Agricultural Effects of Unconventional Oil and Gas Development: A Machine Learning Approach

WORKING PAPER

Adam Patterson[1] and Sandro Steinbach[2,3]

Manuscript Prepared for Publication in Z

[1] Department of Agricultural and Resource Economics, University of Connecticut, 1376 Storrs Road, Storrs, CT 06269-4021

[2] Department of Agricultural and Resource Economics, University of Connecticut, 1376 Storrs Road, Storrs, CT 06269-4021

[3] All correspondence should be addressed to Sandro Steinbach (Email: sandro.steinbach@uconn.edu).

Abstract

Technological advances within the energy industry, specifically in unconventional oil and natural gas development (UOGD), during the 21st century has created a short term economic incentive to produce as horizontally drilled wells have increased by over 1,400% since 2000[4]. Rapid expansion has occurred despite health and environmental impacts being not well understood. The relatively under-examined externalities of UOGD have caused some regions to ban the method of production. Potential adverse impacts to agriculture are particularly understudied given that previous research has documented negative effects of radioactive fallout on crop yield. This study quantifies the effect of hydraulic fracturing on U.S. agricultural productivity from 1950-2020 while accounting for the endogenous relationship with conventional oil and gas development (COGD). We utilize oil and gas data from all 3,946,198 wells included in the Enverus drillinginfo database to estimate the association between agricultural productivity and downwind radiation resulting from UOGD. Identifying variation, the number of upwind wells within 20 km of agricultural county centroids, exploits spatiotemporal differences while mapping annual wind direction with spud year to account for a distribution mechanism. Results suggest an average effect of ** per UOGD well within 20km of county centroids. Our back of the envelope calculations found UOGD to influence an estimated crop loss of $2,910,428 for corn and soybeans nationwide.

[4] (United States Energy Information Administration, 2020)

Introduction

(LongReport*referencing that oil and gas companies have zero regard for communal requests through refusing move pads by as little as X feet*) neighboring residents' health concerns are of secondary importance as firms drill strictly to optimize production (U.S. Environmental Protection Agency, 2017). Companies race to produce as technological advances in unconventional oil and gas development (UOGD) within the 21st century has allowed the energy industry to realize cost efficient horizontal drilling and hydraulic fracturing. As a result, many short term economic benefits have been associated with contemporary evolutions of UOGD (Kearney and Wilson, 2018; Feyrer et al., 2017; Bartik et al., 2019; Allcott and Keniston, 2017; Cascio and Narayan, 2020; Mason et al., 2015; Newell and Raimi, 2015; Hausman and Kellogg, 2015). Newfound incentives have increased the number of horizontally drilled wells in the United States from 10,000 in 2000 to 150,000 in 2019 (United States Energy Information Administration, 2020). The rapid expansion of UOGD, commonly known as hydraulic-fracturing or fracking, has drawn scrutiny as limited research exists regarding the potentially adverse effects on human health and the environment (Werner et al., 2015). This lack of knowledge regarding potential societal costs has led some regions to outright ban UOGD (Black et al., 2019; Currie et al., 2017; Boslett et al., 2016).

Asymmetric information is a key driver for the sparse amount of research as fracking companies are excluded from participating in several federal environmental statues, such as the Safe Drinking Water Act through passage of the 2005 Energy Policy Act (Vengosh et al., 2014; Colborn et al., 2011; Howarth and Ingraffea, 2011). This secrecy protects the industry and creates difficulty in providing evidence of adverse environmental relationships or threats to human health. In addition, elected officials have been allowed to appease the energy industry as according to Boomhower (2021), bundled elections reduce the salience of environmental issues such as oil and gas development. Many important macro topics are often merged into one electoral campaign, creating prioritized voting that often leaves adverse environmental conditions ignored unless directly affecting citizens. Potential effects on agriculture are a particular understudied UOGD externality given the critical nature of food production for global sustainability and the large percentage of wells drilled within agricultural producing counties (Black et al., 2020; Hitaj et al., 2014).

Literature by Rasmussen et al. (2016); McKenzie et al. (2012) indicate the potential for naturally occurring radioactive material (NORM), or ionizing particle radiation, to be emitted into the environmental at all four stages of well production: pad preparation, drilling, perforation, and production. The first step is well pad preparation, typically lasting 30 days, consisting of clearing 3-5 acres and delivery of raw materials needed for construction of the well. Truck traffic and negative externalities associated with equipment delivery provide additional cause for adverse impacts to agriculture as robust research exists regarding harmful encounters. According to Rasmussen et al. (2016), the development phase alone requires over 1000 truck trips per well while a previous study determined that heavy vehicle traffic and trucks idling at well pads were the likely sources of intermittently high dust and benzene concentration (McCawley, 2013; West Virginia Department of Environmental Protection, 2013). According to Gibson (2013), increased truck traffic from fracking results in cattle rejecting dust-ladden feed, refusing to lay in dusty hay, and even dying from dust pneumonia. The second phase occurs when the drilling begins, also known as the spud date of the well. The duration of this phase usually lasts up to a month when the well is drilled both vertically and horizontally. A well is classified as UOGD if it is drilled 5,000 to 9,000 feet vertically and up to 10,000 feet horizontally. The third phase, perforation of the horizontally drilled tunnels, is unique to the fracking process. In this stage, stimulation, or hydraulic fracking, requires 2 to 4 million gallons of water (U.S. Department of Energy, 2009) while sand and other chemical additives are pierced into the rock, creating small holes in which the natural gas can be released from entrapment. According to Mason et al. (2015), water is heavily used only in well drilling and not while in the production phase. The fourth phase of well development consists of the beginning of gas production where the final product is extracted from the shale patch. In addition to NORM and negative externalities associated with building the well pad, related studies identify adverse UOGD impacts of air pollution, water contamination, competition for local inputs, soil fertility and water scarcity to be probable driving factors in agricultural output variation (Muehlenbachs et al., 2015; Hill and Ma, 2017; Marchand and Weber, 2017; Kemball-Cook et al., 2010; Gopalakrishnan and Klaiber, 2013; Colborn et al., 2011).

According to Colborn et al. (2011), the drilling phase, completed before production begins, provides tremendous human health risk from radiation as evidenced by the Crosby well blowout. Hence, coupled with information on well development, we expect our most significant models to result from time variation relating to spud year rather than completion or production year[5]. Probable mechanisms for influence on agricultural productivity during this time interval include both truck traffic and release of adverse chemicals into the environment. During this phase, fugitive methane emissions from fracking development consist of 3.6 to 7.9% of total production whereas COGD emits 2 to 3% (Howarth et al., 2011), making adverse environmental exposure from UOGD a material concern. We also note that the true effect of UOGD on productivity must be stripped away from the effects of COGD. Adverse impacts of fugitive gas emission become even more concerning as exposed livestock are especially sensitive to environmental stressors from UOGD (Bamberger and Oswald, 2012), and according to the U.S. Committee on Interstate and Foreign Commerce (1980), approximately 25% of newborn lambs were stillborn as a result of regional radioactive nuclear testing. Bustad et al. (1957) ran scientific experiments in sheep that resulted in corroborating evidence. Modern research, related specifically to UOGD, by Steinzor et al. (2012) indicates that 20 percent of surveyed households report livestock developing unfavorable health symptoms to include seizures, loss of hair, or death after recent UOGD drilling. Fugitive gases act as a direct concern to productivity as methane and other harmful toxins influence soil fertility. Research indicates that states with intense production of corn and soybean may statistically capture less air-borne methane due to retrieval biases in low surface albedo when agricultural soils are bare (de Gouw et al., 2020), implying that adverse chemicals are absorbed by plants. Worth noting is that fugitive methane emissions contribute directly to global warming (Aben et al., 2017), an indirect concern to long term crop yield as robust research exists documenting the negative impacts of atmospheric warming on agricultural productivity (Schauberger et al., 2017; Kimball, 1983).

