD COUNTY, COLORADO: A HEDONIC ANALYSIS

Oil and gas production using hydraulic fracturing has rapidly spread across the US and moved into suburban and urban neighborhoods. Proximity to residential areas has generated significant concerns by homeowners about water pollution, air pollution, aesthetics, and hence property values. However, the increase in drilling activity has generated sizable gains in local employment and a subsequent increase in demand for housing. In spite of controversies, there is almost no research evaluating whether proximity and level of drilling activity affects house prices on net. We apply the hedonic property method to a sample of 4035 housing transactions between 2009 and 2012 in Weld County, Colorado, the county at the forefront of oil and gas drilling activity in the state.

Results across both the semi-log OLS and semi-log spatial GLS model specifications are consistent. While the count of wells being hydraulically fractured within a half mile of a house has a negative effect on houses in Greeley and other towns, rural households are statistically unaffected by the density of hydraulic fracturing in their immediate area. Employment in the oil and gas sector has a positive and significant effect on house prices in the full county and Greeley model specifications, but not in the rural model specifications. The overall lack of negative effect of hydraulic fracturing on housing prices in Weld County may be a result of the increase in employment associated with drilling operations potentially offsetting some of the disamenity associated with oil and gas drilling.

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Dr. John Loomis, for supporting me in this research and throughout my time at CSU. Working closely with Dr. Loomis on this master's thesis and on a survey design as research assistant have been great experiences, from which I have gained a lot of knowledge. I would also like to thank Dr. Marco Costanigro and Dr. Robin Reich for their econometric and spatial statistic insights and serving on my committee. I feel fortunate to have had a committee that was willing to help me with various steps on this unfunded project.

For funding me as a research assistant in my first year, I would like to thank Dr. John Loomis and Dr. Steve Davies; for funding me as a teaching assistant in my second year, I would like to thank the Department of Agricultural and Resource Economics. Dr. Marshall Frasier and Dr. Joleen Hadrich also deserve to be thanked for being flexible and fun to serve under as a teaching assistant.

Finally, I would like to thank my family and friends for their support through these past two years. My mom has always supported me emotionally in my academic and personal endeavors, and I am incredibly grateful for her involvement in my life. I'm also very thankful for the great cohort of graduate students with whom I have spent the past two years in the classroom and grad lab as well as outside of school. In particular, I am thankful Michelle Kibler for being such a great support and contact in Colorado.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

High energy prices in 2008 accentuated the potential risk of dependence on foreign energy suppliers, spiking interest in American energy independence. The United States of America is making significant progress toward its goal of energy independence by the year 2040, at its current pace it is set to become a net exporter of natural gas by 2020 (Koch, 2012). Unconventional gas is a type of gas that is less economical or more difficult to extract (NaturalGas.org). Exploration of unconventional gas – specifically shale exploration – in the United States is rapidly growing the domestic oil and natural gas sector due to technological advancement in the use of hydraulic fracturing, also commonly known as "fracking," and horizontal drilling to extract natural gas from far below the earth's surface. Fracking is used to stimulate natural gas production after a well has been drilled by pumping fracking fluids – a mixture of water and chemicals – into the well at a high pressure to fracture the shale formation (EPA, 2011). Given that the use of these techniques is still relatively new, existing literature on the costs and benefits of their implementation has been sparse. Cooley and Donnelly (2012) found a shortage of peer-reviewed, scientific information on the use of hydraulic fracturing and its environmental impacts. Weber (2012) studied the effects of the gas boom on employment and income in Colorado, Texas, and Wyoming, and found that the number of jobs created by increased natural gas production may be overstated. Despite the lack of peer-reviewed research on the subject, there has been a noteworthy amount of discourse about potential associated risks to public health and water sources taking place in the media - especially along Colorado's Front Range and western slope. In his article *Controversy over fracking runs deep*, Peter Marcus highlights issues on which industry and environmentalists disagree, focusing on the proposed 350-foot buffer to occupied buildings in urban areas (Marcus, 2012). One issue that has been brought up in the literature is whether the economic benefits of shale gas exploration (e.g. job creation, regional multiplier effects, rents paid to landowners) are actually greater than the costs to society (e.g. risk of water contamination, air pollution, visual issues) (Weber, 2012; US Energy Information Adminstration, 2012). Risks associated with hydraulic fracturing, including water quality and quantity issues and air and noise pollution, may be capitalized into housing prices of homes located near drilling sites[1]. A consensus as to whether shale exploration will help or harm the communities in which it is taking place has not yet been met, indicating that further research in on the topic is imperative in order to construct informed policy on the issue.

Lipscomb, Wang, and Kilpatrick (2012) discuss some major issues associated with unconventional shale gas development and potential real estate valuation issues. They assert that two major legal concerns related to property values and unconventional gas development are mortgageability and insurance. Mortgageability issues may arise due to the perceived risks of contamination associated with fracking, which could cause banks to choose not to grant mortgage loans for the property. If mortgageability issues should arise from fracking operations, sales of single-family residential properties with fracking operations on the property may become unsalable. The other legal concern is the lack of insurance coverage for fracking-related claim, because a limitation of homeowner's insurance may negatively affect a future homeowner's chance of buying a house. While Lipscomb et al. (2012) discusses potential issues related to fracking's impact on residential and potential analysis frameworks (hedonic pricing or contingent valuation), it lacks a quantitative analysis; at the same time, it suggests that there is a need for research about the appropriate distance a shale gas well should be located from a property or drinking water source (Lipscomb, Wang, & Kilpatrick, 2012).

[1] See section 2.1 for a more in depth discussion of these risks.

The Colorado Oil and Gas Conservation Commission (COGCC) reports weekly statistics on well production in the state; as of March 7th, 2013, there are approximately 50,000 active oil and gas wells. Weld County has the highest number of active oil and gas wells in the state of Colorado, with approximately 20,000 fractured wells currently in production. It also has the highest percentage of new permits issued for drilling with 48% of new permits issued in 2012.

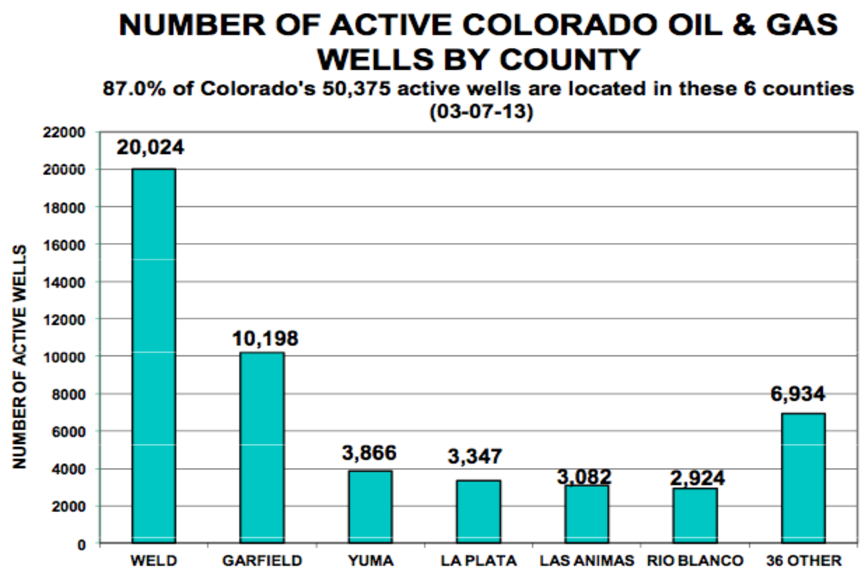

Figure 1.1 Active Oil & Gas Wells in Colorado by County (COGCC, 2013)

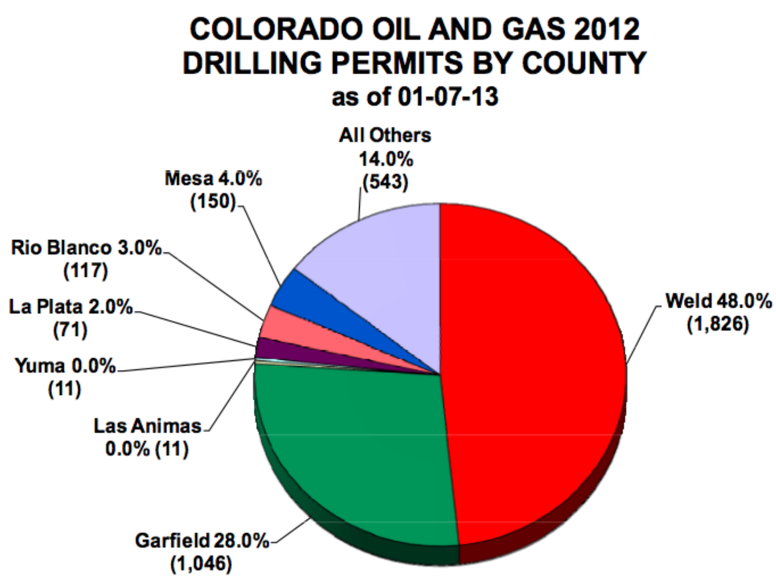

Figure 1.2 Drilling Permits in Colorado by County (COGCC, 2013)

Figure 1.3, a map created in ArcGIS using geographically referenced oil and gas well data downloaded from the COGCC, shows all producing wells (grey points) and all permitted wells (blue points) in Colorado. Note that the density of wells in Weld County is so high that county lines are not visible when looking at the entire state of Colorado, see Figure 1.4 for a closer view of Weld County.

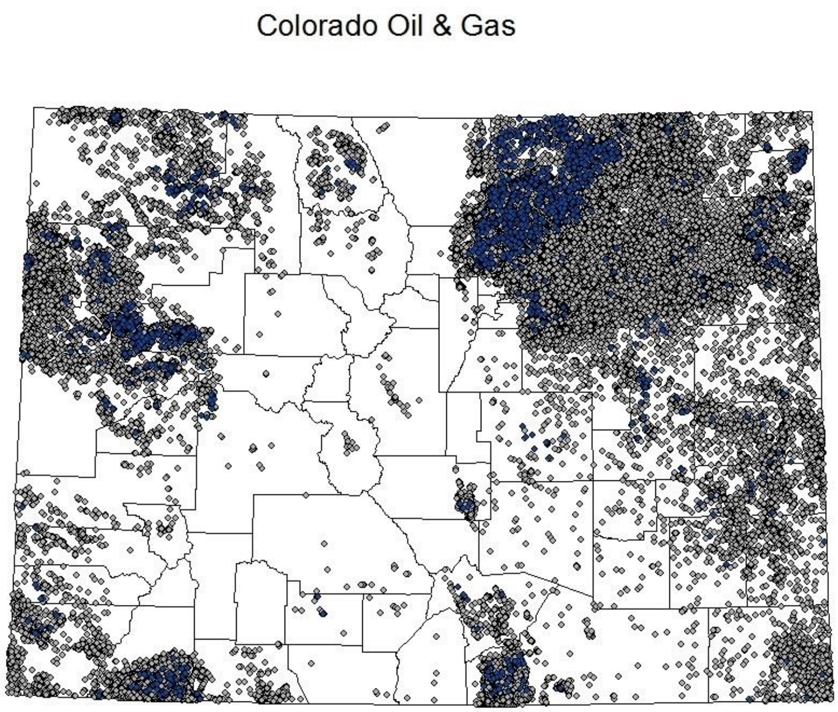

Figure 1.3 All wells in production (grey) and permitted (blue) in Colorado (COGCC)

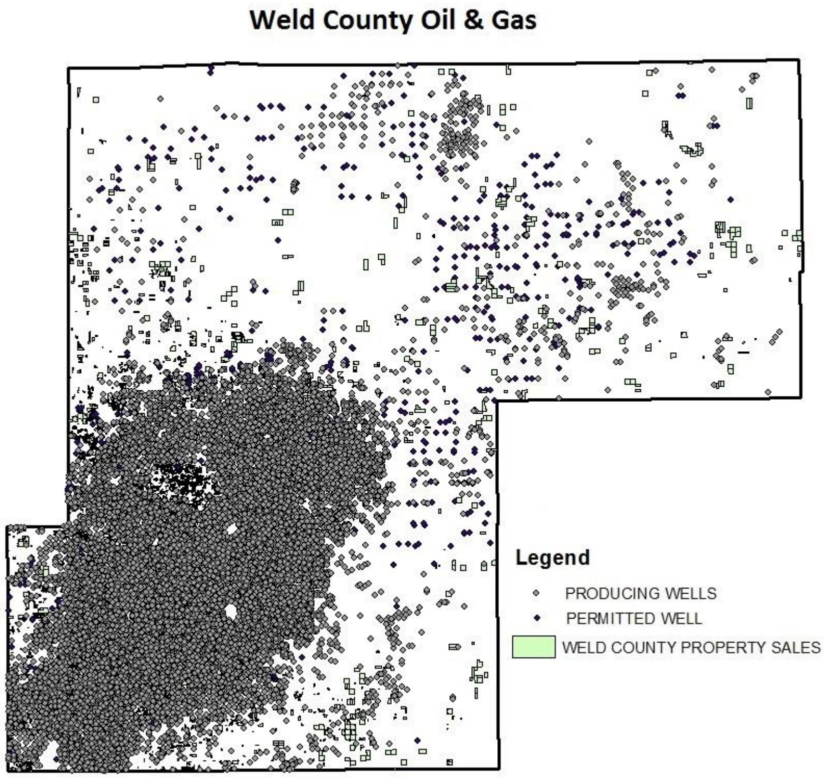

Figure 1.4 All wells in production (grey), permitted wells (blue), and property sales in Weld County (COGCC; Weld County Assessor's Office)

With the largest share of both active and permitted wells in the state (see Figures 1.1 and 1.2 for a detailed breakdown of drilling throughout the state), and given its mix between urban and rural areas and relatively high number of housing transactions, Weld County provides an interesting opportunity to look at the effects unconventional gas and shale exploration are having on those residing near the drilling sites. In July 2011, the Denver Post ran an article that provided details on some of the side effects of drilling that a Weld County couple living near a drill site had to endure while the fractured well is being drilled. The couple was surprised "…when Encana returned to drill and hydrofracture six wells, and the couple was plunged into six months of round-the-clock noise, lights, truck traffic and odors." Encana, a leading energy producer in North America, stated that each well took seven to ten days to drill, and that the drill rig ran 24 hours per day during that period. The City of Arlington Texas states that the fracking process, where trucks are visiting the site and lights are on 24-hours per day, usually lasts (+/-) 30 days (City of Arlington, n.d.). Although the couple thought they were purchasing a dream home in a nice neighborhood, living next to an open space, they describe the supposed open space as now being an “industrial site” and say that the presence of the wells has decreased their property value too much for them to sell under current economic conditions (Jaffe, 2011). In the absence of empirical evidence, it is difficult to assert whether housing values have actually decreased due to proximity to hydraulically fractured wells.

Given the salience of this issue in Colorado, specifically Weld County, it is important that the effects hydraulic fracturing is having on the area be studied. A study was conducted by the Colorado School of Public Health looking at the risks posed to human health by drilling for unconventional gas (McKenzie et al., 2012). The National Science Foundation has awarded a $12 million grant to a team led by Professor Joseph Ryan of CU-Boulder’s civil, environmental and architectural engineering department to study the effects of natural gas development. Colorado State University civil engineering professor Ken Carlson is currently working with Noble Energy to help the energy industry design water treatment plants to recycle waste water, or “flowback,” from oil and gas well drilling and fracking (Magill, 2012). While there are on-going studies about the potential risks associated with hydraulic fracturing in Colorado, there has not yet been a study conducted on its effects on residential property values.

1.1. Goals and Scope of the Study

The goal of this study is to expand the literature on the effects hydraulic fracturing is having on property values in a hedonic analysis, as suggested by Lipscomb, Wang, and Kilpatrick (2012), specifically focusing on its effect on housing values in the leading Colorado county for shale exploration. The hedonic model lends itself well to answering this type of question by applying econometric regression analysis techniques to housing transactions to determine if fracking affects housing prices in a given area. An analysis comparing the influx of drill sites to sale values of homes in the surrounding area is something that has the potential to be useful to various target groups including but not limited to policy makers, local government officials, and the general public. The sample for this study is all single-family residential homes sold between 2009 and 2012 in Weld County, Colorado. This study seeks to determine whether risk perceptions associated with hydraulic fracturing are capitalized into housing prices, and if there exists empirical evidence to support claims that it is negatively impacting neighboring communities.

1.2. Organization of the Study

This paper is comprised of six chapters, each with one or more sections. Chapter 2 contains a review of literature related either to hydraulic fracturing or hedonic analysis, and literature related to both. Discussion of the methodology and empirical specification of the model used in this study will be presented in Chapter 3. An overview of the data and the sources from which the data were obtained will be described in Chapter 4. The results of the analysis under different specifications will be provided in Chapter 5. Finally, Chapter 6 will summarize the study and its limitations, provide policy implications of the results, and give suggestions for extensions and further research on the topic.

CHAPTER 2: LITERATURE REVIEW

This section of the study will focus on reviewing literature relevant to the analysis. Topics that will be reviewed include: background information on fracking and its regulations and policies with a focus on those affecting Colorado, the underlying utility theory to the hedonic method, and the hedonic method as it has been applied to similar studies in existing literature.

2.1. Background information on Fracking and Colorado regulations

The prevalent use of horizontal drilling and hydraulic fracturing is the major difference between modern shale gas development and conventional natural gas drilling. Figure 2.1 (below) shows the sharp increase in shale gas as a share of total production as well as the projected perpetuation of shale gas as a leading type of natural gas production.

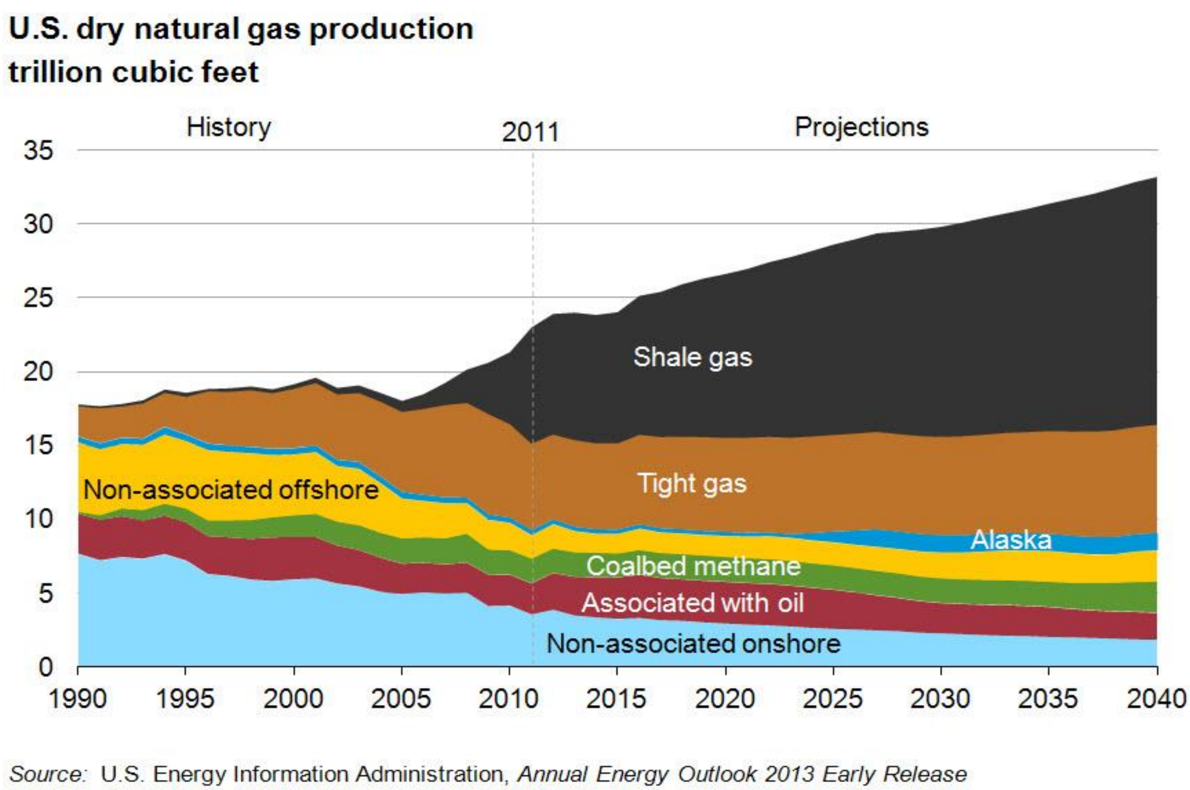

Figure 2.1 History of US natural gas production by type of gas (US EIA, 2012)

Much of the early growth in shale gas was fueled by the use of fracking and horizontal drilling of the Barnett shale in Texas. Through 2008 production of conventional gas was decreasing while prices were rising, which paved the way for an increase in fracking and horizontal drilling in 2009 (Rogers, 2011). Fracking is a drilling technique where fluids (water, chemicals, and sand) are pumped into the well, making it possible to unlock hydrocarbons from the shale and reach natural gas reserves between 6,000 and 10,000 feet below the ground. Horizontal drilling allows the bottom hole to run up to 1000 feet horizontally deep below the ground, which enhances the profitability of hydraulic as a method of natural gas extraction (FracFocus). The main difference between horizontal and vertical drilling is that horizontal drilling allows the number of surface disturbances to be kept comparatively low. In the process of fracturing the shale, miniature earthquakes are created to cause the shale to fracture and release the oil and gas inside. The following diagram provides an in depth look at the typical process of a hydraulic fracturing and horizontal drilling operation.