During the spud phase, fugitive natural gases and other unsafe elements escape the exposed well and mix with nitrogen oxides to produce ground level ozone. After radiation has attached to airborne particles and dispersed through ozone to locations of agricultural production, plants absorb radioactive material while livestock consume contaminated pasture. The resulting ozone, indicated by particle matter when the particles are less than 2.5 μ in diameter, has been demonstrated to be harmful to humans as measured by emergency room admissions during periods of elevation

[5] This is confirmed by mapping average wind direction to completion and production year and running the results. Both models act as a negative control of sorts as no effect is shown at any km distance. Results can be found in Appendix *X

(Peng et al., 2009). While such evidence provides an example of a direct UOGD effect on human health, plant absorption of particle matter has the potential to indirectly effect human health through reduced or contaminated food supplies as prior research by Meyers (2019) detail the adverse effects of exogenous radioactive fallout on agricultural yields. According to Micco et al. (2010), seed growth and reproduction may be obstructed by gamma and beta radiation while agricultural crops can absorb beta particles through both soil and their leaves (Ogundare and Adekoya, 2015). Throughout the century, experiments involving induced mutations of crops using various radiation sources became a common place study (Mba, 2013). According to Stadler (1928), radiation is proven to cause alteration in the genetic make-up of corn and barley while the encompassing collection of literature has found ionizing particles to commonly mutate physical properties of horticulture. Through these mechanisms, we wish to understand if radiation emitted from UOGD alters agricultural yield in a positive or adverse way.

In this study, we quantify the average causal effect of hydraulic drilling on agricultural productivity for two major crops, corn and soybeans, within the contiguous United States. Yield data consists of bushels per acre planted for corn and soy beans. We use yield as Farah (2016) provides evidence of declining agricultural productivity in UOGD areas whereas Thakor (2018) find increases in productivity related to newfound oil leasing payments. Given this contradiction, we are curious as to the true resulting effect of UOGD on agricultural productivity. Using yield per acre planted will capture an unbiased estimate in accounting for farmers that may have shifted production given crop damages from UOGD. Yield as a production variable will also account for potential land expansion amongst producers that experienced new leasing payments. Additionally, yield per acre is robust to farmers that shifted into less water intensive crop production as a result of the massive increase in water demand during perforation. For example, livestock is extremely water intensive which may cause producers to substitute into crops and other alternatives. Nonetheless, producers will temporarily be incentivized to substitute for crops with less necessity of water. Yield per acre is robust to such changes resulting from irrigation as UOGD is documented to emit elevated levels of Radium-226, Uranium-238 and other contaminants in produced water from reservoirs, surface water, local stream sediments, and impounded rivers (Brown, 2014; Torres et al., 2018; Zhang et al., 2015). Thus, yield captures various potential mechanisms in UOGD rather than one specific mechanism of air pollution, radiation, or climate variation.

The main concern related to inferring a causal relationship between agricultural output and UOGD is that UOGD wells overlap spatiotemporally with COGD wells. We are aware that UOGD drilling emits up to 5.9% more of total production than COGD. In an ideal situation, we would have data from separate regions with exclusively UOGD or COGD to determine a consistent effect. Unfortunately, available datasets are heterogeneously observed outcomes. Endogeneity emerges as the effects of NORM resulting from UOGD have historically been related with NORM, and other adverse chemicals, produced from COGD (Hilal et al., 2014; Black et al., 2019; Kolb and Wojcik, 1985). Additionally, truck traffic and other unfavorable effects from well development are homogenous across both UOGD and COGD. In accounting for this confoundedness, we include COGD as a covariate in performing regularization techniques to eliminate confounding (Abadie and Kasy, 2019; Athey, 2018). According to Storm et al. (2019), machine learning provides flexible model selection when treatment is thought to be assigned based on a non-linear combination of observables. We employ the Double Machine Learning (DML) method, proposed by Chernozhukov et al. (2018), to orthogonalize covariates in producing a causal treatment effect in dealing with the endogenous COGD relationship.

Methods

Our primary source of identification in evaluating the impact of agricultural productivity from fracking is the total number of upwind wells within 20 km from each county centroid. Upwind wells, classified by yearly wind direction mapped to spud year, are included to control for atmospheric conditions influencing radiation distribution. We use spud date as our point of reference for wind direction as this is the first, and most intensive, phase in which NORM and other radioactive elements are released from earth's innermore layers. Also, considering that research has found natural gas production to decrease by up to 70% after the first year of production (Howarth and Ingraffea, 2011), we have reason in allowing us to assume homogenous temporal intensity amongst wells after the first year of production. Furthermore, a majority of adverse effects, including fugitive gases, are released pre-production, and this is confirmed in our analysis as substituting spud year wind direction with well completion year wind direction acted as a model negative control of sorts. We found no associate of upwind or downwind wells when mapping yearly completion year direction to each county centroid(SUMMARY STATS OR CREATE ROBUSTNESS). These results act as confirmation in mapping yearly wind direction to each spud year. For obtaining yearly wind direction, we utilize the North American Regional Reanalysis (NARR) dataset to obtain u and v wind in calculating the direction and magnitude of wind. We use $\tan x = \arctan(\frac{v}{u})$ to calculate the radian angle of each wind vector with the interpretation of direction that the wind is blowing from. Azimuth angles are added by 360 and the modulo operator, 360, is performed to transform angles from traditional coordinate bearing interpretations. Given that the resolution grid is 32km (.3 degrees) and our well treatment distance is 20km, we obtain the average yearly direction of the 3 closest resolution grids from each well. To preserve the 360 degree bound of wind direction while calculating an inverse distance weighting (IDW), we create a yearly wind average variable for each county i in time t. First, we obtain a total weight, TW_{it} , for each county centroid per year by summing all IDW values from the nth closest resolution grids :

$$TW_{it} = \sum_{n=1}^{3} IDW_{itn} \tag{1}$$

when

$$IDW_{itn} = \frac{1}{distance_{itn}} \tag{2}$$

Where $distance_n$ denotes the km distance between county centroid and resolution grid latitude and longitude values. We then take the share of each individual IDW weight from the total weight and multiply by the corresponding wind direction observation for each resolution grid.

$$Wind_{itn} = \frac{IDW_{itn}}{TW_{it}} * yearly\ wind\ direction\ observation_{itn} \tag{3}$$

Repeating this step for all 3 grids, and summing the 3 weighted direction observations results in the final numeric wind direction value we employed for analysis.

$$Yearly\ Wind\ Average_{it} = \sum_{n=1}^{3} Wind_{itn} \tag{4}$$

We felt this approach was necessary to preserve the original boundaries of our observation as final direction bearing must be bounded between 0 and 360 and traditional IDW methods scale variables out of the required bounds. We interact wind magnitude with upwind well count as Li et al. (2020) found significant negative interactions between wind velocity and the count of UOGD wells located upwind of RadNet monitors while suggesting lower influence when the wind is strong. Magnitude, or wind speed, is found by using the Pythagorean Theorem where numeric values are calculated by taking the square root of v wind squared and u wind squared summed.