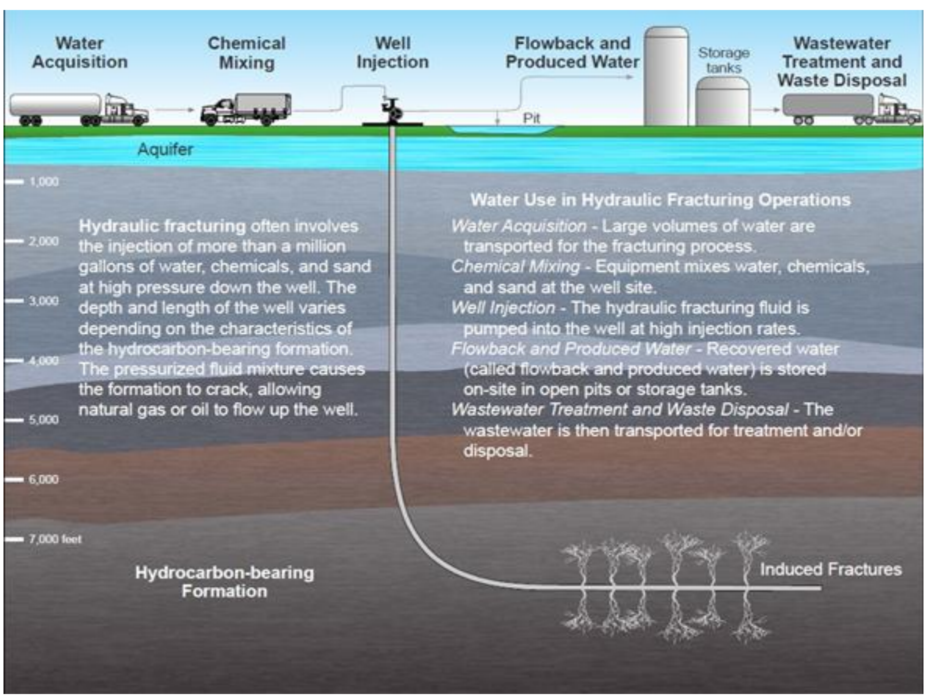

Figure 2.2 Illustration of a horizontal well showing the water lifecycle in hydraulic fracturing (US EPA, 2011)

Weld County encompasses the Niobrara shale, which covers parts of northeastern Colorado, northwestern Kansas, southwestern Nebraska, and southeastern Wyoming. Despite the fact that the Niobrara had already produced an estimated 700 billion cubic feet of natural gas in spring 2011, the attention on the formation has shifted to oil production due to the believed large amount of it. There is still some uncertainty about the amount of oil in the Niobrara formation, but even some of the first wells are producing more than the average oil producing well in the US[2] (Matthews, 2011). In an email conversation with Thom Kerr of the COGCC on February 20th, 2013, he stated, "There are currently no oil and gas wells that are being drilled in Weld County that will produce without the aid of hydraulic fracturing (Kerr, 2013). There are some wells that have been drilled in the past that were capable producers with stimulation, but that is not in recent history." This indicates that it is fair to assume that any new well permitted or drilled is done with the use of hydraulic fracturing.

Potential fracking-related environmental impacts of major concern are the noise and air quality, land use, and potential risks to water sources - many of these are addressed in current policy. The Pacific Institute interviewed 16 representatives of state and federal agencies, academia, industry, environmental groups, and community organizations to find the biggest concerns regarding the use of hydraulic fracturing. They found spills/leaks to be the biggest concern, followed by wastewater, water withdrawals, air emissions, lack of information, and the definition of "fracking" (Cooley, H. & Donnelly, K., 2012). Figure 2.3 shows the full results of Cooley and Donnelly's survey; it depicts graphically how many respondents answered "yes" on whether each relative category was of major concern to them related to the use of hydraulic fracturing.

[2] The first Niobrara horizontal well is still producing 2,500 barrels/month, much higher than the US average of 300 barrels/month.

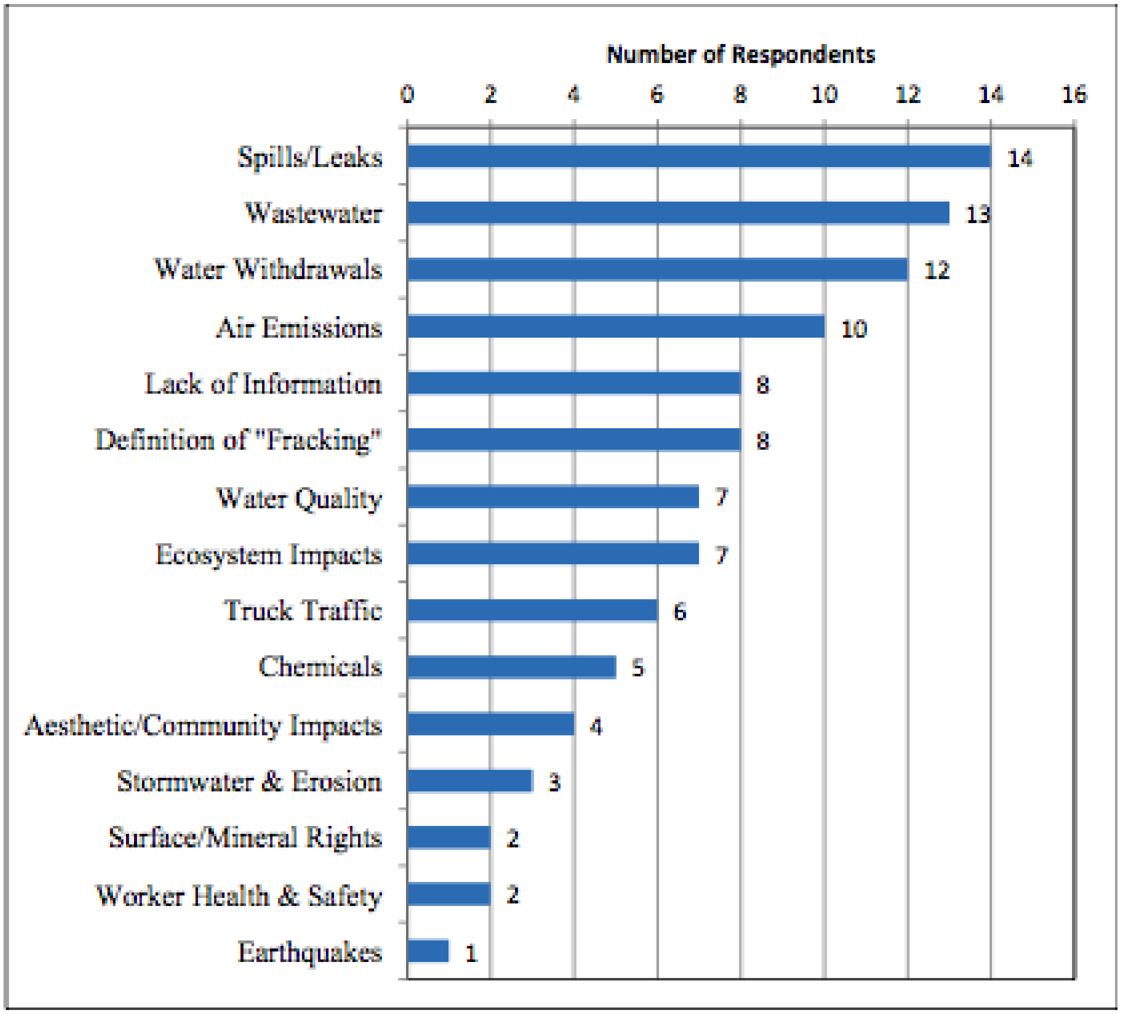

Figure 2.3 Results of Cooley & Donnelly's survey on concerns related to the use of hydraulic fracturing (Cooley and Donnelly, 2012)

McKenzie et al. (2012) studied health effects resulting from air emissions generated in the process of unconventional natural gas development; they found that those living within a half-mile or less of unconventional natural gas development are at greater risk for negative health effects than are those living farther than a half-mile from it.

Rules and regulations for hydraulic fracturing vary by state. Resources for the Future released "A review of shale gas regulations by state" in October 2012, which details each state's regulations and restrictions concerning shale exploration. As this research will focus on Colorado's Weld County, it is important to understand some of the key regulations in Colorado. The main regulating body in Colorado is the Colorado Oil and Gas Conservation Commission.

Colorado requires a buffer of 350-foot buffer to occupied urban buildings and a 150-foot buffer to occupied rural buildings. A permit is needed for all well-related water withdrawals, but pre-drill well water testing is only required in the Wattenberg Field, a large oil and gas field located primarily in Weld County. Wastewater is to be disposed of in a pit with a liner no less than 24 mils thick, and flowback/wastewater disposal tracking recordkeeping is also required. A well can only be temporarily abandoned or left idle for up to 6 months, relatively short in comparison to states like Texas that allow idle time of up to 12 months and unlimited abandonment. Although Colorado ranks on the more conservative side in most regulation categories, it allows up to 24 hours for accident reporting and has one of the higher number of wells per inspector. Local bans do exist in some places in Colorado (e.g. Fort Collins, Longmont), but no statewide bans exist right now (Krupnick et al., n.d.). While the oil and gas industry is politically and increasingly economically important in Colorado, the state is not categorized as "energy dominant" like Texas, and there tends to be a more diverse opinion on the subject. On the whole, Colorado's fracking regulations provide more environmental protection than many states, especially Texas (Davis, 2012).

2.2. Hedonic Property Method

Rosen (1974) first developed a theoretical framework for using non-market valuation to analyze connections between consumer preferences for characteristics of differentiated goods and the hedonic equilibrium price. Taylor (2003) describes the hedonic method as an indirect valuation method in which implicit prices are inferred from market observations in the absence of a direct value for a certain characteristic. Implicit prices computed in these studies may also be referred to as "hedonic" prices. The hedonic method is most often applied to housing – as it is in this study – where the price of a house is determined by its characteristics, neighborhood attributes, and any environmental amenities or disamenities. Full hedonic analyses consist of both first-stage analysis, where the hedonic price function is estimated by regressing characteristics of a good on its price, generating information about implicit and marginal prices, and second-stage analysis in which the implicit prices found in the first-stage are used to estimate the demand functions for the commodity's characteristics. This study will apply the first-stage analysis as most research in this area only does the first-stage analysis (Taylor, 2003).

A basic overview of the utility theory underlying the hedonic model discussed by Rosen (1974) will be described in this paragraph. Let Z represent the differentiated good with the a bundle of attributes, Z = z_1, z_2, z_3, … z_n. Hedonic models start with the assumption that a consumer j, with demographic characteristics α^j, derives utility (U) from some combination of the differentiated good (Z) and the composite commodity (x) that symbolizes all other goods.

$$U_j(x, z_1, z_2, \dots, z_n;\ \alpha^j) \tag{2.1}$$

Assuming that a consumer purchases only one of the differentiated good, the corresponding budget is constraint is: $y^j = x + P(Z)$, where P(Z) represents the hedonic price function, which relates changes in quantities of various attributes (z_i) to changes in the price of the differentiated good (Z) sold in a perfectly competitive market where many buyers and sellers determine an equilibrium price schedule. X is a composite commodity, with price set equal to one. The consumer maximizes utility by choosing the amounts of x and z_i subject to his or her budget constraint -- the consumer is a price taker since market prices cannot be determined individually. Amounts of x and z_i are chosen such that the marginal rate of substitution between any attribute, z_i, and x is equal to the implicit price ratio for z_i.

$$(2.2) \qquad \frac{\partial U / \partial z_i}{\partial U / \partial x} = \frac{\partial P(Z)}{\partial z_i} = P(z_i)$$

The equilibrium marginal implicit price of any housing attribute, z_i, is given by taking the first-order condition (i.e. first partial derivative of the hedonic price function with respect to attribute, z_i) of the utility function for that attribute as displayed in equation (2.2). The bid function, $\theta(z_1, \dots z_n; u, y)$, indicates the way in which a consumer's optimal bid varies in response to changes in z_i if utility (u) and income (y) are held constant. This is represented in the relationship:

$$(2.3) \qquad U^j(y_0 - \theta, z_1, \dots z_n) = U_0,$$

where U_0 represents some initial level of utility. The marginal bid a consumer is willing to make for attribute z_i $(\partial\theta/\partial z_i)$ will equal the marginal rate of substitution between z_i and x. The consumer's optimal choice of z_i in this utility maximization problem occurs when the marginal bid equals the marginal price for z_i. This relationship is presented graphically in Figure 8.

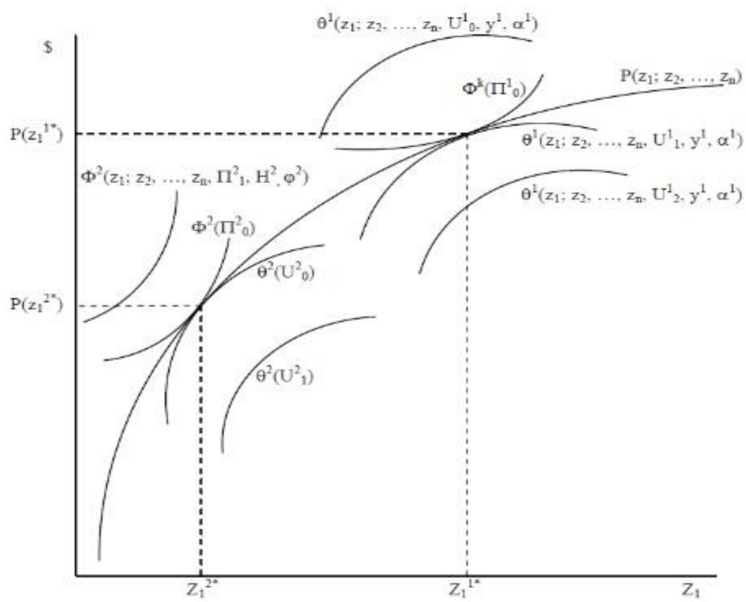

Figure 2.4 The Hedonic Price Function (Taylor, 2003)

As the hedonic method is most often applied to housing, it is important to understand the housing-specific version of the utility maximization problem laid out above. The assumption is made that a person makes a decision to buy a house based on the bundle attributes of the house. Utility derived from housing is a function of the house's structural and property characteristics $(H_1 \ldots H_h)$, demographics and neighborhood characteristics$(N_1 \ldots N_n)$, and location-based characteristics such as proximity to certain amenities or disamenities$(L_1 \ldots L_l)$. House prices are an increasing function of structural and property characteristics (i.e. $U_H > 0$), an increasing function of proximity to amenities $U_L > 0$, and a decreasing function of proximity to disamenities $U_L < 0$ (Loomis, 2004; Taylor, 2003). The hedonic price function in its general form:

(2.4) $$P_{House} = f(H_1 \ldots H_h;\ N_1 \ldots N_n;\ L_1 \ldots L_l)$$

In a hedonic price regression analysis, the appropriate dependent variable is housing sales price as shown in equation (2.4), which represents the present value of all future rents. The simplified linear function functional form is:

(2.5) $$P = \alpha_0 + \sum_{i=1}^{h} \beta_i H_i + \sum_{j=1}^{n} \beta_j N_j + \sum_{k=1}^{l} \beta_k L_k + \varepsilon$$

By regressing the attributes from the right-hand side of equation (2.5) on the dependent variable, sales price, the implicit prices for each attribute are obtained. The regression coefficients, β_i, yielded by running the regression with a linear functional form, measure the incremental change in housing sales price due to a change in one characteristic while holding all others constant. The first-order conditions described in the discussion of the underlying utility function for housing sales help guide the expected signs of the estimated regression coefficients. Motivation for the functional form chosen in this study and the empirical specification will be further discussed in Chapter 3.

2.3. Applications of HPMs to similar topics

Numerous previous studies exist that investigate the impact that local disamenities have on housing prices, such as forest fires (Loomis, 2004), hazardous waste sites (Michaels and Smith, 1990), hog operations (Palmquist et al. 1997), landfills (Hite et al. 2001), and nuclear power plants (Davis, 2011). While there is a vast body of literature on the effects of disamenities on housing prices, very few studies on the impact of proximity to oil and gas wells have been done (Boxall et al., 2005; Klaiber and Gopalkrishnan, 2012; Muehlenbachs et al., 2012). These three studies all find negative impacts on housing values across the different types of model specification and estimation techniques – these differing techniques will be discussed more in depth in this section.

Boxall et al. (2005) examined the impact of small and medium-sized oil and gas facilities on residential property values in rural areas of Alberta, Canada, employing spatial econometric techniques. They used relatively small sample of 532 residential property sales between January 1994 and March 2001 in a hedonic analysis using the prices and characteristics of the properties in the sample, including the property's proximity to different oil and gas facilities. Count variables for the number of wells (flaring oil batteries, sour, sweet, well pipelines) within 4 kilometers[3] of a property were used in addition to a continuous distance variable for the nearest sour gas plant to the house to explore whether proximity to oil and gas facilities affects housing values. They found the double log specification of their model had the best fit in part because it generated price elasticities helpful in interpreting implicit prices. Lagrange multiplier (LM) and robust LM tests were used to identify the existence spatial dependencies in the data. In order to explore the spatial nature of the data, a spatial weights matrix of inverse distances was generated.

[3] Boxall et al. used a range of 4 km as it "was predetermined by energy experts based on evidence regarding the probable maximum range for impacts that extend from typical facilities such as wells…"

Distance specifications (distance bands of a set number of kilometers) were preferred to lattice specifications (number of closest neighbors), and they ultimately chose a distance band of 4 kilometers. Results of the study suggest that the existence of an oil and gas facility near a home can have a negative impact on its value. Measures of hazard and disamenity were found to have statistically significant, negative effects on housing values, reducing the value by between 4 and 8 percent at the mean level of facilities within 4 kilometers. Boxall et al. (2005) provided information that proximity to traditional oil and gas wells has a negative impact on residential property values.

Klaiber and Gopalkrishnan (2012) utilized a hedonic model to assess whether any potential negative externalities associated with shale exploration are capitalized into the values of surrounding residential properties. Their sample includes 3,464 housing sales between 2008 and 2010 in an area south of Pittsburgh, PA that experienced a large influx of horizontal wells in the Marcellus shale play in 2008. This study focused on specifically the permitting and drilling of horizontal wells, exploring potential negative effects of the visibility that occurs between the permit approval and actual drilling of the well. To analyze their data, Klaiber and Gopalkrishnan used a hedonic model with a semi-log specification (logged sales price as the dependent variable). Their independent variable of interest was the number of shale wells within one mile of a property, permitted and drilled up to six months prior to the sale of the house. Effects of horizontal drilling on housing values in rural/suburban areas were found to be important based on different land use categories specified by the researchers (residential, agricultural, forest, or industrial). Although the relationship between sale price and the number of wells near a house reliant on well water was explored, the number of shale wells became statistically insignificant under this model specification. A sensitivity analysis on the time frame and the spatial buffers was performed by testing time windows of 3 and 12 months in addition to 6 months and spatial buffers of 0.75 miles, 1mile, and 2 miles. In this, they found that using a spatial buffer of 0.75 miles and a time window 3 months prior to sale had a very large and statistically significant effect on properties serviced by well water that persists on into the 6 month time window. As they increased the time window prior to sale and /or the spatial buffer, their results became progressively less significant. They indicated that the effect of drilling seemed to disappear if the drill site was 2 miles or farther from the house. Overall, their results show that housing values are negatively impacted in the short term, and that those dependent upon well water and surrounded by agricultural land are disproportionally negatively impacted by drilling.

Muehlenbachs et al. (2012) examined the negative externalities associated with shale gas development across different drinking water sources by using a triple-difference (DDD) estimator in a hedonic analysis. It focused on the interaction between groundwater and hydraulic fracturing, and the potential risks associated with this interaction. The sample included 19,055 housing transactions between 2005 and 2009. Three different specifications were estimated in this study: cross-sectional OLS, property fixed effects, and a triple-difference estimator that uses detailed geographical information. The two explanatory variables of interest in this study are the distance to the nearest well at the time of the sale, a continuous distance variable, and whether the house is a part of the Public Water Service Area (PWSA), a dummy variable. The DDD estimator uses two different treatment and control groups: one based on whether the house is located within 2000 meters of a shale gas well (treatment group) or outside 2000 meters (control group), and one based on whether a house is dependent on groundwater (treatment group) or not (control group) and is within 2 km of a shale gas well. This study found that the risk of groundwater contamination leads to a statistically significant and large reduction of 26.3 percent in the price of a house, depending on what type of water source the house uses. The authors asserted that, based on their results, there may be an increase in the likelihood of foreclosures in areas that are experiencing rapid growth in hydraulic fracturing and shale exploration.

2.4 Summary

These three studies about the effects of oil and gas drilling on housing values found a negative effect of drilling on housing values. However, these negative values reflected a particular range of distance (e.g. 2 kilometers used by Muehlenbachs et al. 2012 or 4 kilometers used by Boxall et al. 2005), or a home that received its water from a domestic water well (Muehlenbachs et al. 2012). One of these studies looked at the effects of sour gas drilling on housing values in rural Alberta, Canada (Boxall et al., 2005) and two looked at Washington County, Pennsylvania (Klaiber and Gopalkrishnan, 2012; Muehlenbachs et al., 2012); none of these studies have involved Colorado or the relatively flat topography of eastern Colorado. Although Alberta has a fairly flat topography, it lacks the high population density of Weld County, Colorado, making a direct comparison difficult. This study will add to the results found by these three related studies by applying a similar type of model to a new study area, Weld County.

CHAPTER 3: EMPIRICAL MODEL

3.1. Variables of Interest

Variables of interest in this study were determined based on hedonic method literature and the availability of the data. More information about the collection of and descriptive statistics for these data will be presented in Chapter 4. The appropriate dependent variable in a hedonic study is the sales price of the house, as suggested by Taylor (2003). There are many independent variables that can be used in this type of study due to the nature of the hedonic price function, an envelope function that connects the sales price of a house to its characteristics. Characteristics that are not thought to have an effect on price, even if they vary across by product type, are not included as regressors in the hedonic price regression. Similarly, characteristics of buyers and sellers of houses are excluded from the regression analysis (Taylor, 2003). There are three over-arching types of independent variables included in the hedonic price regression that align with those specified in the hedonic price function, derived through utility theory: structural and property characteristics variables, neighborhood and demographic variables, and location-based variables measuring proximity to (dis)amenities.

Table 3.1 provides a list of all potential variables of interest with the relative descriptions, expected relationship to the dependent variable (ln_sales or real_salep), the unit of measurement for the variable, and the source from which the data were obtained. The table is broken up into two categories: dependent variable and independent variables. Given the number of potential independent variables that are included in Table 3.1, these variables are further split into categories that better describe them.

Table 3.1 List of variables with descriptions and expected relationship with the dependent variable

Variable Name	Description	Units	Expected Sign	Data Source
Dependent Variable				
ln_salep	Logged real sale price of property (base = 2009)	USD $		Assessor's office[a]
real_salep	Real sale price of property (base = 2009)	USD $		Assessor's office
Independent Variables				
Structural & Property Characteristics				
lotsize	Size of the land associated with the residential structure	Acres	+	Assessor's office
bedrooms	Number of bedrooms	Count	+	Assessor's office
baths	Number of bathrooms	Count	+	Assessor's office
bldgs	Number of buildings on property	Count	+	Assessor's office
ressf	Area of residential structure	ft^2	+	Assessor's office
age	Age of the residential structure at time of sale	Years	-	Assessor's office
outbuildingsf	Area of any outbuildings on the property	ft^2	+	Assessor's office
porchsf	Area of porch	ft^2	+	Assessor's office
fin_bsmnt	DV; =1 if house has a finished basement	0/1	+	Assessor's office
garage	DV; =1 if house has a garage	0/1	+	Assessor's office
remodel	DV; =1 if a house was remodeled	0/1	+	Assessor's office
waterwell	DV: =1 if water well on property	0/1	+	COGCC[b]
horzwell	DV: =1 if horizontal wellbore runs under property	0/1	-	COGCC
Market timing				
y2010	DV; =1 if house sold in 2010	0/1		Assessor's office
y2011	DV; =1 if house sold in 2011	0/1		Assessor's office
y2012	DV; =1 if house sold in 2012	0/1		Assessor's office
allemp	Total number of hours worked by employees in the oil and gas sector in Weld County	Thousands	+	BLS[c]
Census tract demographics				
hh_inc	Mean income of census tract	USD $	+	ACS[d]
pct_white	Percentage Caucasian in census tract	%	+	DOLA[e]
pct_hisp	Percentage Hispanic in census tract	%	-	DOLA
med_age	Median age	Years		DOLA
pct_65plus	Percentage of population over 65 years old	%	-	DOLA
pct_bachlr	Percentage of 25+ population with college degree	%	+	DOLA
pct_HSgrad	Percentage of 25+ population with at least a high school education	%	+	DOLA
pct_own	Percentage of houses owned in census tract	%	+	DOLA

pct_vac	Percentage of houses vacant in census tract	%	-	DOLA
Location characteristics				
RURAL	DV; =1 if house is located in no city or town	0/1	+	Assessor's office
GREELEY	DV; =1 if house is located in Greeley	0/1	-	Assessor's office
hwy_100yd	DV; =1 if nearest interstate is within 100 yards	0/1	-	Assessor's office
hwy_1mile	DV; =1 if nearest interstate is farther than 100 yards and less than 1 mile	0/1	+	Assessor's office
distSPUD	Distance to nearest well drilled within 2 miles and up to 60 days prior to the sale	Meters	+	COGCC
ln_spud	Natural log of distance to nearest spud (ln(distSPUD))	ln(meters)	+	COGCC
spudcount	Number of wells being drilled within a half mile of a house within 60 days of sale	Count	-	COGCC
distPROD	Distance to closest producing well within a half mile at time of sale	Meters	+	COGCC
ln_prod	Natural log of distance to nearest spud (ln(distSPUD))	ln(meters)	+	COGCC
num_producing	Number of wells in production within a half mile of a house at the time of sale	Count	-	COGCC
distPERM	Distance to closest permitted well within 2 miles and 60 days of sale	Meters	-	COGCC
permitcount	Number of permitted wells within a half mile of a house within 60 days of sale	Count	-	COGCC

a = Weld County Office of the Assessor; b = Colorado Oil and Gas Conservation Commission; c = Bureau of Labor Statistics d = American Community Survey (5-year estimates); e = Colorado Department of Local Affairs

3.2. Empirical Specification

Hedonic price functions can take on a number of different functional forms such as linear, semi-log, double-log, or linear or quadratic Box-Cox. Cropper et al. (1988) begin their article, which uses simulation techniques to determine the accuracy of different functional form choices in a hedonic price function, with the following quote that captures the necessity of researcher judgment in specifying functional form in hedonic analyses.

> *"The fact that economics theory places few restrictions on the form of the hedonic price function has led most researchers to use a goodness-of-fit criterion in choosing an appropriate form for the hedonic function. If, however, one's goal is to value product attributes, the form that should be used is the one that most accurately estimates marginal attribute prices." (Cropper et al., 1988, p.668)*

Cropper et al. (1988) found an array of significant empirical results in their study. Multi-collinearity issues were most present in the linear, semi-log, and double-log specifications. They found the linear and quadratic Box-Cox functions consistently performed better than the other specifications in the absence of omitted variable bias. However, when certain variables are omitted from the regression (i.e. it is misspecified), the study found that the linear, semi-log, and double log produced smaller mean errors and less biased results (Cropper et al., 1988). Lansford and Jones (1995) found that marginal values become more difficult to interpret and the calculation of them is challenging when a Box-Cox transformation is used (as cited in Loomis, 2004).

Based on the suggestions of functional form from Cropper et al., two functional forms will be explored in this study, a linear specification and a log-linear specification for the non-linear functional form. Estimating both a linear and non-linear functional form allows for post-estimation testing of the sensitivity of coefficient estimates to different types of functional forms. Linear hedonic price functions, for example, have the advantage that the coefficients yielded by running regression analyses provide the marginal willingness to pay for an incremental increase of one unit in that specific attribute. This assumes that the marginal value of an additional unit of characteristic z_i is constant across all values ofz_i, which is likely not true for most characteristics. As the marginal prices for most housing characteristics are likely non-constant, transformations of the dependent and/or independent variables may be necessary to capture non-linear relationships (Taylor, 2003).

This study explores three types of functional form as suggested by Taylor (2003) and Cropper et al. (1988): linear, log-linear, and double-log. Box-Cox specifications are not explored based on the nature of secondary data and the potential for measurement error in variables and/or unobservable or missing data resulting in omitted variable bias. Each of these specifications is set up as a base model, in which none of the drill-related variables are included in the model, and as a treatment model that includes the drill-related variables[4]. Estimation of both a base and treatment model is carried out in order to capture changes in the coefficients from the base model when fracking variables enter the model. This is necessary as the regressions run that include any of the "spud" variables are conditional on the presence of a well being drilled within 2 miles, dating as far back as 60 days prior to the sale of the house. The method for calculating the implicit prices for each of these functional forms as well as their advantages and disadvantages are described in the following sub-sections 3.2.1, 3.2.2, and 3.2.3.

3.2.1. Linear specification

The linear specification of the hedonic price function is the simplest of the possible functional forms used to estimate it. The linear functional form has the simplest interpretation of marginal implicit prices ($\partial P / \partial z_i = \beta_i$); the regression coefficients are interpreted as the implicit housing price change of a one-unit increase in the given attribute (e.g. a one unit increase in the number of bathrooms would result in β_3 change in the sale price of house i). Since that marginal price is probably not constant across all values of certain attributes, the linear specification of the hedonic price function has the disadvantage of misrepresenting the marginal effects of certain characteristics. Equation 3.1 shows the linear specification that excludes well-activity variables and was estimated in STATA using ordinary least squares (OLS), and equation 3.2. displays the linear specification that included variables about wells in the process of being drilled.

[4] It is important to note that the degrees of freedom in the treatment model are almost 10,000 fewer.

(3.1) *Linear base specification*

$$
\begin{aligned}
salep\_real_i = {} & \alpha_i + \beta_1 lotsize_i + \beta_2 lotsize2_i + \beta_3 baths_i + \beta_4 age_i + \beta_5 ressf_i + \beta_6 ressf2_i \\
& + \beta_7 outbuildingsf_i + \beta_8 porchsf_i + \beta_9 remodel_i + \beta_{10} garage_i \\
& + \beta_{11} finish\_bsmnt_i + \beta_{12} GREELEY_i + \beta_{13} RURAL_i + \beta_{14} y2010_i \\
& + \beta_{15} y2011_i + \beta_{16} y2012_i + \beta_{17} hwy\_mile_i + \beta_{18} pct\_hisp_i + \beta_{19} pct\_own_i \\
& + \beta_{20} hh\_inc_i + \beta_{21} pct\_bachlr_i + \varepsilon_i
\end{aligned}
$$

(3.2) *Linear treatment specification*

$$
\begin{aligned}
salep\_real_i = {} & \alpha_i + \beta_1 lotsize_i + \beta_2 lotsize2_i + \beta_3 baths_i + \beta_4 age_i + \beta_5 ressf_i + \beta_6 ressf2_i \\
& + \beta_7 outbuildingsf_i + \beta_8 porchsf_i + \beta_9 remodel_i + \beta_{10} garage_i \\
& + \beta_{11} finish\_bsmnt_i + \beta_{12} GREELEY_i + \beta_{13} RURAL_i + \beta_{14} y2010_i \\
& + \beta_{15} y2011_i + \beta_{16} y2012_i + \beta_{17} hwy\_mile_i + \beta_{18} pct\_hisp_i + \beta_{19} pct\_own_i \\
& + \beta_{20} hh\_inc_i + \beta_{21} pct\_bachlr_i + \beta_{22} spudcount_i + \beta_{23} num\_producing_i \\
& + \beta_{24} distSPUD_i + \varepsilon_i
\end{aligned}
$$

3.2.2. Semi-log specification

Existing literature on hedonic price functions and regression analyses tend to favor a semi-log (i.e. a log-linear) functional form for estimating the price function (Muehlenbachs et al., 2012; Loomis, 2004; Lewis & Acharya, 2006). The semi-log model involves the transformation of the dependent variable into logs, an act that scales the sales price of the house. This has the advantage of potentially capturing some of the non-linearities in the data and marginal prices for characteristics. Given that marginal prices for semi-log hedonic price functions are variable and depend on the price level, these implicit prices must be estimated at some price level of the housing sales price. For a semi-log specification, Taylor (2003) provides the calculation and intuition of the marginal implicit price, ($^{\partial P}/_{\partial z_i} = \beta_i \cdot P$), where *P* represents a housing sales price level. Interpretation of dummy variable coefficients must be done with care in a semi-log model as it represents an approximate percentage change in the sale price of a house if the characteristic is present (i.e. if that variable is set equal to one)[5]. Equation 3.3 displays the semi-log specification of the model that excludes well-activity variables, and equation 3.4 shows the linear specification that included variables about wells in the process of being drilled.

(3.3) *Semi-log base specification*

$$\begin{aligned} ln\_sales_i = {} & \alpha_i + \beta_1 lotsize_i + \beta_2 lotsize2_i + \beta_3 baths_i + \beta_4 age_i + \beta_5 ressf_i + \beta_6 ressf2_i \\ & + \beta_7 outbuildingsf_i + \beta_8 porchsf_i + \beta_9 remodel_i + \beta_{10} garage_i \\ & + \beta_{11} finish\_bsmnt_i + \beta_{12} GREELEY_i + \beta_{13} RURAL_i + \beta_{14} y2010_i \\ & + \beta_{15} y2011_i + \beta_{16} y2012_i + \beta_{17} hwy\_mile_i + \beta_{18} pct\_hisp_i + \beta_{19} pct\_own_i \\ & + \beta_{20} hh\_inc_i + \beta_{21} pct\_bachlr_i + \varepsilon_i \end{aligned}$$

(3.4) *Semi-log treatment specification*

$$\begin{aligned} ln\_sales_i = {} & \alpha_i + \beta_1 lotsize_i + \beta_2 lotsize2_i + \beta_3 baths_i + \beta_4 age_i + \beta_5 ressf_i + \beta_6 ressf2_i \\ & + \beta_7 outbuildingsf_i + \beta_8 porchsf_i + \beta_9 remodel_i + \beta_{10} garage_i \\ & + \beta_{11} finish\_bsmnt_i + \beta_{12} GREELEY_i + \beta_{13} RURAL_i + \beta_{14} y2010_i \\ & + \beta_{15} y2011_i + \beta_{16} y2012_i + \beta_{17} hwy\_mile_i + \beta_{18} pct\_hisp_i + \beta_{19} pct\_own_i \\ & + \beta_{20} hh\_inc_i + \beta_{21} pct\_bachlr_i + \beta_{22} spudcount_i + \beta_{23} num\_producing_i \\ & + \beta_{24} distSPUD_i + \varepsilon_i \end{aligned}$$

3.2.3. Log-log specification

The log-log specification is similar to the semi-log form in that it includes a log transformation of the dependent variable; however it differs from the semi-log by also transforming continuous right-hand side variables into logs[6]. Marginal implicit prices yielded by running double-log regressions are simple to interpret. These marginal prices are also measures of price elasticity to (i.e. for a one percentage change in attribute z_i, there is a resulting percentage change in the dependent variable by a magnitude of β_i). Due to the lack of houses sold that had an outbuilding on the property (i.e. outbuildingsf = 0), outbuildingsf was not logged because too many missing values were generated in the process. Equation 3.5 shows the

[5] Taylor (2003) asserts that as long as the coefficient estimate is relatively small, the error in interpretation is small.

[6] Discrete variables are not transformed into logs because the log of zero cannot be calculated, thus logging a discrete variable will result in an undefined value – this becomes a missing value in STATA and is then dropped from the regression analysis.

equation estimated under the double-log specification for the entire sample, and equation 3.6 shows the model estimated that contained the sample of houses with well-activity within 2 miles of the house, up to 60 days prior to the sale.

(3.5) *Double-log base specification*

$$\begin{aligned} ln\_sales_i = {} & \alpha_i + \beta_1 \ln\_lotsize_i + \beta_2 \ln\_ressf_i + \beta_3 baths_i + \beta_4 age_i + \beta_5 outbuildingsf_i \\ & + \beta_6 \ln\_porchsf_i + \beta_9 garage_i + \beta_{10} finish\_bsmnt_i + \beta_{11} GREELEY_i \\ & + \beta_{12} RURAL_i + \beta_{13} y2010_i + \beta_{14} y2011_i + \beta_{15} y2012_i + \beta_{16} hwy\_mile_i \\ & + \beta_{17} pct\_hisp_i + \beta_{18} pct\_own_i + \beta_{19} hh\_inc_i + \beta_{20} pct\_bachlr_i + \varepsilon_i \end{aligned}$$

(3.6) *Double-log treatment specification*

$$\begin{aligned} ln\_sales_i = {} & \alpha_i + \beta_1 \ln\_lotsize_i + \beta_2 \ln\_ressf_i + \beta_3 baths_i + \beta_4 age_i + \beta_5 outbuildingsf_i \\ & + \beta_6 \ln\_porchsf_i + \beta_9 garage_i + \beta_{10} finish\_bsmnt_i + \beta_{11} GREELEY_i \\ & + \beta_{12} RURAL_i + \beta_{13} y2010_i + \beta_{14} y2011_i + \beta_{15} y2012_i + \beta_{16} hwy\_mile_i \\ & + \beta_{17} pct\_hisp_i + \beta_{18} pct\_own_i + \beta_{19} hh\_inc_i + \beta_{20} pct\_bachlr_i \\ & + \beta_{21} spudcount_i + \beta_{22} ln\_spud_i + \varepsilon_i \end{aligned}$$

3.3. Spatial Model

Spatial econometrics, a subfield of econometrics that is becoming more widespread, is used to incorporate the geography of the data analyzed in regression analysis. Given the spatial nature of housing transactions data, the data are tested for potential spatial dependencies, also known as spatial autocorrelation (Anselin, 1988). Spatial dependency refers to the positive or negative correlation of observations based on proximity to other observations that suggest process-generating patterns. The Moran's I test statistic is used to determine whether data are spatially autocorrelated.

There are many ways to model spatial dependencies in data; one of these ways is to use a variogram to determine the covariance and generalized least squares (GLS). Residuals are obtained from the OLS model in order to fit a variogram, which describes the degree of spatial dependence in the data. The information from the variogram is then used to compute the covariance matrix (C). By Cholesky decomposition, $C = LL^T$, from which L is taken and all of the regressors and regressands are multiplied by the inverse of L. The variable distSPUD is omitted from the spatial model because it is a continuous distance variable and causes issues. Interaction terms are created for all explanatory variables by GREELEY and by RURAL to capture those effects. Equation 3.7 lays out the general empirical specification of the spatial model.