To be classified as upwind, the well must be going against the direction of the wind in relation to each county centroid. We follow previous research (Li et al., 2020) and add a 45 degree margin on each side of the spud year wind direction to encompass a treatment area. Note (figure x) represents classification of an upwind well in relation to a county centroid. Downwind wells are mapped as a negative control and include wells directionally, in reference to the county centroid, blowing with the wind. A well 20 km away, located in the same direction of wind, should have less to no impact on agriculture comparative to an upwind well. Final calculation of our treatment is performed after aggregating number of wells per each county year match, providing a discrete county level estimate of total wells developed per year. We include all wells, whether their status deemed active or cancelled, due to the possibility that even non-producing wells may have produced radiation from earlier phases. After summing upwind wells per county and mapping the numeric value to each county-spud year pair, we use the following linear baseline specification (dgp) :

$$y_{cit} = \beta_1 \, upwind \, UOGD \, wells_{it} + \beta_2 \, prec_{it} + \beta_3 \, tavg_{it} + \alpha_i + \gamma_{ts} + \eta_{it} \tag{5}$$

where y_{cit} denotes agricultural yield for crop c in county i during time t. Our treatment variable, $upwind \, wells_{it}$, contains a summation of all upwind wells within the km buffer distance for county year combinations. $prec_{it}$ and $tavg_{it}$ contain annual data at the county level for precipitation and temperate average, we note the work of *Studies in Crop Variation III: The influence of rainfall on the yield of wheat at Rothamsted* (1925) detailing the potential for rainfall to exemplify non-linearity. We create polynomial variables to capture potential non-linearity (RUN model squared*). To account for county specific characteristics that influence yield, we utilize α_i to obtain a dummy structure for county fixed effects. To account for spatiotemporal technological change common to all states, a time fixed effects is employed at the state-year level, γ_{ts}. η_{it} represent the unobservables vector. According to Wooldridge (2005), utilization of a fixed-effects estimator allows for causal interpretation while positive covariance exists between unobserved heterogeneity and time-varying regressors. Using a fixed effects approach, our confound will now be composed into a time invariant component, η_i, and a time varying element, η_t. This brings about concern as standard error assumptions of independence, across both panels and time, are invalid as η_i is invariant over time. Standard errors perform poorly under autocorrelation as overestimation of significance levels occurs due to understated standard deviation of estimated treatment effects (Wooldridge, 2003). To account for the within cluster serial correlation of residuals and regressors we employ a clustered standard error at the state-year level (Abadie et al., 2017), a variant of the Frisch-Waugh theorem where partialled out residuals are summed over each panel at the county level by time.We cluster at the state-year level as standard errors performed better than when clustered at the county level. As a test of robustness, we run a loop iterating at each km starting with 1km and ending with 40km. The upwind and downwind coefficient values, with coefficient standard errors, are shown plotted against km distances (Figure 4). Running the code in R, utilizing the fixest (Bergé, 2018) package for high dimensional fixed effects, we encounter the following error: "Variables 'as.factor(year)1951', 'as.factor(year)1952' and 5086 others have been removed because of collinearity". The linear negative control model, containing the same notation[6] as the positive control model, is specified as :

$$y_{cit} = \beta_1 \, downwind \, UOGD \, wells_{it} + \beta_2 \, prec_{it} + \beta_3 \, tavg_{it} + \alpha_i + \gamma_{ts} + \eta_{it} \tag{6}$$

Our primary method of identification utilizes semi-parametric modeling in DML, as detailed in Chernozhukov et al. (2017) to estimate the causal impact of adverse environmental effects from fracking on agricultural yields. We use DML to control for confoundedness, especially given that $\mathbf{X}_{it}$ is not high dimensional, as machine learning methods help prevent overfitting in nonparametric methods through simultaneous training and cross validation. According to Jung et al. (2021), corroborated by simulation results, DML may be used generally for any identifiable causal functions as estimators hold the key property of double robustness. DML uses residual-on-residual regres-

[6] Standard Errors are also clustered at the state-year level

sion as a method for estimating average treatment effects under unconfoundedness. The process performs a regression of outcomes on covariates, and a second regression of the treatment variable on covariates. The residuals from the first regression are regressed on the residuals from the second regression to provide a causal estimate free of nuisance function bias. Machine learning provides a benefit as errors in the fit of nuisance parameters, or flexible functional forms that account for the relationship between endogenous variables, are residualized to make estimation of the average treatment effect orthogonal to nuisance parameters errors. If the unobservables are properly controlled for, we are able to treat emissions from UOGD as an exogenous shock as in Meyers (2019) to provide a causal upwind estimate. DML is the tool in which causality may be realized in spite of endogeneity. According to Athey (2018), doubly robust methods for average treatment effect estimation provide particular advantages as orthogonalization may flexibly estimate the relationship between outcome and endogenous treatment indicators. In accounting for confoundedness representative from COGD, we follow Bach et al. (2021) in using the following partially linear regression models:

$$y_{cit} = \theta \; upwind \; wells_{it} + f_0(\mathbf{X}_{it}) + \eta_{it} \tag{7}$$

where our treatment variable, $upwind \; wells_{it}$, is a function of our covariate and confounder vector, $\mathbf{X}_{it}$:

$$upwind \; wells_{it} = g_0(\mathbf{X}_{it}) + \epsilon_{it} \tag{8}$$

The "partialling out" option for the Neyman-orthogonal propensity score function is utilized in creating a one-dimensional measure of proximity from multi-dimensional objects to estimate θ[7]. Inference of the θ coefficient in equation 2 has the interpretation of a causal parameter if upwind wells are assumed to be randomly assigned conditional on $\mathbf{X}_{it}$ (Bach et al., 2021). Running a Pearson's product-moment correlation test between UOGD upwind wells and COGD upwind wells,

[7] For brevity, we have omitted the description of the scoring function used to satisfy a unique solution to the method-of-moments estimator for θ while obeying Neyman orhogonality. Details of this input procedure are provided in Bach et al. (2021).

the correlation result of 0.016 is significant at $<$ the .001 level. Thus, given that the null was still rejected, we feel comfortable in our assumption linear relationship between our treatment variables. Equation 4 takes into consideration the confounding relationship where our treatment variable is dependent on covariates, primarily COGD. To capture the true effect of UOGD radiation on yield, we must account for spatiotemporal radiation simultaneously released from COGD. This relationship between treatment variable and confounder is modeled with random forest learners. We specify two random forest learners, with 100 trees and 1 variable split at each tree node, using the mlr3 (Lang et al., 2019) and ranger (Wright and Ziegler, 2017) packages to estimate the nuisance functions, $f_0(\mathbf{X}_{it})$ and $g_0(\mathbf{X}_{it})$. Before deploying the learning algorithms, the sample is split into 10 folds for cross-fitting amongst the produced estimator. We use cross-fitted forms in estimation of θ to avoid overfitting our sample with too flexible of a model. Overfitting could produce a biased estimator and sample-splitting allows the elimination of bias introduced by overfit. The learners are specified by :

$$E[\widehat{Y_{cit,k}|\mathbf{X}}_{it,k}] \tag{9}$$

$$E[\widehat{D_{it,k}|\mathbf{X}}_{it,k}] \tag{10}$$

where D is the treatment variable of *upwind wells*$_{it}$, k denotes the fold number, and Y is crop yield, y_{cit}. The vector $\mathbf{X}_{it}$ consists of confounders and covariates. The concept of DML is to offset regularization bias by stripping out the effect of $\mathbf{X}_{it}$ from $\mathbf{D}_{it}$. We realize this by residualizing the true value from conditional expectation, partialling out the effect of COGD to include the error resulting in random forest estimation of the nuisance parameters:

$$\widetilde{Z_k} = Y_{cit,k} - E[\widehat{Y_{cit,k}|\mathbf{X}}_{it,k}] \tag{11}$$

$$\widetilde{W_k} = D_{it,k} - E[\widehat{D_{it,k}|\mathbf{X}}_{it,k}] \tag{12}$$

Finally, we regress Z on W across for each kth fold to obtain causal estimates, θ_{1k}, of UOGD wells on agricultural yield.

$$\widetilde{Z_k} = \theta_{0,k} + \theta_{1,k}\ \widetilde{W_k} \tag{13}$$

Estimation of the causal parameter, $\widehat{\theta_1}$, is an average across the number of k folds :

$$\widehat{\theta_1} = \frac{1}{10} \sum_{k=1}^{10} \theta_{1,k} \tag{14}$$

To appreciate our DML results, we naively add the confounder COGD to our baseline model:

$$y_{cit} = \beta_1 \; upwind\; UOGD\; wells_{it} + \beta_2 \; upwind\; COGD\; wells_{it} + \beta_3 \; prec_{it} + \beta_4 \; tavg_{it} + \alpha_i + \gamma_{ts} + \eta_{it} \tag{15}$$

Findings are reported in Figure 6 and we notice that the corn results are much different than our baseline results. We notice the benefits of DML on our upwind corn model, whereas both downwind DML estimators performed worse than the both baseline negative controls. Rejection of the null in our Pearson product-moment correlation test, signaling potential multicollinearity, may drive these different estimations. We acknowledge the potential limitations of machine learning (Abadie and Kasy, 2019) while understanding that our objects of interest are of many different treatment variables. Our observed data allows comfortability in using Machine Learning as (i) our treatment variable is not considered a rare event and (ii) the variation in outcome variables are relatively robust (Storm et al., 2019). Literature details additional concerns of endogeneity resulting from shale plays being drill depth heterogenous, thus creating variation in vertical distance drilled throughout each individual formation (Hill, 2018). Variation in drilling depth may provide heterogenous impacts to agriculture productivity through air pollution and water contamination (Vidic et al., 2013). We estimate an average treatment effect across all shale plays while acknowledging that treatment effects may be heterogenous amongst region. A further limitation of this study is that the large variation in county size exhibits bias toward capturing causal effects in smaller counties. As the size of county increases, points of production will lie further from the county centroid resulting in a biased estimate. According to the USDA's Natural Resources Conservation Service (United States Department of Agriculture, 2020*a*), there are 3,111 total counties within the continental United States whereas the total area is 3,129,426 square miles (United States Census Reference Files, 2020). Converting from miles to km results in a value of 8105176.132. $\frac{8105176.132}{3{,}111}$ infers an average county size of $\approx$ 2605 square km. Fully acknowledging the limitations within using county centroids as point of production, we suggest further, more intensive($grant?), research using machine learning and artificial intelligence to map true agricultural yield locations to UOGD wells.

Data

We create an annual dataset from various governmental entities, at the county level, encompassing 1950 to 2020. Oil and gas development observations are obtained from Enverus drillinginfo to include 3,946,198 vertically, horizontally, directionally, and unidentified drilled wells in the United States from 1900 to 2021. Although our output starts in year 1950, we include all drill well years in hopes of training a more accurate classifier model in predicting directionally and unidentified wells. For classification of UOGD wells within the Enverus dataset, we utilize the 336,646 observations labeled horizontally drilled while deploying a random forest algorithm to separate the 1,331,081 directionally and unidentified drilled wells into horizontally drilled or vertically drilled. We utilize the randomForest package in R (Liaw and Wiener, 2002), with Breiman's classic algorithm, and train the random forest classifier with 2,615,116 horizontal and vertical wells using 100 trees (40) and 12 random variables sampled at each split. The classifier exemplified an average out of bag sample prediction error of approximately .6% using our 100 tree model (Figure 7). When deployed upon unidentified and directional wells, the model identified 138,359 horizontally drilled wells giving us a total of 475,005 wells classified as UOGD. NA values ($\approx 25\%$ percent of the dataset) for spud year were interpolated using a k nearest neighbors imputation algorithm where k=1 to interpolate the closest neighbor given a set of predictor variables in Enverus (approximately 34 when trimmed to exclude columns with larger NA percentages). We use the k nearest neighbor method as our random forest regression to interpolate spud year had an unacceptable mean residual error of approximately 5 years. After all variables have been used to include breadth of prediction, we omit all spud years prior to 1978 given that our NARR wind data starts in 1978. Thus, the final dataset used for analysis includes 389,323 UOGD wells while the descriptive statistics were taken using the full UOGD dataset. Producing status of the well was not taken into consideration as excluding incomplete or nonactive wells would introduce a limitation in our study. Logic behind this is that the potential mechanism for radiation emission remains wide, all documented activity should be included in analysis as hazardous materials are released during all stages of well construction.

For agricultural yield data, we utilize United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) agricultural data encompassing 1950 to 2010 from Burke and

Emerick (2016)[8] while observations from 2011 to 2021 are obtained directly from the NASS website (United States Department of Agriculture, 2020*b*). Agricultural yield data includes corn, soybeans, and cotton, however, cotton was not available at the county-annual level online [9]. Similarly, we obtain area planted information for each county in calculating a net production loss.

We standardize covariate data as we obtain two separate, but similar, data sources. Precipitation and temperature average is included within the dataset used from Burke and Emerick (2016). For years 2011-2019, we utilize information contained within the Parameter elevation Regression on Independent Slopes Model (PRISM) dataset provided by the National Center for Atmospheric Research (2021)(NCAR). The PRISM spatial resolution consists of a 2.5 x 2.5 mile grid. A balanced panel is created in keeping the set of weather stations constant over our period of interest. Missing values are interpolated by using a distance-weighted average of the cumulative density function for surrounding stations [10].

[8] Burke accredits, with gratitude, Michael Roberts and Wolfram Schlenker for sharing their data.

[9] Inference of yield was attempted, but acres planted and production data for cotton is not available at the county-year level. We use only corn and soybeans yield in our analysis for completeness.