(3.7) *Spatial specification*

$$L^{-1} \times \ln\_sales_i = \beta_i(L^{-1} \times X_i) + \beta_i(L^{-1} \times (GREELEY * X_i)) + \beta_i(L^{-1} \times (RURAL * X_i)) + (L^{-1} \times \varepsilon_i)$$

3.4. Statement of Hypotheses

Theory, the data, and the results of previous studies guide the hypotheses made about the relationships between each independent variable and the dependent variable. The expected sign of each candidate variable as it relates to the dependent variable is provided along with a description in Table 1 in section 3.1. Most structural characteristics are considered to have a positive effect on housing prices with the exclusion of age, because as the age of the house increases, sales prices are expected to decrease. Lot size and residential square footage are expected to have a non-linear relationship with price, because housing prices are thought to increase with these variables at a decreasing rate (Taylor, 2003). A house being located in Greeley, the county seat, is expected to have a negative price due to the relative poorness of the city to the rest of the county. Likewise, if a house is located in no town it takes on a value of one in the variable RURAL; it is expected to have a positive sign for many reasons including less air and noise pollution and larger lot sizes. Having a highway located within 100 yards of your home may cause some level of disutility due to traffic and negatively affect the price of your house; however it may be positive to have a highway located within a mile of your home for commuting purposes.

Discerning the expected relationship between well-activity variables and housing sales price is the focus of this study. For all of the count variables – spudcount and num_producing – the expected relationship is negative. As the density of wells or wells-to-be-drilled within a half mile of a home increases, it is expected that the house would lose value, suggesting a negative relationship. The specifications of the model that include distance variables to the nearest spud (well-to-be-drilled) within two miles and the nearest producing well within a half mile are limited to the observations that are complete and contain housing transaction data and well-activity within those specified distance bands. It is expected that having a well within that range has a negative impact on housing values, thus since Euclidean distance is used in the linear and semi-log specifications, the estimated coefficient should have a positive sign. As distance to the nearest well being drilled within two miles increases, housing prices should theoretically increase. The same logic is applied to the distance to the nearest well in production, although the scope is smaller at one half-mile. These expectations are guided primarily based on the results of Boxall et al. (2005) and Muehlenbachs et al. (2012), both of whom found negative effects of drilling on housing values within a specified distance band, and also by the concerns that have been arising in the local and (inter)national news.

The significance of individual variables in each model is evaluated by calculating t-tests on every variable, while overall model significance is evaluated using an f-test to test for statistical significance of the whole model. Since heteroskedasticity is always something to be aware of when dealing with cross-sectional data, post-estimation testing of regression residuals is performed using the Breusch-Pagan test statistic for heteroskedasticity. If the model is found to have heteroskedasticity, it will be corrected for using White's robust standard errors. Correct model specification is tested using Ramsey's RESET test for omitted variables. Additionally, consistency of coefficients estimates across the base and treatment model is analyzed to determine whether omitted biases may be present in the base specification of the model.

CHAPTER 4: DATA

Several data sets were collected from a handful of sources and joined together to form a database comprised of housing sales date and price, housing characteristics, location, census tract demographics, and proximity to fracked wells. Data on housing sales and characteristics used in this study were obtained through the Weld County Office of the Assessor. GIS data containing geographical information as well as sale date and price were provided directly by the Office of the Assessor, while data on property characteristics were downloaded from the office of the assessor's website and merged with the GIS data based on the housing account number. Data on fractured wells, horizontal wells, permitted wells, and water wells were downloaded from the website of the Colorado Oil and Gas Conservation Commission (COGCC). The data on well production, which include various dates (spud date, status of well bore, and first date of production) in the drill process, were merged with GIS data containing information on the location of the well to complete the well data. Permitted well GIS data contained information on both location and the date the permit was granted, therefore these data could be extracted from the .dbf file associated with the shape file downloaded from COGCC.

4.1 Housing Sales Data

Information about all properties sold in Weld County, Colorado from 2009 to 2012 is provided in the housing data. The original housing transaction data set contains 23,117 observations available for sampling – these data were provided directly as a GIS layer by the Office of the Assessor. Figure 4.1, created using the data provided and ArcGIS, shows the locations of these housing transactions; since these are primarily concentrated in Greeley, Figure 4.2 shows a closer up version of Greeley.

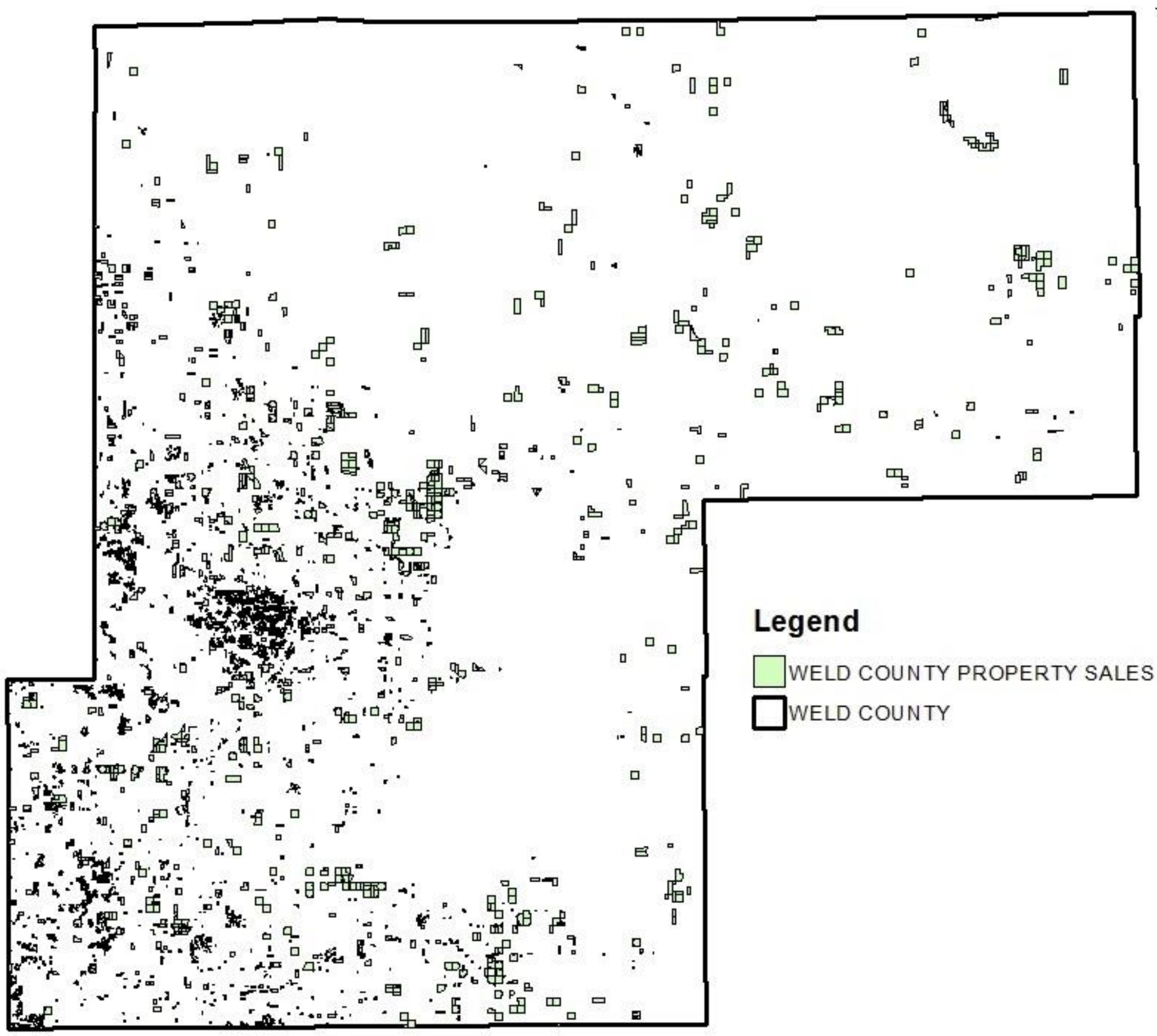

Figure 4.1 All housing transactions in Weld County, 2009-2012 (Weld County Office of the Assessor)

Figure 4.2 Greeley housing transactions in Weld County, 2009-2012 (Weld County Office of the Assessor)

Sales price data are deflated[7] (2009 = 100) using the annual Housing Price Index for the Denver-Boulder-Greeley area, which was obtained from the US Bureau of Labor Statistics using the series id CUUSA433SAH (U.S. Bureau of Labor Statistics, n.d.). Since the UTM

[7] Deflated house prices generated using the following formula: salep_real = salep*($HPI_{current}$/HPI_{base})

coordinates identify the centroid of the property, it is understood that the house may be located at a different point on the parcel and that a slight measurement error may come as a result of this.

Upon merging this data set with the data sets containing housing characteristics in STATA, incomplete observations (i.e. those missing either sales date and price or housing characteristics such as lot size, residential square feet, number of bedrooms, etc.) were dropped from the data set. Property sales other than single-family residential housing such as commercial properties and apartment buildings were also dropped from the dataset in order to capture the effects of drilling only on residential housing sales. Any property with a lot size of zero acres or over 500 acres was considered an outlier and was removed from the dataset. Houses with under 400 residential square feet were dropped because it was unclear whether these were errors in the data entry. In order to create a data set of arm's length, single-residential housing transactions, houses that sold for under $\$50/ft^2$ were removed from the data. A few housing sales present in the assessor's dataset were actually located outside of county boundaries based on GIS mapping; these were dropped when the GIS property sales centroid point data was spatially joined with the census tract shapefile downloaded from the Colorado Department of Local Affairs' (DOLA) website. Duplexes, manufactured homes, and townhouses were removed from the sample in order to capture the effects of drilling on only those with an occupancy code of "single-family residential" homes sold. The final eligible sample before removing duplexes, manufactured homes, and townhouses contained 14,222 observations, for which summary statistics are provided in Table 1A in the Appendix. Summary statistics for the final sample are provided in Table 4.1.

Table 4.1 Summary statistics for structural and property characteristics

Variable	N	Mean	Std. Dev.	Min	Max	Count
salep_real	13531	215230.40	120543.6	30174.49	2413959	
ln_sales	13531	12.158	0.491219	10.31475	14.69678	
lotsize	13531	0.46	1.43663	0	40	
baths	13531	2.62	0.924454	0	23	
age	13531	16.56	20.00692	0	147	
ressf	13531	1708.25	646.5102	520	7774	
outbuildingsf	13531	88.20	582.6612	0	22092	
porchsf	13531	264.46	256.1646	0	4824	
garage	13531	0.95	0.210543	0	1	
finish_bsmnt	13531	0.39	0.487933	0	1	
y2010		0.23	0.423376	0	1	3166
y2011		0.24	0.426944	0	1	3244
y2012		0.30	0.456568	0	1	4007
y2009 (omitted)						3114

In order to capture any effects on price associated with the year of a sale, the variables y2010, y2011, and y2012 were generated. Sales from the year 2009, the reference year, are omitted in order to avoid creating a dummy variable trap. If a house was sold in 2010, the variable y2010 is set equal to one. If a house was sold in 2011, y2011 is set equal to one. If a house was sold in 2012, y2012 is set equal to one. Remodel, garage, and finish_bsmnt were generated in STATA, and set equal to one if the house had that characteristic, zero if it did not.

4.2 Demographic & Neighborhood Data

Colorado's DOLA website has downloadable GIS shapefiles that contain data on demographics from the 2010 US Census by census block, census tract, county, place, school district, and zip code (Colorado Department of Local Affairs, n.d.). These data contain various demographic characteristics of these population groupings. As past literature has suggested (Taylor, 2003) and implemented (Muehlenbachs et al., 2012; Lewis and Acharya, 2006; Klaiber and Gopalakrishnan, 2012), census tracts were chosen as the appropriate level to be used for demographics data. When the housing transactions centroid data were spatially joined to the census tract data in ArcGIS, the demographic characteristics specific to each house were added to the housing sales attribute table. One key demographic was missing from the DOLA shapefile attributes - a measure of household income. Data on mean household income were obtained from the American Community Survey, specifically the 5-year mean household income over the past 12 months[8], and matched to each census tract by tract number. Mean income was obtained to be used in accordance with Muehlenbachs et al. (2012), which utilized mean household income for the census tract.

Dummy variables were generated to control for a few location characteristics associated with the houses sold. Using the logic that a house located within one mile of a major interstate may derive some benefit from this proximity but that being located within 100 yards of an interstate may impose a cost on those living in the home, two dummy variables were created using the buffer and intersect tools in ArcGIS to reflect these distances. HWY_100yd is equal to one when a house falls within the 100 yard buffer around a major interstate. HWY_1mile is equal to one when a house falls within the mile buffer but not within the 100 yard buffer around a major interstate; if HWY_1mile is equal to zero, the property is located farther than one mile from the nearest major highway in Weld County. A dummy variable (GREELEY) is created to indicate whether the house was within Greeley city limits. In addition, for any house located outside of all city and town limits in the county, the dummy variable (RURAL) is set equal to one. After consideration of potential effects of increased employment in the oil and gas sector on housing prices, data on total employment in the oil and gas sector were obtained from the Bureau of Labor Statistics using series id CES1021100001 (US Bureau of Labor Statistics, n.d.).

[8] 5-year estimates are used because they are recommended for this type of study under the ACS's "Guidance for Data Users" available on their site (http://www.census.gov/acs/www/guidance_for_data_users/estimates/)

Allemp is a variable collected on a monthly basis that captures the total number of employees (in thousands) currently working in the oil and gas sector. These data were matched to housing sales data based on the date the house was sold, and range from approximately 156 to 192 thousand employees in the Weld County oil and gas sector between 2009 and 2012. Table 4.2 provides the complete summary statistics for all neighborhood and demographic data.

Table 4.2 Summary statistics for neighborhood and demographic variables

Variable	N	Mean	Std. Dev.	Min	Max
GREELEY	13531	0.28	0.448044	0	1
RURAL	13531	0.09	0.283657	0	1
hwy_mile	13531	0.41	0.492723	0	1
pct_hisp	13531	0.22	0.156326	0.0658	0.8635
pct_own	13531	0.77	0.134501	0.0588	0.9563
hh_inc	13531	78940.46	21925.2	23052	157490
pct_bachlr	13531	0.29	0.122685	0.019064	0.6
allemp	13531	170.536	12.31295	156.1	191.7

4.3 Well Data

All data on hydraulically fractured wells, directional wells, permitted wells, and water wells were downloaded from the website of the Colorado Oil and Gas Conservation Commission (COGCC). In order to capture the effects different stages in the drilling and natural gas extraction might have on housing prices, it is imperative to include data on permitted wells, wells in the process of being drilled, and wells in production. GIS point data files, updated daily, were downloaded from the COGCC's maps "GIS downloads" section of the website. Well completion data, including various dates in the progression of a well drill from spud date to current status date, were downloaded from the COGCC Library in the statistics section under the heading *Production and Prices*. Although there are data sets for each year's well completions, the newest data set (2012) contains all completions from prior years and the new 2012 completions; therefore that data set was used for information on all well completions between 2009 and 2012. The original data set contains 55,653 observations between 1911 and 2012. Wells with repeated API numbers are considered duplicate observations and removed from the data set in order to successfully merge geographical information to well completion information. The API number is defined by the COGCC as “A well identifier assigned as defined in API (American Petroleum Institute) Bulletin D12A, as amended. The API Well Numbers are assigned by the appropriate state or federal regulatory agency.” Two data sets, one for producing wells and one for wells being drilled, are created by merging this completion data set with GIS files that include the geographical references of wells.

The data set including information on wells in the process of being drilled was created by merging the completion data with the GIS fractured well point data. Incomplete observations - observations missing either UTM coordinates or a spud date - are removed from the data set. Spud is defined by investopedia as “in the oil and gas industry, the process of beginning to drill a well” (Spud Definition | Investopedia). The spud date in the well completion data provides the date on which the drilling process for a given well began. The spud well data set had 4,035 observations after the repeated API numbers were removed from the data set.

Continuous distance variables (i.e. distSPUD) were calculated by spatially joining the spud well data with housing sales data using the point distance tool in ArcToolbox’s analysis tools to calculate distances based on unique observation identifiers assigned by ArcGIS when the data are uploaded. Distances between houses and all spud wells within a two-mile radius around the housing sales centroid are calculated and then sorted so that the minimum distance can be kept. A two-mile radius was chosen as it kept the size of the data set manageable and fell between the distance of two kilometers (1.24 miles) to the nearest shale well used by Muehlenbachs et al. (2012) and the four-kilometer (2.49 miles) radius around a property used by Boxall et al. (2005) to get a count of the number of wells within that distance to a house. A dummy variable was created to signify if the spud date fell within a 60 day window prior to the sale of the house. A time period of 60 days prior to the sale is chosen to capture activity that may be taking place during the period of escrow. As McKenzie et al. (2012) report that the disturbance of a well is highest when it is within one half mile of a house, a count variable of the number of wells being drilled within 60 days and a half mile radius of house was created in ArcGIS, spudcount. Dummy variables were generated and set equal to one for wells that were drilled within one mile of a home (spud_1mileDV) and a half mile (spud_halfmileDV) in order to examine whether significance changes as the distance band is decreased. The spud data were merged with the full housing sales data to identify which houses had a well drilled within two miles (distSPUD) and how many wells were drilled within a half mile of the house in the 60 days prior to the sale -- summary statistics are provided in Table 4.3.

Table 4.3 Summary statistics for relevant oil and gas variables

Variable	Obs	Mean	Std. Dev.	Min	Max
spudcount	13531	0.0723524	0.578895	0	11
distSPUD	4035	2029.046	777.3933	56.14307	3218.615
ln_spud	4035	7.513003	0.503907	4.027903	8.076706
spud_halfDV	13531	0.022393	0.147964	0	1
spud_1mileDV	13531	0.0936368	0.291334	0	1
num_producing	13531	4.856626	5.772465	0	41
distPROD	8802	361.4456	187.8114	6.940779	804.5902
permitcount	13531	0.024758	0.343959	0	9
distPERM	1739	2050.613	191.1916	3218.151	735.4974

To create the wells in production data set, a similar process to the wells in the process of being drilled was used, starting with merging the completion file with the GIS fractured well point data. Observations that lack geographical UTM coordinates are dropped from data set, because the distance to the well from a house cannot be measured in the absence of location data. The formation status is used to determine whether a well is in production based on formation codes provided on the COGCC website. Those with a status of abandoned well bore or completion, abandoned location, closed, dry and abandoned, plugged and abandoned, waiting on completion are dropped from the data set as they not considered producing wells. Although few wells were without a status code, these wells were also dropped from the data set due to uncertainty of whether the well is in production. The final production data set, comprised of 18,481 producing wells, was uploaded into ArcGIS using the UTM X and UTM Y coordinates to geographically reference the wells on a map and to be able to match the wells up with housing sales data. These data are spatially joined to housing sales data using the point distance tool in ArcGIS to calculate distances between points. Using the unique well and property identifiers, the data were then merged with the housing sales data to identify which properties had a producing well within half mile of the home at the time it was sold, and to get a count of the number of producing wells exist within one half mile at the time of sale. Due to the volume of producing wells relative to wells being drilled and the perceived level of disturbance associated with a well in production compared to the drilling process[9], the distance to the nearest well in production within a half mile was calculated, unlike the spud and permitted wells with a two mile search radius, was chosen. Summary statistics for the variables num_producing, a count variable generated in STATA to get a count of the number of producing wells within a half mile of any

[9] The level of truck traffic and the amount of visual disturbance decreases significantly once a well is finished being drilled and it is moved into production.

house at the time of sale, and distPROD, a variable measuring the distance to the nearest producing well within a half mile at the time of sale, are provided in Table 4.3.

Data on permitted wells was downloaded from the COGCC website's GIS maps section. Unfortunately, these data only date back to January 25th, 2011 and do not span over the entire temporal range of this study. As the GIS database file contained both geographical and temporal data, these data did not need to be merged as was necessary with the other well data. With just 1,739 observations, the permitted well data set was much smaller than the other well data sets. The permitted well data were spatially joined with housing sales data in the same manner in which the spud and production data were. A count of wells permitted within a half mile of a property and within 60 days of the sale as well as a minimum distance to the nearest permitted well within a two miles and 60 days of sale were both variables generated using these and the housing sales data. Summary statistics on these two variables, countpermit and distPERM, are provided in Table 4.3 along with the statistics on other well data.

A dummy variable (waterwell) indicating whether there is a domestic-use water well on the property was generated by spatially joining the property sales shape file with water well point data in ArcGIS, where a distance of zero meters to the nearest water well signifies the presence of one on the property. Waterwell is activated and equal to one when there is a water well on the property, and it is equal to zero if there exists no water well on the property. This variable serves as a proxy to indicate whether having a domestic-use water well has a relationship with the sale price of a house. However, as these are secondary data, there is no certainty that the inhabitants of the home are not also dependent upon municipal water. Data on whether a house is served by a public water service district and uses municipal water were unfortunately not available[10]. This variable was generated using GIS sales polygons.

[10] From an email correspondence with the Weld County office of the assessor from Feb. 17, 2013:

Summary statistics for the two separate groups, sorted by distSPUD, are also provided in the Appendix. Table 4A provides summary statistics for the all variables with a well being drilled within 2 miles and within 60 days prior to the sale transaction date. Table 5A provides summary statistics for all variables if there was not a well being drilled within 2 miles, up to 60 days prior to the date of sale. One key observation to be made in these tables is that the mean real sale price is within just under $5,000 dollars whether there was a well present ($212,560) or not ($ 216,365), with transactions that did not have a well being drilled on the higher end, and the minimum and maximum prices line up nearly identically as well. These tables are provided in order to give an idea what differences in the data exist between the control versus treatment model specifications.

"Our office does not track what type of water properties use. I believe the Division of Water Resources issues well permits, so they may have information on which properties use well water. However, properties with wells often have municipal water as well. "

CHAPTER 5: RESULTS

The six specifications of the model outlined in Chapter 3 were run and tested for overall model statistical significance, statistical significance of the individual parameter estimates, proper specification, and for disturbances in the error term using the hypothesis tests discussed in section 3.3. The parameter estimates, their standard errors, and level of significance are reported in Table 5.2. Results of the base and treatment specifications for the linear, semi-log, and double-log models are presented in that order in the table. The R-squared increased significantly from the linear to the semi-log specification, and then slightly from the semi-log to the double-log. This is to be expected, however, because the log transformation of the dependent variable compresses the scale of the dependent variable, which reduces the variability in it overall. Drill activity treatment variables were tested for cross-correlations between the variables to determine whether they might cause multicollinearity issues in the regression analysis. The cross-correlations – reported in Table 5.1 – between the three variables were low for the large sample size, therefore all of them were included in the treatment regressions.

Table 5.1 Cross-correlations of treatment variables

	spudcount	num_producing	distSPUD
spudcount	1.0000		
num_producing	0.1743	1.0000	
distSPUD	-0.4478	-0.1230	1.0000

Table 5.2 Results of the base and treatment models for the linear, semi-log, and double-log models

	Linear Base	*Linear Treatment*	*Semi-log Base*	*Semi-log Treatment*	*Double-log Base*	*Double-log Treatment*
	salep_real	**salep_real**	**ln_sales**	**ln_sales**	**ln_sales**	**ln_sales**
lotsize	18467.8***	14796.3***	0.0643***	0.0550***		
	(1719.0)	(2636.5)	(0.0052)	(0.0052)		
lotsize2	-496.9***	-336.7***	-0.00181***	-0.00133***		
	(67.2)	(70.1)	(0.00023)	(0.00016)		
baths	3306.9*	3972.3+	0.00997*	0.00852	0.0158***	0.0112
	(1513.7)	(2385.7)	(0.0043)	(0.0059)	(0.0043)	(0.0071)
age	-802.8***	-781.2***	-0.00518***	-0.00511***	-0.00529***	-0.00506***
	(46.6)	(93.0)	(0.00021)	(0.00025)	(0.00021)	(0.00038)
ressf	45.71***	18.38	0.000590***	0.000574***		
	(10.9)	(23.2)	(0.000019)	(0.000027)		
ressf2	0.0110***	0.0175**	-5.30e-08***	-5.01e-08***		
	(0.0026)	(0.0057)	(4.1e-09)	(5.8e-09)		
outbuildings f	6.980**	7.866**	0.0000523** *	0.0000554***	0.0000354***	0.0000324***
	(2.33)	(2.88)	(0.0000084)	(0.0000073)	(0.0000059)	(0.0000084)
porchsf	41.75***	45.26***	0.000125***	0.000148***		
	(5.25)	(9.79)	(0.000011)	(0.000015)		
remodel	-3904.6*	-2354.8	0.00977	0.0175	0.00795	0.0135
	(1714.6)	(2806.4)	(0.0083)	(0.012)	(0.0083)	(0.014)
garage	7689.9*	12816.0**	0.0869***	0.119***	0.0755***	0.110***
	(3246.7)	(4931.4)	(0.016)	(0.019)	(0.017)	(0.030)
finish_bsmnt	30232.3***	30806.1***	0.140***	0.144***	0.124***	0.131***
	(1664.1)	(2606.7)	(0.0055)	(0.0088)	(0.0054)	(0.0093)
GREELEY	-25402.7***	-25833.6***	-0.0812***	-0.102***	-0.0854***	-0.104***
	(1451.4)	(2560.7)	(0.0050)	(0.010)	(0.0049)	(0.0096)
RURAL	2895.5	7637.8	0.0167	0.0115	-0.0307**	-0.0470*
	(3421.2)	(6438.3)	(0.012)	(0.016)	(0.011)	(0.020)
y2010	2947.0+	639.4	0.00684	0.00663	0.00611	0.00274
	(1726.7)	(2596.6)	(0.0056)	(0.010)	(0.0055)	(0.010)

y2011	3241.4*	3762.7	0.00184	0.00723	0.000366	0.00378
	(1593.0)	(2718.7)	(0.0056)	(0.010)	(0.0055)	(0.0099)
y2012	19529.5***	15889.5***	0.0921***	0.0810***	0.0886***	0.0724***
	(1659.7)	(2655.5)	(0.0054)	(0.0098)	(0.0053)	(0.0099)
hwy_mile	1871.7	3959.4+	-0.00274	0.00802	0.00199	0.0140+
	(1467.9)	(2112.2)	(0.0043)	(0.0075)	(0.0043)	(0.0076)
pct_hisp	-31960.5***	-62127.0***	-0.567***	-0.603***	-0.490***	-0.556***
	(6852.6)	(10760.6)	(0.027)	(0.044)	(0.027)	(0.046)
pct_own	-58902.4***	-48186.1***	-0.138***	-0.135**	-0.151***	-0.113*
	(8552.9)	(11872.5)	(0.031)	(0.044)	(0.031)	(0.046)
hh_inc	0.614***	0.467***	0.00000184* **	0.00000129***	0.00000173***	0.00000102***
	(0.044)	(0.072)	(0.00000015)	(0.00000027)	(0.00000015)	(0.00000025)
pct_bachlr	119646.8***	104249.5***	0.365***	0.401***	0.427***	0.456***
	(9464.0)	(12292.9)	(0.026)	(0.048)	(0.026)	(0.044)
spudcount		-1910.9*		-0.00619+		-0.00620
		(874.5)		(0.0037)		(0.0040)
distSPUD		-2.739*		-0.00000648		
		(1.37)		(0.0000050)		
num_producing		-131.4		0.000398		0.00103+
		(171.5)		(0.00060)		(0.00060)
ln_lotsize						0.115***
					(0.0045)	(0.0081)
ln_ressf					0.628***	0.624***
					(0.011)	(0.020)
ln_porchsf						0.0221***
					(0.0024)	(0.0044)
ln_spud						-0.0114
						(0.0086)
Intercept	35945.7**	71924.9**	11.16***	11.19***	7.468***	7.565***
	(13175.0)	(25536.1)	(0.040)	(0.056)	(0.081)	(0.17)

N	13531	4035	13531	4035	13223	3940
R^2	0.61	0.704	0.782	0.781	0.791	0.79
adj. R^2	0.61	0.703	0.782	0.78	0.79	0.789
F-statistic	854	336.2	1955	595.8	2160.7	556.9

Robust standard errors in parentheses + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The linear specification performs decently in both the control and treatment models. Parameter coefficient estimates align with the expected signs in Table 3.1for all variables with the exception of percentage of homeowners. Pct_own is highly statistically significant at the 0.0001 level with negative coefficients of - $58,902 and - $48,186 for the base and treatment models respectively, meaning that a one percent increase in homeownership decreases real housing sales price by $58,902 in the absence of well activity and $48,186 with it, at the mean. The addition of the well-activity variables increases the adjusted R-squared from .61 to .70, indicating that the addition of the well-related variables improves the predictive power of the model due to the large increase in the R-squared. Simply adding more explanatory variables to a model usually increases the R-squared even if the variables are not significant, which is the reason for the focus on the adjusted R-squared. Spudcount, however, is statistically significant at the 95% confidence level, and has the expected [negative] sign despite being fairly small in magnitude. The implicit price associated with spudcount is -$1,911, the estimate of β_{22}. This indicates that, ceteris paribus, each additional well drilled within one half mile of a house up to 60 days prior to its sale decreases the sale price by $1,911, a 0.899 percent decrease in house prices at the mean. DistSPUD is also statistically significant in the linear model, but has a very low effect on housing prices and the wrong sign on the coefficient; for each meter farther from a house the well is drilled, the price decreases by $2.74. The null hypothesis on the t-tests for significance on num_producing fails to be rejected, meaning it is not statistically different from zero.

The inclusion of these well-related variables leads to a better fit of the overall linear model, shown by an increase in the R-squared of .094. However, graphical analysis of the residuals show a non-constant variance in them, exhibited in Figures 5.1 and 5.2. The post-estimation Breusch Pagan tests of both linear models also revealed heteroskedasticity as the null hypothesis of homoskedasticity could not be rejected at the 0.0001 confidence level. The regression results reported in Table 5.1 are those from the model run using White's robust standard errors, as are Figures 5.1 and 5.2, indicating that the robust standard errors do not provide a solution for the heteroskedasticity in the error term. The linear specifications are considered the worst of the three for this reason.

To determine whether multicollinearity presented an issue in the regression results, variance inflation factors (VIFs) were estimated post-regression. As a rule of thumb, a variable with a VIF of 5 or larger is considered highly multicollinear. Three variables had VIFs greater than 5 in both the base and treatment models: residential square feet, residential square feet squared, and lot size. These variables are important structural variables describing the property and are not removed from the regression as it is likely the correlation between the ressf (lotsize) and ressf^2 (lotsize^2) that is causing the VIF to be so high.

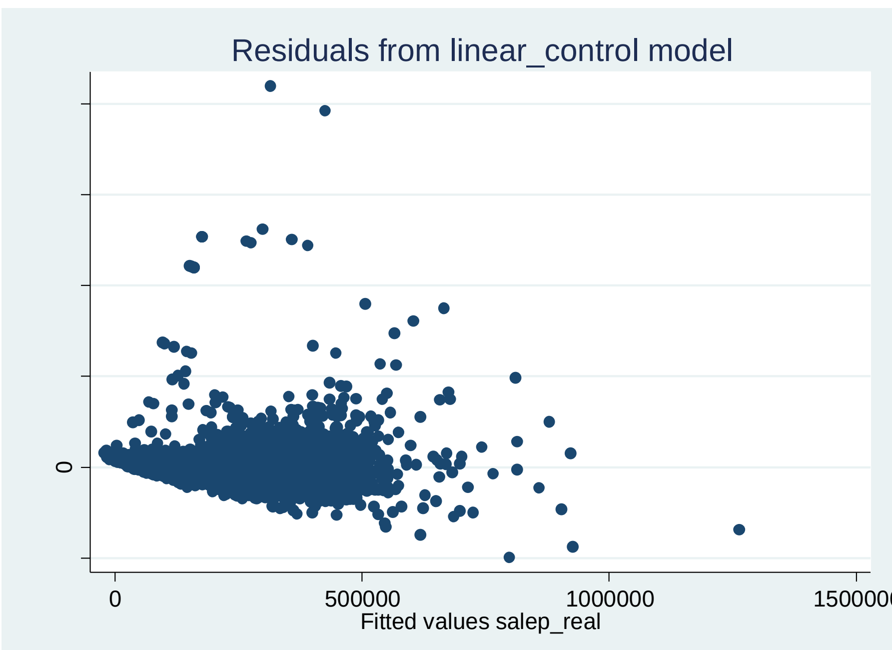

Figure 5.1 Residuals obtained from the robust estimation of the linear control_model

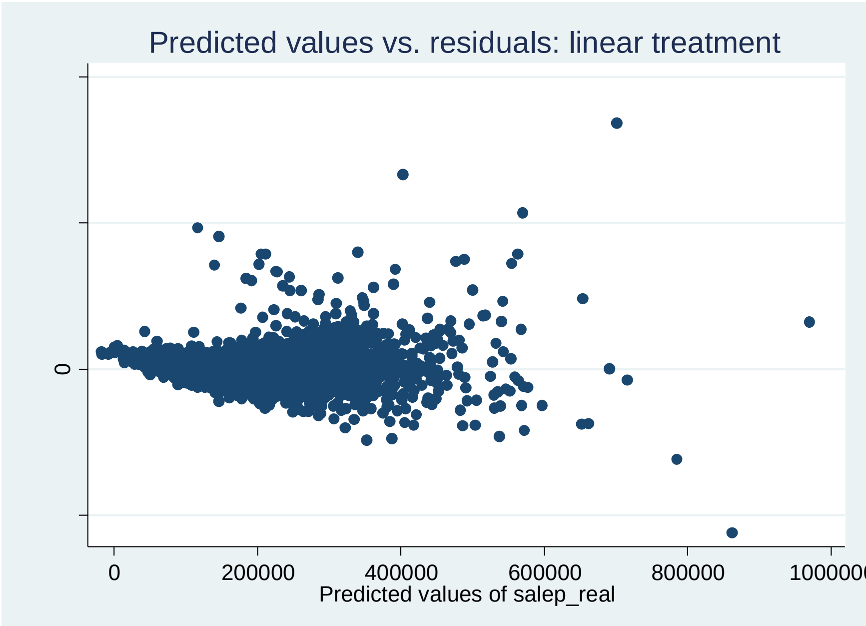

Figure 5.2 Residuals obtained from the robust estimation of the linear treatment model

The semi-log specifications provide an improvement over the linear specifications in terms of goodness-of-fit of the functional form to the data apparent in the increase in the R-squared. Parameter estimates maintain their levels of significance and signs as compared to the linear specifications except for the garage coefficient estimate, which becomes significant at a higher level of confidence than in the linear models. Spudcount drops from being significant at the 95% confidence level to just the 90% confidence level – specifically it is significant at a confidence level of 93%. Unlike in the linear specification, the R-squared between the semi-log base and treatment models actually decreases very slightly by 0.001 as shown in Table 5.2. While the variables of interest lose significance in this model, it provides a better fit to the data over the linear specification. The F-statistic obtained from the Ramsey RESET test run post-estimation of the semi-log treatment model of 1.99 provides statistical evidence to fail to reject the null hypothesis of no omitted variables. The residuals also do not appear to follow a heteroskedastistic pattern, as is shown in Figure 5.3, which are the residuals obtained from the semi-log treatment estimated model. Because the Breusch-Pagan test indicated heteroskedasticity, the regression was run using White's robust standard errors (these are the results reported in Table 5.2). The errors obtained from the robust standard error semi-log regression are shown in Figure 5.4 and appear to be normally distributed.

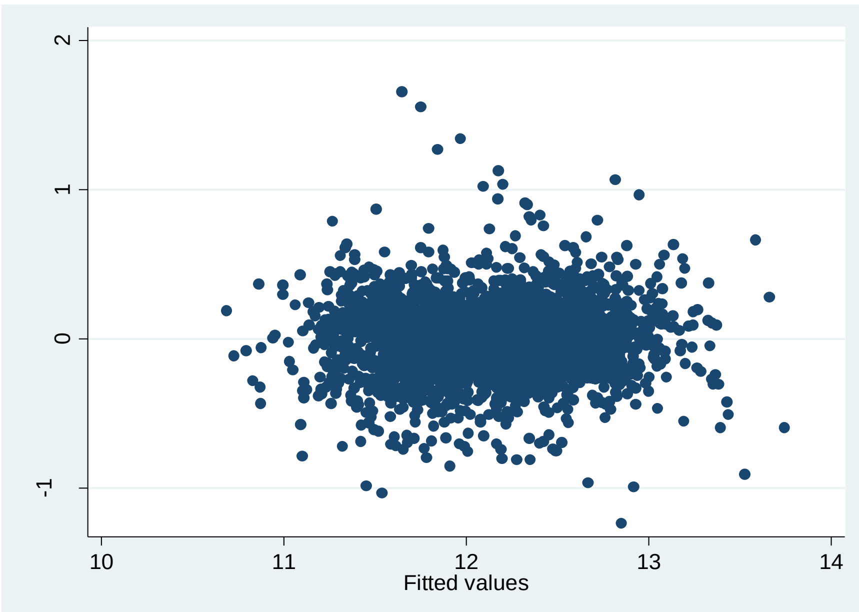

Figure 5.3 Residuals obtained from the robust estimation of the semi-log treatment model

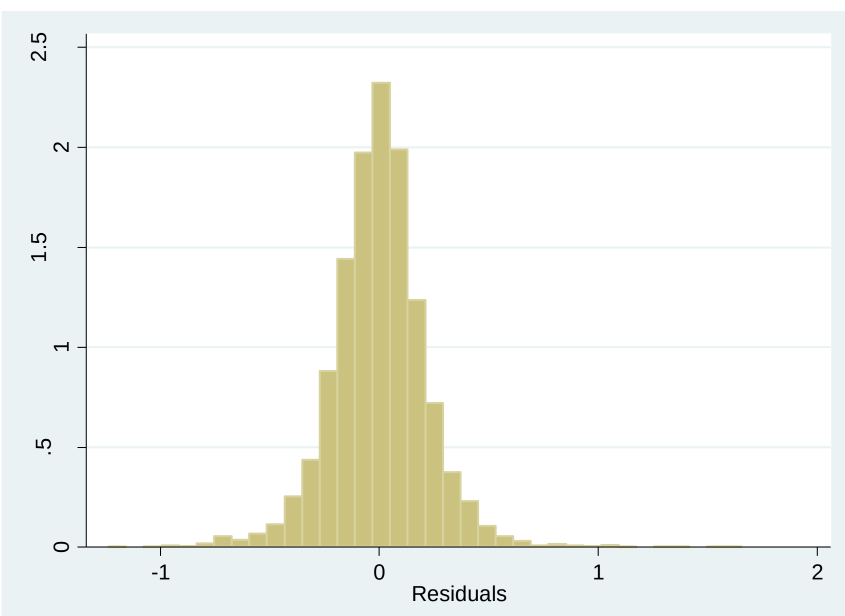

Figure 5.4 Distribution of residuals obtained from the robust estimation of the semi-log treatment model

The double log specifications produce results very similar to those of the semi-log specifications. Aside from the logged-independent variables that were transformed, the parameter estimates from the double-log model are nearly the same as the semi-log estimates for the two respective models, control and treatment. The distribution graph of the residuals and scatter plot of the residuals versus predicted values of logged sales prices look almost identical to their semi-log counterparts. This can be seen by comparing Figure 5.3 to Figure 5.5 and Figure 5.4 to Figure 5.6, although there appear to be longer tails in the distribution of the double-log model. The Ramsey RESET test indicates that there are omitted variables in both the base and treatment models, as it did in the other specifications. However, recall that Cropper et al. (1988) suggested that the double-log model performs better than many other specifications when there are missing or omitted variables. The long tail in the distribution of the disturbance term in the double log specification and the statistically significant result on the test for omitted variables give reason to favor the semi-log specification over the double-log specification, despite the lower R-squared measure.