[10] We acknowledge Wolfram Schlenker for his method of balancing PRISM station data

Results

According to Bartik et al. (2019), local communities that have allowed fracking have realized net benefits. Negative impacts of fracking have the potential for large costs in non-local communities through radiation exposure in food as evidenced by reduced yield. While these non-local communities do not experience the same economic benefits, they do incur economic losses associated with UOGD. To calculate monetary damages resulting from UOGD, we use 2017 national census data from USDA's NASS for total crop area in the U.S[11]. We use our 20km distance mapping to create a dummy variable for counties with upwind wells > 1 in 2017. A total of 52 county centroids had at least 1 UOGD well within 25km in 2017. Using an aggregation of area harvested for all of the 52 counties results in 431,381 impacted acres. An upwind well will cause, on average, a .08 decrease in bushels of soybeans per acre [12]. This results in a net loss of 34,510 soybean bushels. Using the average price of a soybean bushel for 2017 of \$9.78, we calculate the crop loss resulting from UOGD to be \$337,508 in 2017. Given that $\approx 5\%$ of total UOGD wells were spudded in 2017, we extrapolate an overall loss of \$1,687,539[13]. Using similar methodologies for calculating corn loss, a total of 727,910 acres were used for production within 25km of UOGD wells. We found a net loss of 72,791 corn bushels with an average yearly price of \$3.36. Net production losses to corn include a revenue shortfall of \$244,577 in 2017. Extrapolating these results to all years results in a \$1,222,889 loss to agricultural producers. We found estimated damages from UOGD for both crops to be $\approx$ \$2,910,428

Bootstrapping our DML results at 20km provides a disappointing result. I assume, due to way the dataset is constructed, this happened because many of the yield observations are not impacted by UOGD. Thus, our bootstrapped data set with replacement may be biased toward non-UOGD related counties. The coefficient distribution is positive, not good.

Algorithm 1 for DML performs well on soy upwind, but the downwind measures for corn and soy

[11] 2017 is the most recent Census data, which contains information on all producers, whereas 2020 survey data may have non-reported information

[12] The average acre in 2017 produced 160 bushels of corn and 47 bushels of soy.

[13] Constructing a price index from the IQR of UOGD spud year, 1995-2015, we find an average value of \$3.32 per corn bushel and \$8.08 per soybean bushel (United States Department of Agriculture: Economics, Statistics, and Market Information System, 2020)

are all messed up. Algorithm 2 for DML performs better on downwind observations, and corn. Algorithm 2 optimizes Z* for the score function (Take formulas from barth.

References

Bibliography

Abadie, A., Athey, S., Imbens, G. and Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering ?. Working Paper. National Bureau of Economic Research.

Abadie, A. and Kasy, M. 2019. Choosing Among Regularized Estimators in Empirical Economics: The Risk of Machine Learning. *The Review of Economics and Statistics* 101 (5): 743–762.

Aben, R. C. H., Barros, N., van Donk, E., Frenken, T., Hilt, S., Kazanjian, G., Lamers, L. P. M., Peeters, E. T. H. M., Roelofs, J. G. M., de Senerpont Domis, L. N., Stephan, S., Velthuis, M., de Waal, D. B. V., Wik, M., Thornton, B. F., Wilkinson, J., DelSontro, T. and Kosten, S. 2017. Cross Continental Increase in Methane Ebullition Under Climate Change. *Nature Communications* 8 (1).

Allcott, H. and Keniston, D. 2017. Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America. *The Review of Economic Studies* 85 (2): 695–731.

Athey, S. (2018). The Impact of Machine Learning on Economics. *in* 'The Economics of Artificial Intelligence: An Agenda'. NBER Chapters. National Bureau of Economic Research, Inc. pp. 507–547.

Bach, P., Chernozhukov, V., Kurz, M. S. and Spindler, M. (2021). 'DoubleML – An Object-Oriented Implementation of Double Machine Learning in R'.

Bamberger, M. and Oswald, R. E. 2012. Impacts of Gas Drilling on Human and Animal Health. *New Solutions: A Journal of Environmental and Occupational Health Policy* 22 (1): 51–77.

Bartik, A. W., Currie, J., Greenstone, M. and Knittel, C. R. 2019. The local economic and welfare consequences of hydraulic fracturing. *American Economic Journal: Applied Economics* 11 (4): 105–155.

Bergé, L. 2018. Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package . *CREA Discussion Papers* (13).

Black, K. J., Boslett, A., Hill, E., Ma, L. and McCoy, S. 2020. A Review of the Economic, Environmental, and Health Impacts of the Fracking Boom. *SSRN Electronic Journal* .

Black, K. J., McCoy, S. J. and Weber, J. G. 2019. Fracking and Indoor Radon: Spurious Correlation or Cause for Concern?. *Journal of Environmental Economics and Management* 96: 255–273.

Boomhower, J. (2021). When Do Environmental Externalities Have Electoral Consequences? Evidence from Fracking. Technical report. National Bureau of Economic Research.

Boslett, A., Guilfoos, T. and Lang, C. 2016. Valuation of Expectations: A Hedonic Study of Shale Gas Development and New York's Moratorium. *Journal of Environmental Economics and Management* 77: 14–30.

Brown, V. J. 2014. Radionuclides in Fracking Wastewater: Managing a ToxicBlend. *Environmental Health Perspectives* 122 (2).

Burke, M. and Emerick, K. (2016). 'Replication data for: Adaptation to climate change: Evidence from us agriculture'.

Bustad, L. K., George, L. A., Marks, S., Warner, D. E., Barnes, C. M., Herde, K. E., Kornberg, H. A. and Parker, H. M. 1957. Biological Effects of I 131 Continuously Administered to Sheep. *Radiation Research* 6 (3): 380.

Cascio, E. U. and Narayan, A. 2020. Who Needs a Fracking Education? The Educational Response to Low-Skill-Biased Technological Change. *ILR Review* .

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C. and Newey, W. 2017. Double/Debiased/Neyman Machine Learning of Treatment Effects. *American Economic Review* 107 (5): 261–65.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J. 2018. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21 (1): C1–C68.

Colborn, T., Kwiatkowski, C., Schultz, K. and Bachran, M. 2011. Natural Gas Operations from a Public Health Perspective. *Human and Ecological Risk Assessment: An International Journal* 17 (5): 1039–1056.

Currie, J., Greenstone, M. and Meckel, K. 2017. Hydraulic Fracturing and Infant Health: New Evidence from Pennsylvania. *Science Advances* 3 (12): e1603021.

de Gouw, J. A., Veefkind, J. P., Roosenbrand, E., Dix, B., Lin, J. C., Landgraf, J. and Levelt, P. F. 2020. Daily Satellite Observations of Methane from Oil and Gas Production Regions in the United States. *Nature Research : Scientific Reports* 10 (1).

Farah, N. (2016). The Effects of Hydraulic Fracturing on Agricultural Productivity. *in* 'In Implications of North American Energy Self-Sufficiency, 34th USAEE/IAEE North American Conference'.

Feyrer, J., Mansur, E. T. and Sacerdote, B. 2017. Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution. *American Economic Review* 107 (4): 1313–1334.

Gibson, J. 2013. Bombing North Dakota: Living Amid the Bakken Oil Boom. *Earth Island Journal* 27 (4).

Gopalakrishnan, S. and Klaiber, H. A. 2013. Is the Shale Energy Boom a Bust for Nearby Residents? Evidence from Housing Values in Pennsylvania. *American Journal of Agricultural Economics* 96 (1): 43–66.

Hausman, C. and Kellogg, R. 2015. Welfare and Distributional Implications of Shale Gas. *NBER Working Paper* .

Hilal, M., Attallah, M., Mohamed, G. Y. and Fayez-Hassan, M. 2014. Evaluation of Radiation Hazard Potential of TENORM Waste from Oil and Natural Gas Production. *Journal of Environmental Radioactivity* 136: 121–126.

Hill, E. L. 2018. Shale gas development and infant health: Evidence from Pennsylvania. *Journal of Health Economics* 61: 134–150.

Hill, E. and Ma, L. 2017. Shale Gas Development and Drinking Water Quality. *American Economic Review* 107 (5): 522–525.

Hitaj, C., Boslett, A. and Weber, J. 2014. Shale Development and Agriculture. *Choices* .

Howarth, R. W. and Ingraffea, A. 2011. Should Fracking Stop?. *Nature* 477: 271–275.

Howarth, R. W., Santoro, R. and Ingraffea, A. 2011. Methane and the Greenhouse-Gas Footprint of Natural Gas from Shale Formations. *Climatic Change* 106 (4): 679–690.

Jung, Y., Tian, J. and Bareinboim, E. 2021. Estimating Identifiable Causal Effects through Double Machine Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 35 (13): 12113–12122.

Kearney, M. S. and Wilson, R. 2018. Male Earnings, Marriageable Men, and Nonmarital Fertility: Evidence from the Fracking Boom. *The Review of Economics and Statistics* 100 (4): 678–690.

Kemball-Cook, S., Bar-Ilan, A., Grant, J., Parker, L., Jung, J., Santamaria, W., Mathews, J. and Yarwood, G. 2010. Ozone Impacts of Natural Gas Development in the Haynesville Shale. *Environmental Science & Technology* 44 (24): 9357–9363.

Kimball, B. A. 1983. Carbon Dioxide and Agricultural Yield: An Assemblage and Analysis of 430 Prior Observations 1. *Agronomy Journal* 75 (5): 779–788.

Kolb, W. and Wojcik, M. 1985. Enhanced radioactivity due to natural oil and gas production and related radiological problems. *Science of The Total Environment* 45: 77–84.

Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., Au, Q., Casalicchio, G., Kotthoff, L. and Bischl, B. 2019. mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software* .

Li, L., Blomberg, A. J., Spengler, J. D., Coull, B. A., Schwartz, J. D. and Koutrakis, P. 2020. Unconventional oil and gas development and ambient particle radioactivity. *Nature Communications* 11 (1).

Liaw, A. and Wiener, M. 2002. Classification and Regression by randomForest. *R News* 2 (3): 18–22.

Marchand, J. and Weber, J. 2017. Local Labor Markets and Natural Resources : A synthesis of the Literature. *Journal of Economic Surveys* 32 (2): 469–490.

Mason, C. F., Muehlenbachs, L. A. and Olmstead, S. M. 2015. The Economics of Shale Gas Development. *Annual Review of Resource Economics* 7 (1): 269–289.

Mba, C. 2013. Induced Mutations Unleash the Potentials of Plant Genetic Resources for Food and Agriculture. *Agronomy* 3 (1): 200–231.

McCawley, M. (2013). Air, Noise, and Light Monitoring Results for Assessing Environmental Impacts of Horizontal Gas Well Drilling Operations. Technical report.

McKenzie, L. M., Witter, R. Z., Newman, L. S. and Adgate, J. L. 2012. Human Health Risk Assessment of Air Emissions from Development of Unconventional Natural Gas Resources. *Science of The Total Environment* 424: 79–87.

Meyers, K. 2019. In the Shadow of the Mushroom Cloud: Nuclear Testing, Radioactive Fallout, and Damage to U.S. Agriculture, 1945 to 1970. *The Journal of Economic History* 79 (1): 244–274.

Micco, V. D., Arena, C., Pignalosa, D. and Durante, M. 2010. Effects of sparsely and densely ionizing radiation on plants. *Radiation and Environmental Biophysics* 50 (1): 1–19.

Muehlenbachs, L., Spiller, E. and Timmins, C. 2015. The Housing Market Impacts of Shale Gas Development. *American Economic Review* 105 (12): 3633–3659.

National Center for Atmospheric Research 2021. Climate Data.

Newell, R. and Raimi, D. 2015. Shale Public Finance: Local Government Revenues and Costs Associated with Oil and Gas Development. *NBER Working Paper* .

Ogundare, F. and Adekoya, O. 2015. Gross Alpha and Beta Radioactivity in Surface Soil and Drinkable Water Around a Steel Processing Facility. *Journal of Radiation Research and Applied Sciences* 8 (3): 411–417.

Peng, R. D., Bell, M. L., Geyh, A. S., McDermott, A., Zeger, S. L., Samet, J. M. and Dominici, F. 2009. Emergency Admissions for Cardiovascular and Respiratory Diseases and the Chemical Composition of Fine Particle Air Pollution. *Environmental Health Perspectives* 117 (6): 957–963.

Rasmussen, S. G., Ogburn, E. L., McCormack, M., Casey, J. A., Bandeen-Roche, K., Mercer, D. G. and Schwartz, B. S. 2016. Association Between Unconventional Natural Gas Development in the Marcellus Shale and Asthma Exacerbations. *JAMA Internal Medicine* 176 (9): 1334.

Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., Pugh, T. A. M., Rolinski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W. and Frieler, K. 2017. Consistent Negative Response of US Crops to High Temperatures in Observations and Crop Models. *Nature Communications* 8 (1).

Stadler, L. 1928. Genetic Effects of X-rays in maize. *Proc. Nat. Acad. Sci.* .

Steinzor, N., Subra, W. and Sumi, L. 2012. Gas Patch Roulette: How Shale Gas Development Risks Public Health in Pennsylvania. *Earthworks Oil and Gas Accountability Project* .

Storm, H., Baylis, K. and Heckelei, T. 2019. Machine Learning in Agricultural and Applied Economics. *European Review of Agricultural Economics* 47 (3): 849–892.

Studies in Crop Variation III: The influence of rainfall on the yield of wheat at Rothamsted 1925. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* 213 (402-410): 89–142.

Thakor, R. (2018). Liquidity Windfalls and Reallocation: Evidence from Farming and Fracking. Technical report.

Torres, L., Yadav, O. P. and Khan, E. 2018. Risk assessment of human exposure to Ra-226 in oil produced water from the Bakken Shale. *Science of The Total Environment* 626: 867–874.

United States Census Reference Files 2020. State Area Measurements and Internal Point Coordinates. `https://www.census.gov/geographies/reference-files/2010/geo/state-area.html`.

United States Department of Agriculture 2020*a*. County FIPS codes. `https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/home/?cid=nrcs143_013697`.

United States Department of Agriculture 2020*b*. National Agriculture Statistics Survey. `https://quickstats.nass.usda.gov/`.

United States Department of Agriculture: Economics, Statistics, and Market Information System 2020. U.S. Corn and Soybean Prices. `https://afdc.energy.gov/data/10338`.