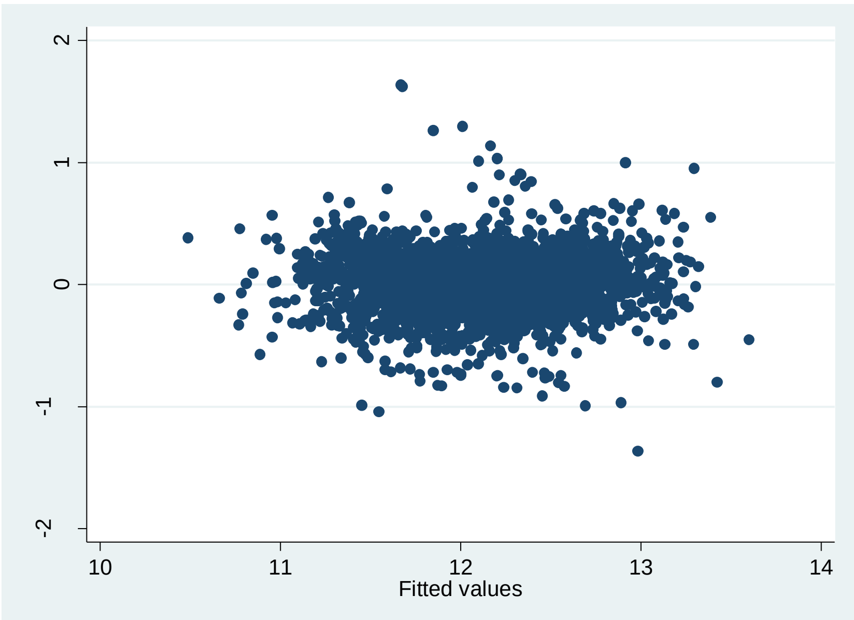

Figure 5.5 Residuals obtained from the robust estimation of the double-log treatment model

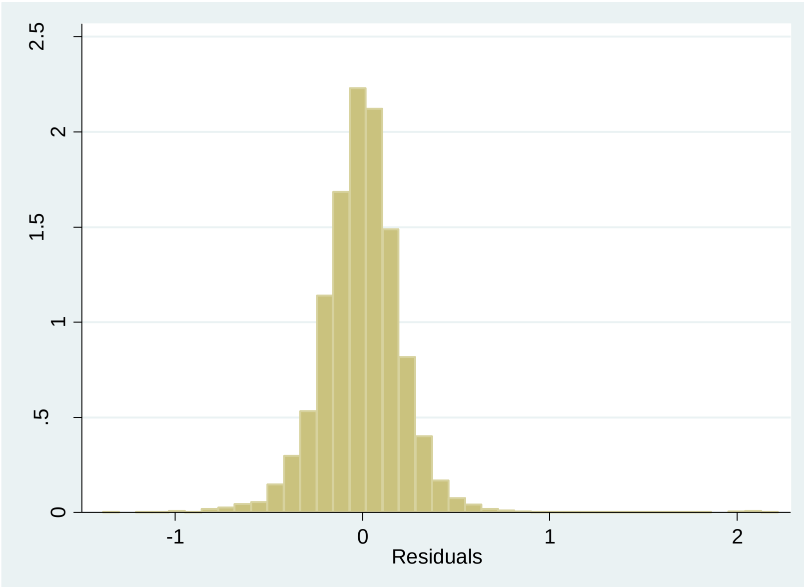

Figure 5.6 Distribution of residuals obtained from the robust estimation of the double-log treatment model

Table 5.3 Treatment variable implicit prices for statistically significant parameters[11]

	Linear	Semi Log	Double Log
spudcount	- $1,911	- $1,316	N.S.
num_producing	N.S.	N.S.	$219
distSPUD	-$2.74	N.S.	--
ln_spud	--	--	N.S.
Gain in R^2 with treatment	+.094	-.001	-.001

The implicit prices were calculated using the mean housing sale price ($212,560) for all transactions that had a well being drilled within 2 miles of the home up to 60 days prior to the sale of the home[12]. For the double log model, the implicit price must be calculated at a level of the variable, num_producing, for which the mean value of 6.6 wells producing within a half-mile of the house at the time of sale is used. The implicit price for the variable spudcount is interpreted as a decrease in the sale price of the house of $1,911 for the linear specification and

[11] N.S. is reported if the coefficient estimate was not statistically different from zero; -- is reported for variables not present in the specification.

[12] Full summary statistics for all variables based on this contingency are reported in Table 4A in the Appendix.

slightly less at $1,316 for the semi-log specification per new well being drilled within a half-mile of the house up to 60 days prior to the sale of the house. The interpretation of the implicit price for num_producing, the only statistically significant treatment variable in the double log model, is similar to the interpretation of the spudcount variable. However, num_producing is estimated to increase housing values by $219 per well in production within a half mile of the house at the time of the sale. distSPUD, significant only in the linear specification, has a somewhat confusing interpretation of the implicit price calculation. For each meter farther from a house the drilling occurs, up to a maximum of 2000 meters and up to 60 days prior to the sale date, the value of the house decreases by $2.74. While the implicit price on spudcount matches expectations and is negative, the implicit prices on distSPUD and num_producing are the opposite of their hypothesized effect on housing values. Due to the absence of data on mineral rights and royalty payments in this data set, however, this may be accurately capturing the relative effects on housing values if the landowners are receiving money for the drilling.

The results from the original empirical specifications of the model exhibited little change due to the inclusion of the oil and gas related variables, although the number of wells being drilled within 60 days of the sale and within a half-mile radius of the house was statistically significant with the expected negative sign in the linear and semi-log specifications. One potential reason for the general insignificance of the well-related variables is the increased workforce and subsequent demand for housing in the area, which may be driving housing values up as it has in areas of North Dakota (Platt, 2013). Alternate specifications, which included an oil and gas sector employment variable, to the six original model specifications described in Chapter 3 were tested to determine if the relationship between oil and gas employment and housing values held statistical significance. The only difference in these six specifications to the original six is the addition of the variable allemp and the absence of the year dummy variables (y2010, 2011, and y2012), which were removed due to high correlation to allemp. Of the three year dummy variables, y2012 was the one that was consistently statistically significant across different model specifications, whereas allemp was statistically significant across all model specifications. These high correlation coefficients are reported in Table 5.4 below.

Table 5.4 Correlations between oil and gas sector employment and year dummy variables

	y2010	y2011	y2012	allemp
y2010	1.0000			
y2011	-0.3104	1.0000		
y2012	-0.3585	-0.3642	1.0000	
allemp	-0.5319	0.0885	0.8647	1.0000

The regression results from the alternate specifications, including coefficient estimates and standard errors as well as the number of observations (N), R-squared, and F-statistic, are reported in Table 5.5. In comparing the results from the original six specifications (Table 5.2) to the alternate specifications (Table 5.5), one observes almost no difference in the overall model results or parameter estimates. The semi-log treatment model that included employment data, however, was the only model specification that did not test positive for omitted variable bias by the Ramsey RESET test. The null hypothesis that the model had no omitted variables failed to be rejected as the F-statistic for $F_{(3, 4009)}$ was 1.88. While the results from the models that included employment data were, overall, very similar to the original six specifications that included year dummy variable instead of employment data. Changes in treatment variable significance and R-squared for the alternate models, reported in Table 5.5, nearly mirror the results displayed in Table 5.2 for the original six models.

These models were also run with the inclusion of an interaction between wellwater, a dummy variable set equal to one if there was a domestic-use water well on the property, and distSPUD. Muehlenbachs et al. (2012) found a stronger negative effect on housing values if the house was served by well water than if it was served by the municipal water supply. The results from these regressions were very similar to those run without the wellwater variable: there was zero change in the adjusted R-squared values for the linear, semi-log, and double log specifications with the addition of wellwater to the same models reported in Table 5.5. Waterspud was only statistically different from zero at the 90% confidence level in the double log model. Full results from these regressions are reported in Table 6A in the Appendix instead of in this section due to the lack of change. Interacting wellwater with spudcount resulted in too few non-zero observations and too little variation in the data to include in regression analysis.

Table 5.5 Results of the base and treatment models for the linear, semi-log, and double-log models including Weld County Oil and Gas sector employment data

	Linear Base	*Linear Treatment*	*Semi-log Base*	*Semi-log Treatment*	*Double-log Base*	*Double-log Treatment*
Dep. Var.	**salep_real**	**salep_real**	**ln_sales**	**ln_sales**	**ln_sales**	**ln_sales**
lotsize	18624.2***	14884.4***	0.0650***	0.0555***		
	(-1721.5)	(-2635.6)	(-0.0052)	(-0.008)		
lotsize2	-499.7***	-337.5***	-0.00182***	-0.00134***		
	(-67.7)	(-70.1)	(-0.00023)	(-0.00025)		
baths	3399.6*	4047.9+	0.0104*	0.00902	0.0162***	0.0115
	(-1511.8)	(-2395.2)	(-0.0043)	(-0.0067)	(-0.0043)	(-0.0071)
age	-801.0***	-780.0***	-0.00517***	-0.00510***	-0.00528***	-0.00506***
	(-46.6)	(-92.9)	(-0.00021)	(-0.00038)	(-0.00021)	(-0.00038)
ressf	45.71***	18.42	0.000589***	0.000574***		
	(-10.9)	(-23.1)	(-1.9E-05)	(-4.3E-05)		
ressf2	0.0109***	0.0175**	-5.29e-08***	-5.01e-08***		
	(-0.0026)	(-0.0057)	(-4.10E-09)	(-9.80E-09)		
outbuildingsf	6.906**	7.796**	0.0000519***	0.0000548***	0.0000354***	0.0000323***
	(-2.33)	(-2.89)	(-8.5E-06)	(-1.1E-05)	(-6E-06)	(-8.5E-06)
porchsf	41.89***	45.41***	0.000126***	0.000149***		
	(-5.24)	(-9.78)	(-1.1E-05)	(-0.00002)		
remodel	-4069.1*	-2453.4	0.00829	0.0171	0.00656	0.0132
	(-1714.6)	(-2808.3)	(-0.0084)	(-0.014)	(-0.0083)	(-0.014)
garage	7542.0*	12853.2**	0.0856***	0.118***	0.0745***	0.110***
	(-3246.3)	(-4905.1)	(-0.016)	(-0.027)	(-0.017)	(-0.03)
finish_bsmnt	30095.6***	30595.5***	0.140***	0.142***	0.124***	0.130***
	(-1663.4)	(-2613.7)	(-0.0055)	(-0.0092)	(-0.0054)	(-0.0094)
GREELEY	-25484.4***	-25837.1***	-0.0815***	-0.102***	-0.0857***	-0.104***
	(-1452.3)	(-2568.1)	(-0.005)	(-0.0097)	(-0.005)	(-0.0097)
RURAL	2398.5	7210.1	0.0139	0.00869	-0.0334**	-0.0497*
	(-3425.6)	(-6439.5)	(-0.012)	(-0.022)	(-0.011)	(-0.02)
hwy_mile	1982.4	4301.8*	-0.00207	0.0102	0.00255	0.0159*

	(-1471)	(-2120.8)	(-0.0044)	(-0.0078)	(-0.0043)	(-0.0076)
pct_hisp	-32309.7***	-62408.5***	-0.569***	-0.604***	-0.491***	-0.557***
	(-6850.2)	(-10835.6)	(-0.027)	(-0.046)	(-0.027)	(-0.046)
pct_own	-59038.5***	-48621.0***	-0.139***	-0.138**	-0.153***	-0.115*
	(-8543.1)	(-11815.6)	(-0.031)	(-0.047)	(-0.031)	(-0.046)
hh_inc	0.616***	0.467***	0.00000185***	0.00000129***	0.00000174***	0.00000102***
	(-0.044)	(-0.072)	(-1.5E-07)	(-2.6E-07)	(-1.5E-07)	(-2.5E-07)
pct_bachlr	119547.0***	104763.3***	0.364***	0.405***	0.426***	0.461***
	(-9459.7)	(-12311.2)	(-0.026)	(-0.044)	(-0.026)	(-0.044)
allemp	582.2***	490.9***	0.00289***	0.00247***	0.00282***	0.00228***
	(-53.4)	(-77.6)	(-0.00016)	(-0.00029)	(-0.00016)	(-0.00029)
spudcount		-1856.0*		-0.00571		-0.00563
		(-876.8)		(-0.0038)		(-0.004)
num_producing		-151.9		0.000236		0.000876+
		(-170.7)		(-0.00061)		(-0.0006)
distSPUD		-2.633+		-5.5E-06		
		(-1.36)		(-5E-06)		
ln_lotsize					0.116***	0.116***
					(-0.0045)	(-0.0081)
ln_ressf					0.628***	0.622***
					(-0.011)	(-0.02)
ln_porchsf					0.0222***	0.0265***
					(-0.0024)	(-0.0044)
ln_spud						-0.0099
						(-0.0086)
Intercept	-56140.1***	-6073.9	10.70***	10.80***	7.018***	7.199***
	(-14206.5)	(-29425)	(-0.048)	(-0.088)	(-0.084)	(-0.18)
N	13531	4035	13531	4035	13223	3940
R^2	0.61	0.704	0.781	0.779	0.789	0.789
adj. R^2	0.609	0.702	0.78	0.778	0.789	0.788
F-statistic	903.7	365	2131.9	563.6	2388	606.3

Robust standard errors in parentheses + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.6 Treatment variable implicit prices for statistically significant parameters[13]

	Linear	Semi Log	Double Log
allemp	$ 491	$ 525	$ 485
spudcount	- $1,856	- $1,214[14]	N.S.
num_producing	N.S.	N.S.	$186
distSPUD	- $2.63	N.S.	--
ln_spud	--	--	N.S.
Gain in R^2 with treatment	+.094	-.002	No change

Implicit prices, calculated using the mean housing sales price for the sample for the semi-log and double-log models as suggested by Taylor (2003), are reported for the statistically significant treatment variables from the alternate models in Table 5.6. The implicit prices reported in Table 5.6 are very similar to those from Table 5.3, however the implicit price has dropped slightly in each of the categories. Since allemp is also a treatment variable, implicit prices for it are also reported in Table 5.6. Allemp represents thousands of hours worked by all oil and gas sector employees in Weld County, and corresponds to each housing transaction based on that figure during the month of the sale. An increase of 1000 hours worked in the sector, a proxy for increased housing demand, increases the sale price of the home by around $500 for all specifications. Employment in the oil and gas sector appears to increase housing values only slightly, while an increase in the number of fracking sites around the home near the time of sale has a larger and negative effect.

Removing the continuous distance variable, distSPUD, increases the sample size from 4035 to the original 13531 observations available for sampling, in order to remove any biases imposed on the data due to sample selection. Running the same three functional forms (linear, semi-log, and double log) without distSPUD yields nearly identical results to those from the

[13] N.S. is reported if the coefficient estimate was not statistically different from zero; -- is reported for variables not present in the specification.

[14] While spudcount is not statistically significant in the semi log specification, it is very close to being significant (robust standard error p-value = 0.131).

models that include the distSPUD variable. There is a slight increase in the R-squared value, but that can be attributed to the larger and different sample used in these regressions. Results from these regressions are reported in Table 5.7. Implicit prices for the variable spudcount align very closely with those from the regressions that included distSPUD and had a smaller sample size. The implicit price of an additional well being drilled within a half-mile of a house for the full sample obtained from the linear regression model was - $1,791, slightly smaller than in other specification at - $1,911. Spudcount, while not quite statistically significant in the semi-log model with a p-value of 0.121, had an implicit price of - $1,117, again slightly less than the smaller sample for which the implicit price was - $1,214. Using the full sample of single-family residential sales yields slightly lower implicit prices associated with spudcount but similar levels of statistical significance, indicating a lower impact on housing values throughout the county.

Table 5.7 Results of the linear, semi-log, and double-log models

	Linear full sample	*Semi-log full sample*	*Double log full sample*
	salep_real	**ln_sales**	**ln_sales**
lotsize	18634.8***	0.0651***	
	(1720.9)	(0.0052)	
lotsize2	-499.8***	-0.00182***	
	(67.8)	(0.00023)	
baths	3355.7*	0.0102*	0.0160***
	(1514.7)	(0.0044)	(0.0043)
age	-802.2***	-0.00518***	-0.00530***
	(47.1)	(0.00021)	(0.00021)
ressf	45.69***	0.000589***	
	(10.9)	(0.000019)	
ressf2	0.0110***	-5.28e-08***	
	(0.0026)	(4.1e-09)	
outbuildingsf	6.908**	0.0000519***	0.0000352***
	(2.33)	(0.0000085)	(0.0000060)
porchsf	41.92***	0.000126***	
	(5.26)	(0.000011)	
remodel	-4030.1*	0.00841	0.00673
	(1715.8)	(0.0084)	(0.0083)

garage	7561.6*	0.0857***	0.0745***
	(3247.5)	(0.016)	(0.017)
finish_bsmnt	30162.0***	0.140***	0.124***
	(1657.3)	(0.0055)	(0.0055)
GREELEY	-25466.9***	-0.0809***	-0.0839***
	(1470.0)	(0.0050)	(0.0050)
RURAL	2304.8	0.0134	-0.0352**
	(3438.3)	(0.012)	(0.011)
hwy_mile	1999.9	-0.00192	0.00291
	(1466.3)	(0.0044)	(0.0043)
pct_hisp	-32497.3***	-0.571***	-0.494***
	(6991.8)	(0.027)	(0.027)
pct_own	-59405.4***	-0.142***	-0.158***
	(8651.2)	(0.031)	(0.031)
hh_inc	0.615***	0.00000185***	0.00000174***
	(0.044)	(0.00000015)	(0.00000015)
pct_bachlr	119449.0***	0.363***	0.425***
	(9495.1)	(0.026)	(0.026)
allemp	576.4***	0.00285***	0.00276***
	(51.2)	(0.00016)	(0.00016)
spudcount	-1791.4*	-0.00519	-0.00438
	(789.0)	(0.0034)	(0.0033)
num_producing	66.31	0.000498	0.00114**
	(137.0)	(0.00038)	(0.00038)
ln_lotsize			0.116***
			(0.0045)
ln_ressf			0.627***
			(0.011)
ln_porchsf			0.0226***
			(0.0024)
Intercept	-54870.1***	10.71***	7.032***
	(14335.9)	(0.048)	(0.085)
N	13531	13531	13223
R^2	0.610	0.781	0.789
Adj. R^2	0.609	0.780	0.789
F	828.3	1936.3	2148.3

Robust standard errors in parentheses =" + p<0.10 * p<0.05 ** p<0.01 *** p<0.001"

Table 5.8 Results of the treatment models for the linear and semi-log models by rural vs. urban & Greeley vs. non-Greeley

	Urban semi log	*Rural semi log*	*Urban linear*	*Rural linear*	*Greeley semi log*	*Non-Greeley semi log*	*Greeley linear*	*Non-Greeley linear*
Dep. var	**ln_sales**	**ln_sales**	**salep_real**	**salep_real**	**ln_sales**	**ln_sales**	**salep_real**	**salep_real**
lotsize	0.295***	0.0462***	102757.7***	8694.3**	0.548***	0.0559***	137411.6***	15313.9***
	(0.022)	(0.0091)	(5515.2)	(2724.6)	(0.12)	(0.0047)	(24634.0)	(1292.3)
lotsize2	-0.0396***	-0.00109***	-14353.0***	-198.3*	-0.122	-0.00133***	-32936.5*	-344.9***
	(0.0044)	(0.00026)	(1114.2)	(78.5)	(0.080)	(0.00015)	(16566.7)	(40.8)
baths	0.00334	0.0747*	2132.5	27889.5**	0.0198	0.00725	10089.6***	3446.0+
	(0.0057)	(0.029)	(1450.5)	(8618.6)	(0.013)	(0.0066)	(2738.7)	(1824.6)
age	-0.00612***	-0.00146	-1052.4***	62.28	-0.00642***	-0.00460***	-990.9***	-646.6***
	(0.00025)	(0.00092)	(64.5)	(274.2)	(0.00047)	(0.00030)	(97.4)	(82.5)
ressf	0.000463***	0.000963***	-15.76*	142.4***	0.000498***	0.000601***	-11.07	31.93***
	(0.000028)	(0.00011)	(7.00)	(32.0)	(0.000053)	(0.000032)	(11.0)	(8.89)
ressf2	-3.05e-08***	-.00000013***	0.0234***	-0.00817	-4.70e-08***	-5.59e-08***	0.0189***	0.0148***
	(5.9e-09)	(0.000000021)	(0.0015)	(0.0062)	(0.000000012)	(6.8e-09)	(0.0024)	(0.0019)
outbuildingsf	0.0000443+	0.0000471***	-13.43*	9.295**	-0.0000210	0.0000535***	-1.690	7.762***
	(0.000025)	(0.000011)	(6.40)	(3.37)	(0.000069)	(0.0000075)	(14.3)	(2.09)
porchsf	0.000117***	0.0000723	37.96***	-1.535	0.000121***	0.000161***	26.99***	53.13***
	(0.000016)	(0.000056)	(3.99)	(16.8)	(0.000027)	(0.000018)	(5.63)	(5.09)
remodel	-0.00235	0.0725+	-5741.0+	-3376.8	0.0453*	-0.0104	3925.2	-9017.0*
	(0.012)	(0.043)	(3053.7)	(12962.7)	(0.018)	(0.015)	(3636.5)	(4291.3)
garage	0.117***	0.0241	3690.2	-1296.6	0.0928**	0.119***	4053.0	11419.7+
	(0.022)	(0.051)	(5526.0)	(15333.2)	(0.036)	(0.022)	(7402.1)	(6221.6)
finish_bsmnt	0.111***	0.213***	20389.0***	49229.2***	0.139***	0.142***	17353.6***	34385.3***
	(0.0085)	(0.043)	(2154.6)	(12991.6)	(0.017)	(0.010)	(3596.2)	(2831.1)
hwy_mile	-0.00232	-0.000576	197.0	6015.7	0.00781	0.0102	-460.1	4327.5+
	(0.0072)	(0.038)	(1836.7)	(11488.8)	(0.014)	(0.0089)	(2899.0)	(2458.8)
pct_hisp	-0.773***	0.160	-101544***	24069.4	-0.763***	-0.607***	-64990.7***	-86336.6***
	(0.042)	(0.24)	(10639.8)	(70785.6)	(0.086)	(0.057)	(17860.9)	(15785.1)
pct_own	-0.0878*	0.340	-31140.9**	46653.2	-0.212***	-0.132+	-42988.2***	-67757.5**