United States Energy Information Administration 2020. U.S. Total Horizontal Wells by Production Rate Bracket. `https://www.eia.gov/petroleum/wells/`.

U.S. Committee on Interstate and Foreign Commerce 1980. The Forgotten Guinea Pigs" A Report on Health Effects of Low-Level Radiation Sustained as a Result of the Nuclear Weapons Testing Program. *United States Government* .

U.S. Department of Energy 2009. Modern Shale Gas Development in the United States: A Primer. *Work prepared under DE-FG26- 04NT15455* .

U.S. Environmental Protection Agency 2017. Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States. *Office of Research and Development Executive Summary* .

Vengosh, A., Jackson, R. B., Warner, N., Darrah, T. H. and Kondash, A. 2014. A Critical Review of the Risks to Water Resources from Unconventional Shale Gas Development and Hydraulic Fracturing in the United States. *Environmental Science & Technology* 48 (15): 8334–8348.

Vidic, R. D., Brantley, S. L., Vandenbossche, J. M., Yoxtheimer, D. and Abad, J. D. 2013. Impact of Shale Gas Development on Regional Water Quality. *Science* 340 (6134): 1235009–1235009.

Werner, A. K., Vink, S., Watt, K. and Jagals, P. 2015. Environmental Health Impacts of Unconventional Natural Gas Development: A Review of the Current Strength of Evidence. *Science of The Total Environment* 505: 1127–1141.

West Virginia Department of Environmental Protection (2013). Noise, Light, Dust, and Volatile Organic Compounds Generated by the Drilling of Horizontal Wells Related to the Well Location Restriction Regarding Occupied Dwelling Structures. Technical report.

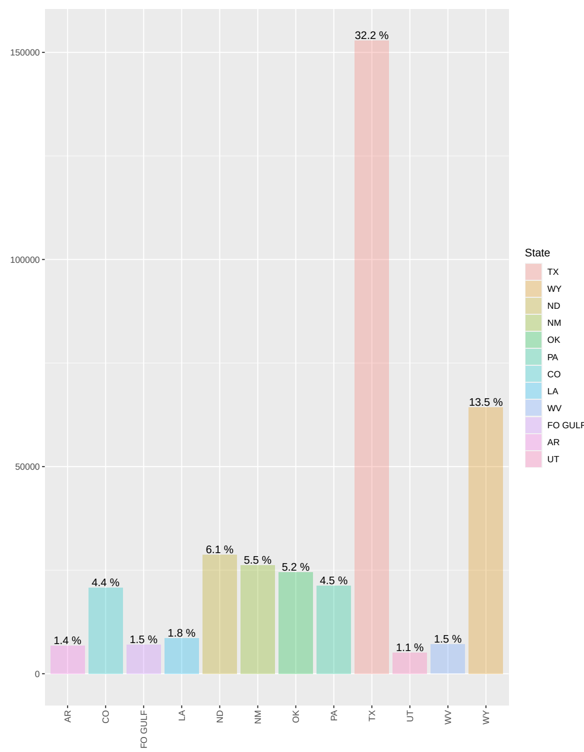
Wooldridge, J. M. 2003. Cluster-sample methods in applied econometrics. *American Economic Review* 93 (2): 133–138.

Wooldridge, J. M. 2005. Fixed-Effects and Related Estimators for Correlated Random-Coefficient and Treatment-Effect Panel Data Models. *The Review of Economics and Statistics* 87 (2): 385–390.

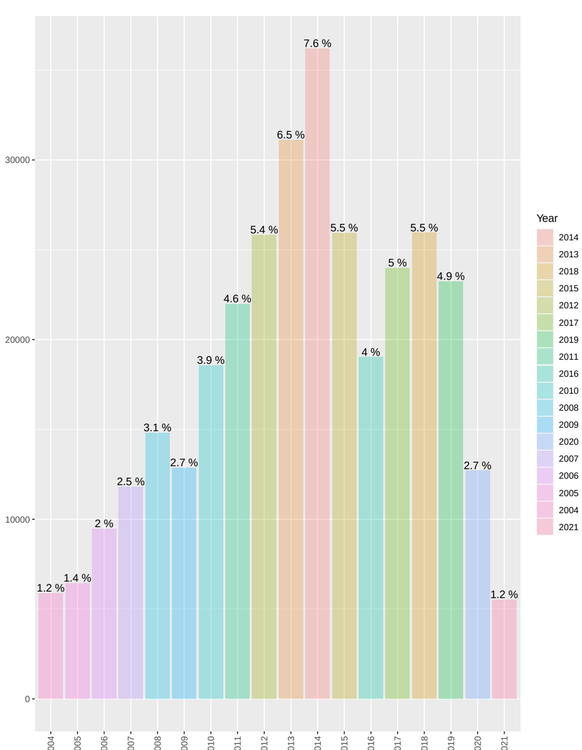
Wright, M. N. and Ziegler, A. 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* 77 (1).

Zhang, T., Hammack, R. W. and Vidic, R. D. 2015. Fate of Radium in Marcellus Shale Flowback Water Impoundments and Assessment of Associated Health Risks. *Environmental Science & Technology* 49 (15): 9347–9354.