	(0.041)	(0.34)	(10502.6)	(100779.0)	(0.062)	(0.074)	(12795.6)	(20600.9)
hh_inc	0.00000103***	0.00000381*	0.374***	0.803	0.00000140*	0.00000112***	0.467***	0.432***
	(0.00000026)	(0.0000017)	(0.067)	(0.51)	(0.00000059)	(0.00000032)	(0.12)	(0.088)
pct_bachlr	0.264***	0.792*	73146.5***	168969.5+	0.0654	0.480***	31500.0	120768.6***
	(0.044)	(0.32)	(11189.7)	(94219.0)	(0.10)	(0.056)	(21395.6)	(15513.9)
allemp	0.00274***	0.00139	545.4***	409.7	0.00255***	0.00250***	377.6***	492.8***
	(0.00028)	(0.0015)	(70.9)	(435.5)	(0.00054)	(0.00034)	(112.5)	(94.2)
spudcount	-0.00650+	0.0110	-1804.5+	3742.2	-0.0146+	-0.00299	-2354.3	-1557.2
	(0.0036)	(0.018)	(924.6)	(5458.3)	(0.0080)	(0.0042)	(1656.1)	(1175.9)
num_producing	0.00139*	-0.00137	244.8	-1357.0	-0.000743	0.000417	93.91	-67.69
	(0.00059)	(0.0030)	(149.8)	(883.4)	(0.0012)	(0.00071)	(259.0)	(197.2)
distSPUD	-0.0000126**	0.0000475+	-3.714**	6.776	-0.0000126	-0.00000258	-3.661+	-2.153
	(0.0000049)	(0.000025)	(1.23)	(7.35)	(0.0000091)	(0.0000059)	(1.88)	(1.64)
intercept	10.93***	9.482***	35988.5+	-306571.8*	10.94***	10.74***	37964.4	-4378.3
	(0.075)	(0.42)	(19105.5)	(125115.1)	(0.13)	(0.10)	(27058.0)	(27960.6)
N	3678	357	3678	357	933	3102	933	3102
R^2	0.796	0.743	0.737	0.660	0.811	0.756	0.775	0.683
adj. R^2	0.795	0.728	0.736	0.640	0.807	0.754	0.770	0.681
F	711.9	48.56	512.3	32.66	195.6	476.7	157.0	331.8

Robust standard errors in parentheses + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To capture any discrepancies in the effects of hydraulic fracturing on rural versus urban households, two sets of linear and semi log contingency models were run based on either a) whether dummy variable RURAL equaled one or zero, or b) whether dummy variable GREELEY equaled one or zero. Since Weld County is a diverse county that is comprised of a pseudo- urban area including Greeley and bedroom communities to Denver, and rural agricultural land, accounting for the differences in these areas provides further insight into the true effects of drilling on different parts of the county. Greeley and its surrounds are growing rapidly; most of the single-family residential housing transactions between 2009 and 2012 occurred in or around Greeley. The two dummy variables, GREELEY and RURAL, are not mutually exclusive (i.e. if a sale was not in Greeley, it was not necessarily considered rural and vice versa). Full results from these regressions are reported in Table 5.8, where models reported with urban (RURAL = 0) in the headline are the counterpart to rural (RURAL = 1) and non-Greeley (GREELEY = 0) are the counterpart to Greeley (GREELEY = 1).

The results of these regressions show that fracking does appear to affect rural and non-Greeley residents in Weld County differently than those residing in Greeley and other urban and suburban areas. 357 houses sold were not incorporated in any township and considered rural, while 3678 houses sold were considered urban or suburban. There were 933 housing sales in the city of Greeley and 3102 outside of Greeley (non-Greeley) over the 4-year time span. In addition to very different overall model explanatory power, both statistical significance and coefficient magnitude varied across the rural and urban, and Greeley and non-Greeley transactions.

The biggest disparities in results were apparent between the rural and urban models, especially in the treatment variables. Allemp was statistically significant at the <0.001 level in the non-rural model and statistically insignificant in the rural model, under both the linear and semi-log specifications; however, allemp was highly statistically significant in the Greeley and non-Greeley models. Parameter estimates for other treatment variables obtained in the urban linear and semi-log specifications matched up closely to the original parameter estimates from the full regression. This is not true of the parameter estimates obtained from the rural models in which the sign on each of the drill variables switched from the full model. Of these variables, only distSPUD was estimated to be statistically different from zero, but finally with the hypothesized positive sign – indicating that for every meter farther away from a house the drilling occurs, the value increases by $ 12.21 – much larger in magnitude than in all other specifications. Fracking operations located 1000 meters (1 km) away from a house increase the house's sale price by approximately 6%. The adjusted R-squared values from the urban linear and semi-log models were 0.736 and 0.795; from the rural models they were lower at 0.640 and 0.728. Implicit prices, evaluated at the relative real housing sale price mean for that subsection of the data, for all statistically significant oil and gas sector activity variables are reported in Table 5.9.

Table 5.9 Treatment variable implicit prices for statistically significant parameters[15]

	Rural Linear	Rural Semi Log	Urban Linear	Urban Semi Log
allemp	N.S.	N.S.	$545	$ 571
spudcount	N.S.	N.S.	- $1805	- $ 1,354
num_producing	N.S.	N.S.	N.S.	$ 289
distSPUD	N.S.	$ 12.21	- $3.71	- $2.62

Fewer differences existed between the Greeley and non-Greeley models. Parameter estimates for non-treatment variables were similar in both magnitude and statistical significance, with the exception of outbuildingsf. Outbuildingsf was positive for non-Greeley housing sales

[15] N.S. is reported if the coefficient estimate was not statistically different from zero; -- is reported for variables not present in the specification.

and negative for those within Greeley city limits, likely a result of few housing transactions within Greeley having an outbuilding on the property. In general, the drilling treatment variables in these models lacked explanatory power. Spudcount was statistically significant at the 90% confidence level for the Greeley, semi-log specification, and distSPUD was statistically significant at the 90% confidence level for the Greeley, linear specification. Allemp had a positive sign, was statistically significant at the 99.9% confidence level in all Greeley and non-Greeley specifications, and of similar magnitude. The impact of an additional well being drilled within a half mile of a house in Greeley, up to 60 days prior to the sale data, is a decrease of $2,489 per well. Implicit prices, evaluated at the relative real housing sale price mean for that subsection of the data, for all statistically significant treatment variables from the Greeley and non-Greeley specifications are reported in Table 5.10.

Table 5.10 Treatment variable implicit prices for statistically significant parameters[16]

	Non-Greeley Linear	Non-Greeley Semi Log	Greeley Linear	Greeley Semi Log
allemp	$ 493	$ 563	$378	$ 435
spudcount	N.S.	N.S.	N.S.	- $2,489
num_producing	N.S.	N.S.	N.S.	N.S.
distSPUD	N.S.	N.S.	- $3.66	N.S.

Finally, the results of the spatial model are compared to an OLS model that contains interactions between RURAL and GREELEY with each explanatory variable. Those that are not considered rural or are not in Greeley are the baseline category. The OLS regression including these interactions was run as a stepwise OLS regression that iteratively removes and reintroduces combinations of explanatory variables minimize the AIC statistic. If a variable is not included in the stepwise estimation of a model, it can be interpreted as the variable lacking statistical significance in the explanation of the dependent variable. The results of the OLS regression

[16] N.S. is reported if the coefficient estimate was not statistically different from zero; -- is reported for variables not present in the specification.

show that neither spudcount nor num_producing are statistically significant, and they are dropped from the model. Otherwise, results from the stepwise OLS regression align well with other semi-log specifications that were run. Results from the OLS and GLS regressions are reported in Table 5.11. The step-wise GLS regression provides further evidence that the number of wells being drilled within a half mile of a house at up to 60 negatively affects the sale price of a house, as spudcount is statistically significant and has a negative sign. By accounting for the negative spatial autocorrelation using GLS, spudcount becomes significant in the regression. Spudcount was negatively spatially autocorrelated, which causes variances to be higher and things that would have been statistically significant show up statistically insignificant in a t-test. The implicit price for spudcount, evaluated at the mean sale price, in the other category is -$212.55, however it is much higher for the Greeley category at - $212.55/- $4,166. Accounting for the negative spatial autocorrelation in the data shows that there is larger effect on housing values in Greeley due increases in the number of wells being drilled within a half mile of the house.

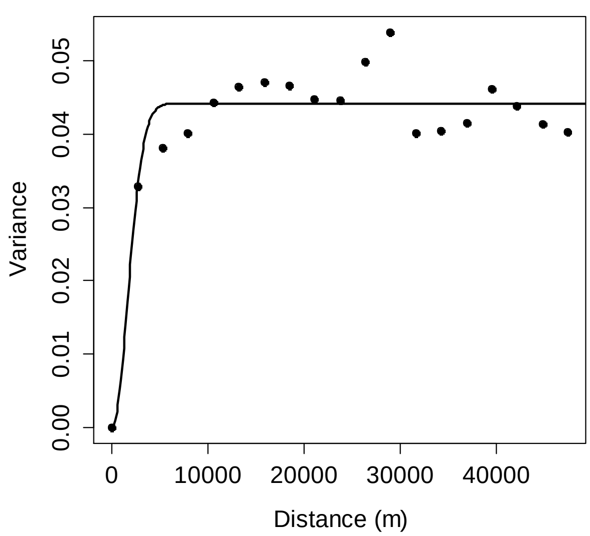

Figure 5.7 Gaussian variogram of OLS residuals created in R

Table 5.11 Results from OLS and GLS spatial model regressions

	OLS	GLS
Intercept	10.8900***	11.0100***
	(0.08)	(0.08)
lotsize	0.2174***	0.2841***
	(0.02)	(0.02)
GREELEY	-0.0835***	-0.1699***
	(0.01)	(0.04)
lotsize2	-0.0195***	-0.0385***
	(0.00)	(0.00)
baths	0.0033	-0.0041
	(0.01)	(0.01)
age	-0.0060***	-0.0058***
	(0.00)	(0.00)
ressf	0.0005***	0.0005***
	(0.00)	(0.00)
ressf2	-0.00000004***	-0.00000003***
	(0.00)	(0.00)
outbuildingsf	0.0001***	0.000045***
	(0.00)	(0.00)
porchsf	0.0001***	0.0001***
	(0.00)	(0.00)
remodel	-0.0009	-0.0453**
	(0.01)	(0.02)
garage	0.1482***	0.1187***
	(0.02)	(0.02)
finish_bsmnt	0.1227***	0.1236***
	(0.01)	(0.01)
pct_hisp	-0.6257***	-0.7266
	(0.06)	(0.04)
pct_own	-0.1619**	-0.1883***
	(0.05)	(0.05)
hh_inc	0.000001***	0.000001+
	(0.00)	(0.00)
pct_bachlr	0.3714***	0.4865***
	(0.06)	(0.05)
allemp	0.0026***	0.0025***
	(0.00)	(0.00)
spudcount		-0.0010
		(0.00)
RURAL	0.0228	-1.5770***
	(0.02)	(0.22)
GREELEY:lotsize2	0.0122**	0.2740*
	(0.00)	(0.12)
GREELEY:baths	0.0234	-0.1233.
	(0.02)	(0.07)
GREELEY:ressf	-0.0002**	0.0337**
	(0.00)	(0.01)

GREELEY:ressf2	0.0000001**	
	(0.00)	
GREELEY:outbuildingsf	0.0001*	-0.00000001*
	(0.00)	(0.00)
GREELEY:remodel		0.09***
		(0.02)
GREELEY:porchsf	0.0001	
	(0.00)	
GREELEY:hh_inc		0.000001*
		(0.00)
GREELEY:pct_bachlr		-0.43***
		(0.09)
GREELEY:allemp		0.0007+
		(0.00)
GREELEY:spudcount		-0.0196+
		(0.01)
lotsize:RURAL	-0.1843***	-0.2400***
	(0.02)	(0.02)
lotsize2:RURAL	0.0189***	0.0375***
	(0.00)	(0.00)
baths:RURAL	0.0819***	0.0793***
	(0.02)	(0.02)
age:RURAL	0.0041***	0.0041***
	(0.00)	(0.00)
ressf:RURAL	0.0003***	0.0005***
	(0.00)	(0.00)
ressf2:RURAL	-0.0000001***	0.0000***
	(0.00)	(0.00)
remodel:RURAL	0.0608+	0.1170***
	(0.03)	0.03
garage:RURAL	-0.1047*	-0.0980*
	(0.04)	(0.04)
finish_bsmnt:RURAL	0.0820*	0.0959**
	(0.03)	(0.03)
pct_hisp:RURAL	0.4490**	0.7863***
	(0.16)	(0.14)
pct_own:RURAL	-0.3358+	0.5699*
	(0.18)	(0.23)
pct_bachlr:RURAL	0.6331**	0.000004**
	(0.24)	(0.00)
allemp:RURAL	-0.0030***	
	(0.00)	
hh_inc:RURAL		0.000003578**
		(0.00)
R^2	0.795	0.777
N	4035	4035
AIC	-1075.4	-1190.5

Robust standard errors in parentheses + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001

CHAPTER 6: CONCLUSION

Although the variables of interest, the well-activity variables, are consistently of low significance across the models, they still add to the R-squared and thus explanatory power of the regressors as a group. These variables are essentially telling the story that there is not a relationship between housing transactions prices and the proximity to the nearest well being drilled up to 60 days prior to the sale. One of the issues is that the data and regression analysis can only provide correlations between variables, not causation, which is often the concern expressed in the media concerning fracking. Another issue may be that wells very close to houses may be receiving royalties for the drilling if they own the mineral rights, something on which there were no data available for this study.

6.1. Findings of the Study

In this study, the results show a low level of impact on housing values due to fracking related activities. Including the three well-related variables on which there was adequate data, spudcount, num_producing, and distSPUD, did not have any impact on the R-squared in any of the semi-log and double-log treatment models, regardless of whether or not oil and gas employment was included. The distance to the nearest well being drilled within 60 days and 2 miles of a property at the time of the sale consistently lacked statistical significance in the different models estimated. Spudcount had the hypothesized, negative sign - indicating that an increase in the number of wells being drilled within 60 days and a half mile of a housing transaction has a negative effect on the sale price of the house - but while it was significant in the specifications with a logged dependent variable, it remained near the 10% significance level across the board. The number of wells in production within a half mile of a house at the time of the sale tested statistically insignificant in all specifications of the model. Residential square footage and the house's lot size were two of the most powerful explanatory variables in the housing price regression analysis.

Alternate specifications were tested that included a variable that captured oil and gas sector employment. The year dummy variables were not included in these specifications because of the high level of cross-correlation between these variables and employment. While the addition of a variable that captured employment did not improve overall regression results, indicative of nearly identical R-squared values across the board on the original and alternate specifications. The largest difference between the six original specifications, with year dummy variables, and the six alternate specifications, with oil and gas sector employment data to replace the year dummy variables, was that the sign on the employment variable was positive and significant, while the year dummy variables were not all statistically significant. The alternate specifications including employment data are preferred to the original specifications, due to the statistical significance of the employment variable.

To analyze overall model explanatory power, R-squared values obtained from these regressions are compared to those from similar studies. These model specifications – both those with year dummy variables and those with oil and gas sector employment – tend to perform quite well comparatively. Klaiber and Gopalkrishnan (2012) reported R-squared values from their regression analyses, with the log of sale price as the dependent variable, of around .69, while the R-squared values for the semi-log specifications in this study were consistently around .77 to .78. Boxall et al. (2005) reported a similar R-squared of .67 for the linear regression run in their study.

This study finds that hydraulic fracturing has different impacts on rural housing values than urban housing values in Weld County. Breaking the data up based on whether the house sold was a in a rural location or located in Greeley had statistical implications that running a full regression including all areas of the county did not. For rural housing values, the volume of drill sites within a half mile radius of the house did not have a statistically significant effect on housing values, while the an increase of a meter to nearest well being drilled increased values by about $12 per meter. However, this is relatively small number considering that the mean sale price of rural Weld County houses between 2009 and 2012 was $257,085, indicating a low economic impact of fracking on housing values. Single-family residential properties sold within the Greeley city limits were, while statistically unaffected by the distance to the nearest well being drilled, negatively impacted by an increase in the density of drill sites around the home within a half mile by $2,489 per drill site. Considering that the mean housing sale price in Greeley was $ 170,499, this number is not trivial. Rural property owners are affected by distance to drill sites, on the other hand, urban (Greeley) residential properties are impacted by the volume of drill sites near the home. These discrepancies between rural and urban have policy implications, specifically implications that policies are needed to target each group accordingly. To protect home owners in Greeley, policies may be needed to regulate the maximum number of drill sites within a certain distance from another drill site. Minimum distances from residential properties may need to be set and/or increased in rural areas. Horizontal drilling techniques allow the number of well pads to be kept down while increasing the efficiency of extraction. The use of more horizontal drilling in higher population density areas may help solve this issue.

Moran's I was used to determine whether spatial dependencies were present in the data. The residuals from the OLS model tested positive for spatial autocorrelation, violating the assumption that observations are independent. In the presence of spatially autocorrelated residuals, coefficient estimates often remain unbiased despite OLS being inefficient. Inefficiency is the result of incorrect variances, which can cause OLS estimators' standard errors to be underestimated. Incorrect standard errors and variances can lead to type I or type II errors on t-tests for variable significance. Results from the GLS model accounting for spatial autocorrelation were consistent with those from the non-stepwise OLS models, both in terms of parameter estimates, significance, and overall model fit in terms of R^2. The stepwise GLS model, where all variables are weighted by the covariance of the residuals, includes the fracking-related variable spudcount, while the stepwise OLS does not due to the inflated variance of the negative spatial autocorrelation (Anselin, 1988).

It is important for the reader of this study to understand that these results represent a specific geographic area. While results show a slight negative impact of fracking on property values in Weld County – specifically in Greeley – these values relate only to housing sales in Weld County and may not apply to other areas of the state and/or country. Property values and political views vary throughout the state of Colorado, and living in a specific county may already have implications of buyers' tastes and preferences.

6.2. Limitations of the Study

There are two main limitations of this study; one that relates to data availability and one that relates to modeling techniques employed. While other studies analyzing the effects of fracking on housing values looked at the effects depending on the water source serving the house (Muehlenbachs et al., 2012), these data were not available for Weld County. Since water issues are some of the most prevalent issues associated with fracking in Colorado, the absence of data on water supply may be part of the reason for the low-level of significance found in drilling variables. Given that a lot of people are concerned about the public health effects of fracking (i.e. from air pollution and water pollution), the lack of data available on these factors may be downplaying some of the perceived risks associated with fracking that were hypothesized would be capitalized into housing values.

On a similar note, the lack of data on whether the owner of the house also owns the mineral rights, thus receiving royalties from the oil and gas extraction, may be part of the cause of the positive sign on num_producing and negative sign on distSPUD, the opposite of the hypothesized sign, respectively. If the owner is receiving royalties from the extraction of oil and gas, then it would be expected that the drilling would increase the value of the home should the mineral rights transfer over to the new owner. The lack of data on this may be muddling results.

Due to the limited number of observations that had a permit for drilling issued up to 60 days prior to the sale and within 2 miles of a house sold, data on permitted wells were omitted from the regressions. At the same time, it is hard to say whether there is an issue of asymmetric information with permitting data, and if people know about permitted wells and their proximity to the house.

6.3. Suggestions for Further Research

Many of the suggestions for future extensions of this research stem from the limitations of the study as specified in section 6.2. If the data on what type of water is serving a household were available, including interaction terms between it and well-drilling variables in a regression might improve the results. Similarly, if data on the mineral rights owners and whether these rights will transfer to the new owner were readily available, the use of this information may capture better results in analysis of the effects of fracking on housing values.

Collecting data that goes back to before the start of the fracking boom, around 2005, would increase the already large sample size and potentially capture more effects of the drilling influx. By increasing the time span or the search radius for fracking around the house from 2 miles to a higher distance of 5 miles, the analysis might be improved upon. Instead of using continuous distance, one could use specific distance bands to test sensitivity to certain distances and potentially identify a distance at which drilling no longer matters to a homeowner. Additionally, testing to see if the number of wells in a specific neighborhood affects the prices of the houses in that neighborhood is another thing that could be done to look at the effect of drilling on residents.

LITERATURE CITED

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Boxall, P. C., Chan, W. H., & McMillan, M. L. (2005). The impact of oil and natural gas facilitites on rural residential property values: a spatial hedonic analysis. *Resource and Energy Economics*, *27*, 248-269.

Brown, S. P.A., & Krupnick, A. J. (2010). Abundant Shale Gas Resources: Long-Term Implications For U.S. Natural Gas Markets. RFF DP 10-41. Resources For the Future, Washington DC.

City of Arlington, TX :: Government :: Community Development and Planning :: Gas Drilling and Production :: FAQs. (n.d.). City of Arlington, TX. Retrieved May 24, 2013, from http://www.arlingtontx.gov/planning/gas_drilling_FAQs.html#9

Colorado Department of Local Affairs - GIS Data. (n.d.). *Colorado.gov: The Official State Web Portal*. Retrieved April 9, 2013, from http://www.colorado.gov/cs/Satellite/DOLA-Main/CBON/1251595720248

Cooley, H. & Donnelly, K. (2012, June). *Hydraulic Fracturing and Water Resources: Separating the Frack from Fiction*. Oakland, CA: The Pacific Institute. Retrieved November 21, 2012, from http://www.pacinst.org/reports/fracking/index.htm

Cropper, M. L., Deck, L. B., & McConnell, K. E. (1988). On the Choice Functional Form for Hedonic Price Functions. *The Review of Economics and Statistics*, *70*(4), 668-675.

Davis, C. (2012). The Politics of "Fracking": Regulating Natural Gas Drilling Practices in Colorado and Texas. *Review of Policy Research*, *2*, 177-191.

Davis, L.W. (2011). The effect of power plants on local housing values and rents. *Review of Economics and Statistics*, 93(4), 1391-1402.

Hite, D., Chern, W., Hitzhusen, F., & Randall, A. (2001). Property-Value Impacts of an Environmental Disamenity: The Case of Landfills. *Journal of Real Estate Finance and Economics*, 22:2/3, 185-202.

Hydraulic Fracturing: The Process | FracFocus Chemical Disclosure Registry. (n.d.). *Home | FracFocus Chemical Disclosure Registry*. Retrieved October 13, 2012, from http://fracfocus.org/hydraulic-fracturing-how-it-works/hydraulic-fracturing-process

Klaiber, H. A., & Gopalakrishnan, S. (August, 2012)**.** *The Impact of Shale Exploration on Housing Values in Pennsylvania***.** Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2012 AAEA Annual Meeting, Seattle, Washington.

Jaffe, M. (2011, January 17). Colorado suburban homeowners face invasion of oil and gas wells. *The Denver Post* . Retrieved October 26, 2012, from http://www.denverpost.com/commented/ci_18493742?source=commented-business

Kerr, T. (2013, February 20). Personal email communication.

Koch, W. (2012, December 5). U.S. forecasts rising energy independence. *USA Today*. Retrieved February 1, 2013, from http://www.usatoday.com/story/news/nation/2012/12/05/usa-energy-independence-renewable/1749073/

Krupnick, A et al. (n.d.). Shale Maps. *Resources for the Future - RFF.org* . Retrieved December 15, 2012, from http://www.rff.org/centers/energy_economics_and_policy/Pages/Shale_Maps.aspx

Lewis, L. Y., & Acharya, G. (2006). Environmental Quality and Housing Markets: Does Lot Size Matter?. *Marine Resource Economics*, *21*, 317-330.

Loomis, J. (2004). Do nearby forest fires cause a reduction in residential property values? *Journal of Forest Economics*, 10, 149-157.

Magill, B. (2012, November 8). Could fracking use less water? CSU wants to find out. *the Coloradoan*. Retrieved January 14, 2013, from http://www.coloradoan.com/article/20121108/NEWS01/311080036/Could-fracking-use-less-water-CSU-wants-find-out?gcheck=1

Marcus, P. (2012, December 14). Controversy over fracking in Colorado runs deep. *The Colorado Statesman*. Retrieved January 5, 2013, from http://www.coloradostatesman.com/content/993907-controversy-over-fracking-colorado-runs-deep

Matthews, V. (2011, Spring). Rock talk: Colorado's new oil boom - the Niobrara. Colorado. Geological Survey Newsletter, 13(1).

McKenzie, L.M., et al. (2012). Human Health Risk Assessment of Air Emissions From Development of Unconventional Natural Gas Resources. *The Science of the Total Environment*, May 1;424:79-87. doi: 10.1016/j.scitotenv.2012.02.018. Epub 2012 Mar 22.

Michaels, R.G., & Smith, V.K. (1990). Market Segmentation and Valuing Amenities Hedonic Models: The Case of Hazardous Waste Sites. *Journal of Urban Economics*, 28, 223-242.

Muehlenbachs, L., Spiller, E., & Timmins, C. (2012). Shale Gas Development and Property Values: Differences across Drinking Water Sources. RFF DP 12-40. Resources For the Future, Washington DC.

Mueller, J.M., & Loomis, J.B. (2008) Spatial Dependence in Hedonic Property Models:Do Different Corrections for Spatial Dependence Result in Economically Significant Differences in Estimated Implicit Prices? *Journal of Agricultural and Resource Economics*, 33(2):212-231

"NaturalGas.org." *NaturalGas.org*. N.p., n.d. Web. 4 Sept. 2012. <http://www.naturalgas.org/overview/unconvent_ng_resource.asp>.

Palmquist, R. B., Roka, F. M., & Vukina, T. (1997). Hog operations, environmental effects, and residential property values. *Land Economics*, 73, 114-124.

Platt, J. (2013, May 9). Oil and fracking booms creating housing busts | MNN - Mother Nature Network. *Environmental News and Information | MNN - Mother Nature Network*. Retrieved May 16, 2013, from http://www.mnn.com/earth-matters/energy/stories/oil-and-fracking-booms-creating-housing-busts

Rogers, H. (2011). Shale gas--the unfolding story. *Oxford Review of Economic Policy*, 27(1), 117 - 143.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *The Journal of Political Economy*, 82(1), 34-55.

Spud Definition | Investopedia. (n.d.). *Investopedia - Educating the world about finance*. Retrieved December 12, 2012, from http://www.investopedia.com/terms/s/spud.asp

Taylor, L.O. (2003) The hedonic method. In P.A. Champ, Boyle K.J., & Brown T.C. (Eds.), *A Primer on Nonmarket Valuation* (pp. 331-393). Norwell, MA: Kluwer Academic Publishers.

U.S. Bureau of Labor Statistics. (n.d.). *Consumer Price Index - All Urban Consumers*. Retrieved February 24, 2013, from http://data.bls.gov/cgi-bin/dsrv

U.S. Bureau of Labor Statistics. (n.d.). *Consumer Price Index - All Urban Consumers*. Retrieved April 11, 2013, from http://data.bls.gov/cgi-bin/dsrv

US Environmental Protection Agency. (2011, November). Plan to Study the Potential Impacts of Hydraulic Fracturing on Drinking Water Resources (EPA/600/R-11/122). Washington DC: Office of Research and Development.

US Energy Information Administration. (2012, December 5). *Annual Energy Outlook 2013 Early Release* (DOE/EIA-0383ER(2013)). Washington, DC: U.S. Government Printing Office.

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

APPENDIX

Table 1.A Eligible sample summary statistics for structural and property variables

Variable	N	Mean	Min	Max	Std. Dev.	Variance
ln_sales	14222	12.13731	10.31475	15.05109	.4978698	.2478743
salep_real	14222	211644.4	30174.49	3440356	122772.1	1.51e+10
lotsize	14222	.484093	0	72.56	1.691441	2.860973
age	14222	16.5038	0	147	19.65888	386.4716
ressf	14222	1694.536	520	25337	669.3111	447977.4
ressf2	14222	3319397	270400	6.42e+08	6040178	3.65e+13
bedrooms	14222	3.403741	0	8	.9566852	.9152465
baths	14222	2.601884	0	23	.9190523	.8446572
bldgs	14222	1.08297	1	16	.4676509	.2186974
drywall	14222	.9874842	0	1	.1111757	.01236
finish_bsmnt	14222	.3848966	0	1	.486588	.2367679
outbld	14222	.0496414	0	1	.2172106	.0471804
garage	14222	.9449445	0	1	.2280967	.0520281
remodel	14222	.1158065	0	1	.320004	.1024026
y2010	14222	.2344255	0	1	.4236541	.1794828
y2011	14222	.2408241	0	1	.4275988	.1828407
y2012	14222	.2951062	0	1	.4561065	.2080331
RURAL	14222	.0912671	0	1	.2879986	.0829432
GREELEY	14222	.2779497	0	1	.4480042	.2007078
hwy_mile	14222	.4156237	0	1	.4928465	.2428977
hwy_100yd	14222	.0290395	0	1	.1679232	.0281982

Table 2.A Summary statistics for neighborhood demographic variables

Variable	N	Mean	Min	Max	Std. Dev.	Variance
median_hhinc	14222	66679.5	13906	184821	19489.8	3.80E+08
pct_white	14222	73.4707	11.99	91.37	15.4751	239.478
pct_hisp	14222	22.5485	6.58	86.35	15.5801	242.738
pct_male	14222	50.0007	45.12	56.86	1.88523	3.55409
pct_65plus	14222	9.38073	1.12	31.75	4.55592	20.7564
pct_vac	14222	6.4375	1.44	23.48	2.93256	8.59989
pct_own	14222	76.6978	5.88	95.63	13.4395	180.619
pc_nonUScit	14222	0.04693	0	0.2505	0.05013	0.00251
pct_hsgrad	14222	0.59771	0.36292	0.81	0.07737	0.00599
pct_bachlr	14222	0.28597	0.01906	0.6	0.12398	0.01537

Table 3.A Summary statistics for relevant oil and gas variables

Variable	N	Mean	Min	Max	Std. Dev.	Variance
wellwater	14222	.0217972	0	1	.146026	.0213236
horzwell	14222	.0287583	0	1	.1671322	.0279332
distSPUD	4250	2029.834	56.14307	3218.615	779.1168	607022.9
ID_spud	4250	.0006459	.0003107	.0178116	.0005806	3.37e-07
spud2	4250	4727108	3152.044	1.04e+07	3020415	9.12e+12
DS_halfmile	319	557.1741	56.14307	801.9086	165.864	27510.87
DS_1mile	1334	1074.349	56.14307	1609.021	363.2971	131984.8
spudcount	14222	.0709464	0	11	.5689512	.3237054
distPROD	9290	362.1434	1.340132	804.5902	188.512	35536.79
ID_prod	9290	.0041021	.0012429	.7461951	.0087901	.0000773
prod2	9290	166680.8	1.795954	647365.4	158748.1	2.52e+10
num_producing	14222	4.854591	0	41	5.77892	33.39591
distPERM	1739	2050.613	191.1916	3218.151	735.4974	540956.4
ID_perm	1739	.000601	.0003107	.0052304	.0003928	1.54e-07
perm2	1739	4745658	36554.22	1.04e+07	2901289	8.42e+12
permitcount	14222	.0240473	0	9	.3370742	.113619

Table 4.A Summary statistics for all variables if there was a spud within 2 miles of a house up to 60 days prior to the sale

Variable	Obs	Mean	Std. Dev.	Min	Max
salep_real	4035	212559.7	106775.9	30174.49	1542392
ln_sales	4035	12.15981	0.463643	10.31475	14.24885
ln_lotsize	4029	-1.49592	0.833112	-3.91202	3.688879
lotsize	4035	0.486798	1.769378	0	40
lotsize2	4035	3.366896	51.70964	0	1600
baths	4035	2.628253	0.925106	0	23
age	4035	15.7482	19.48985	0	130
ln_ressf	4035	7.383626	0.347309	6.327937	8.646114
ressf	4035	1710.387	619.2882	560	5688
ressf2	4035	3308848	2641167	313600	3.24E+07
outbuildingsf	4035	89.00397	627.9527	0	22092
porchsf	4035	258.5187	247.2978	0	4824
remodel	4035	0.119207	0.324072	0	1
garage	4035	0.957125	0.2026	0	1
finish_bsmnt	4035	0.372986	0.483659	0	1
GREELEY	4035	0.231227	0.421669	0	1
RURAL	4035	0.088476	0.284021	0	1
y2010	4035	0.237423	0.425556	0	1
y2011	4035	0.247088	0.431372	0	1
y2012	4035	0.305081	0.460499	0	1
hwy_mile	4035	0.405205	0.490992	0	1
pct_hisp	4035	0.219681	0.144144	0.0658	0.8635
pct_own	4035	0.773565	0.126415	0.0588	0.9563
hh_inc	4035	79233.92	21476.25	23052	157490
pct_bachlr	4035	0.284447	0.119516	0.019064	0.6
allemp |	4035	170.6813	12.07883	156.1	191.7
spudcount	4035	0.242627	1.040512	0	11
distSPUD	4035	2029.046	777.3933	56.14307	3218.615
spud2	4035	4721216	3015524	3152.044	1.04E+07
ln_spud	4035	7.513003	0.503907	4.027903	8.076706
spud_halfDV	4035	0.075093	0.263574	0	1
spud_1mileDV	4035	0.314003	0.464175	0	1
num_producing	4035	6.60347	6.04416	0	39
distPROD	3256	348.563	183.0098	10.66153	804.5902
prod2	3256	154978.5	151611.2	113.6683	647365.4
permitcount	4035	0.041884	0.440265	0	9

Table 5.A Summary statistics for all variables if there was *no* spud within 2 miles of a house up to 60 days prior to the sale

Variable	Obs	Mean	Std. Dev.	Min	Max
salep_real	9496	216365.3	125927.1	30476.24	2413959
ln_sales	9496	12.15725	0.5025	10.3247	14.69678
ln_lotsize	9487	-1.47963	0.812097	-3.91202	3.568969
lotsize	9496	0.44859	1.269037	0	35.48
lotsize2	9496	1.811519	23.05669	0	1258.83
baths	9496	2.610678	0.924175	0	20
age	9496	16.90543	20.21376	0	147
ln_ressf	9496	7.374708	0.366603	6.253829	8.95854
ressf	9496	1707.344	657.766	520	7774
ressf2	9496	3347635	2913002	270400	6.04E+07
outbuildingsf	9496	87.85689	562.3472	0	18672
porchsf	9496	266.9805	259.8123	0	3779
remodel	9496	0.120156	0.325161	0	1
garage	9496	0.95198	0.21382	0	1
finish_bsmnt	9496	0.398273	0.489568	0	1
GREELEY	9496	0.297915	0.457366	0	1
RURAL	9496	0.088142	0.283517	0	1
y2010	9496	0.232519	0.42246	0	1
y2011	9496	0.236626	0.425033	0	1
y2012	9496	0.292334	0.454859	0	1
hwy_mile	9496	0.419019	0.493424	0	1
pct_hisp	9496	0.226866	0.161183	0.0658	0.8635
pct_own	9496	0.765669	0.137733	0.0588	0.9563
hh_inc	9496	78815.77	22114.16	23052	157490
pct_bachlr	9496	0.288355	0.123995	0.019064	0.6
allemp	9496	170.4742	12.41121	156.1	191.7
spudcount	9496	0	0	0	0
distSPUD	0				
spud2	0				
ln_spud	0				
spud_halfDV	9496	0	0	0	0
spud_1mileDV	9496	0	0	0	0
num_producing	9496	4.114364	5.48754	0	41
distPROD	5546	369.0089	190.1841	6.940779	804.4459
prod2	5546	172331	161424.1	48.17442	647133.1
permitcount	9496	0.017481	0.293352	0	9

Table 6.A Regression results for treatment models that included an interaction between a domestic-use water well (wellwater) and the distance to the nearest fracking site (distSPUD) from the property

	Linear	*Semilog*	*Double log*
	salep_real	**ln_sales**	**ln_sales**
lotsize	15669.2***	0.0558***	
	(-2823.1)	(-0.0084)	
lotsize2	-352.7***	-0.00134***	
	(-73.5)	(-0.00026)	
baths	4044.4+	0.00923	0.0115
	(-2379.6)	(-0.0067)	(-0.0071)
age	-773.5***	-0.00509***	-0.00505***
	(-92.4)	(-0.00038)	(-0.00038)
ressf	18.32	0.000573***	
	(-23.2)	(-0.000043)	
ressf2	0.0175**	-5.00e-08***	
	(-0.0057)	(-9.80E-09)	
outbuildingsf	8.249**	0.0000549***	0.0000357***
	(-2.94)	(-0.000011)	(-0.0000088)
porchsf	45.32***	0.000149***	
	(-9.82)	(-0.00002)	
remodel	-2582	0.0167	0.0129
	(-2792.5)	(-0.014)	(-0.014)
garage	12276.1*	0.118***	0.108***
	(-4915.8)	(-0.027)	(-0.03)
finish_bsmnt	30633.5***	0.142***	0.129***
	(-2603.2)	(-0.0092)	(-0.0093)
GREELEY	-26311.0***	-0.102***	-0.105***
	(-2542.4)	(-0.0096)	(-0.0096)
RURAL	7625.8	0.00875	-0.0479*
	(-6456.8)	(-0.022)	(-0.02)
hwy_mile	4071.5+	0.00982	0.0153*
	(-2105.7)	(-0.0077)	(-0.0076)
pct_hisp	-62426.0***	-0.604***	-0.558***
	(-10847.3)	(-0.046)	(-0.046)
pct_own	-50062.6***	-0.140**	-0.118*
	(-11705.9)	(-0.047)	(-0.046)
hh_inc	0.467***	0.00000130***	0.00000102***
	(-0.072)	(-0.00000026)	(-0.00000025)
pct_bachlr	104001.3***	0.405***	0.459***
	(-12318.1)	(-0.044)	(-0.044)
allemp	489.9***	0.00248***	0.00229***
	(-76.9)	(-0.00029)	(-0.00029)
spudcount30	-1992.0+	-0.00444	-0.00357
	(-1064.6)	(-0.005)	(-0.0046)
num_producing	-128.6	0.000215	0.000946
	(-172.9)	(-0.00062)	(-0.00061)
waterspud	-6.355	-0.00000157	-0.0000309+
	(-4.37)	(-0.000018)	(-0.000017)
ln_lotsize			0.120***
			(-0.0084)
ln_ressf			0.620***
			(-0.02)
ln_porchsf			0.0265***
			(-0.0044)
Intercept	-9506.7	10.79***	7.150***
	(-29706.7)	(-0.088)	(-0.16)

N	4035	4035	3940
R-sq	0.704	0.779	0.789
adj. R-sq	0.702	0.778	0.788
F	361.9	564.7	606.7

Robust standard errors in parentheses "+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001"