Main Figures and Tables

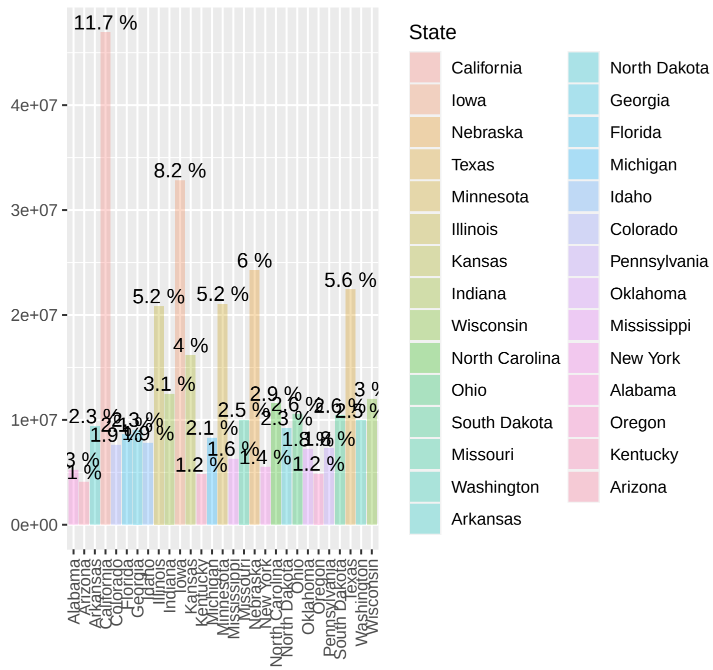


(a) UOGD Well Count by State, including states with > 1% of total wells

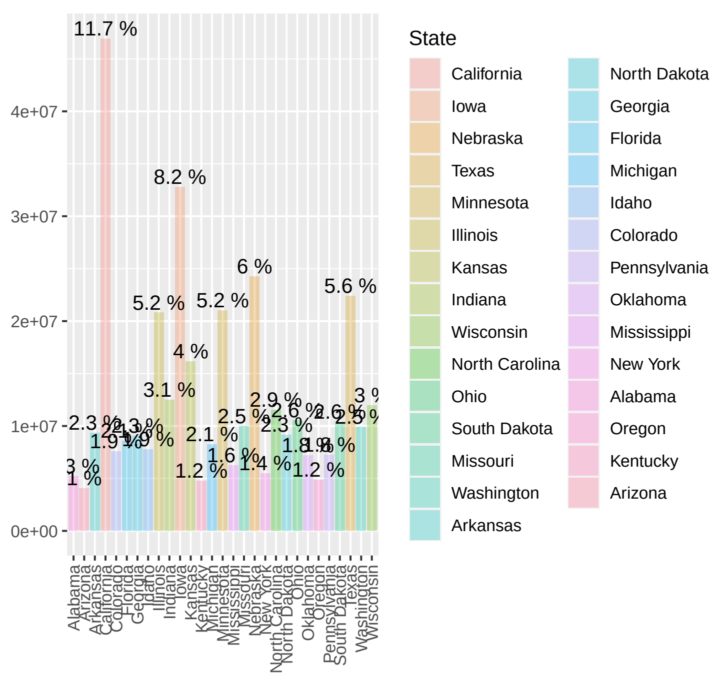


(b) UOGD Well Count by Year, including years with > 1% of total wells

Figure 1: UOGD Well Count by State and Year

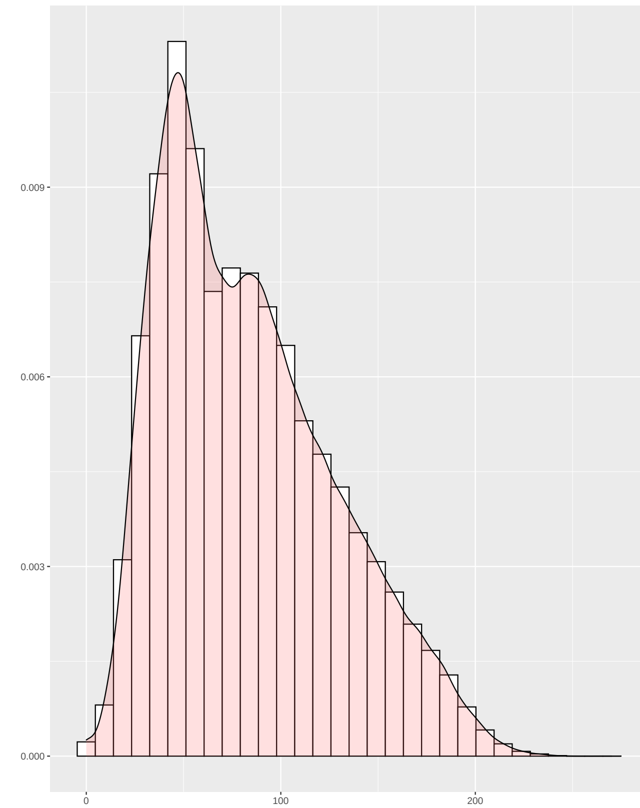


(a) Cash Receipts by State, including states with > 1% of total receipts

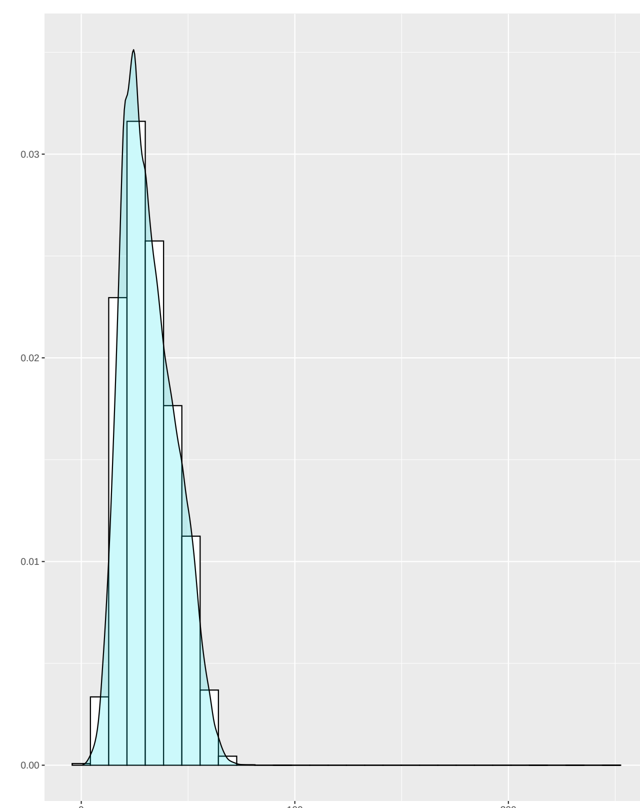


(b) Cash Receipts by Commodity, including states with > 1% of total receipts

Figure 2: 2013 Commodity Cash Receipts



(a) Density of Corn



(b) Density of Soy

Figure 3: Distribution of Agricultural Yield

Table 1: Descriptive Statistics

Variable	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Corn Yield	141,239	87.69	44.75	0.00	51.60	118.00	296.30
Soy Yield	117,397	31.66	12.29	0.70	22.00	40.10	249.00
UOGD Spud Year	475,005	1999	24	1900	1995	2015	2021
COGD Spud Year	1,245,350	1994	11	1900	1984	2005	2021

Table 2: Baseline Results at 20km

Dependent Variables: Model:	Soy	Soy	Soy	Corn	Corn	Corn
upwnd_well	-0.0414**	-0.0414*	-0.0414*	0.1020**	0.1020	0.1020
	(0.0189)	(0.0242)	(0.0242)	(0.0430)	(0.0659)	(0.0659)
Fixed-effects						
state×year	Yes	Yes		Yes	Yes	
fips	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Standard-Errors	state×year	fips	fips	state×year	fips	fips
Observations	117,361	117,361	117,361	153,108	153,108	153,108
R^2	0.88333	0.88333	0.88331	0.65316	0.65316	0.65316
Within R^2	7.25×10^{-5}	7.25×10^{-5}	0.83128	3.21×10^{-5}	3.21×10^{-5}	0.52561

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 3: Baseline Results with COGD at 20km

Dependent Variables:						
Model:	Soy	Soy	Soy	Corn	Corn	Corn
upwnd_well	-0.0421**	-0.0421*	-0.0421*	0.1020**	0.1020	0.1020
	(0.0190)	(0.0243)	(0.0243)	(0.0431)	(0.0659)	(0.0659)
upwnd_well_vertical	0.0157**	0.0157**	0.0157**	-0.0006	-0.0006	-0.0006
	(0.0072)	(0.0063)	(0.0063)	(0.0115)	(0.0082)	(0.0082)
Fixed-effects						
state×year	Yes	Yes		Yes	Yes	
fips	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Standard-Errors	state×year	fips	fips	state×year	fips	fips
Observations	117,361	117,361	117,361	153,108	153,108	153,108
R^2	0.88334	0.88334	0.88332	0.65316	0.65316	0.65316
Within R^2	0.00017	0.00017	0.83130	3.21×10^{-5}	3.21×10^{-5}	0.52561

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

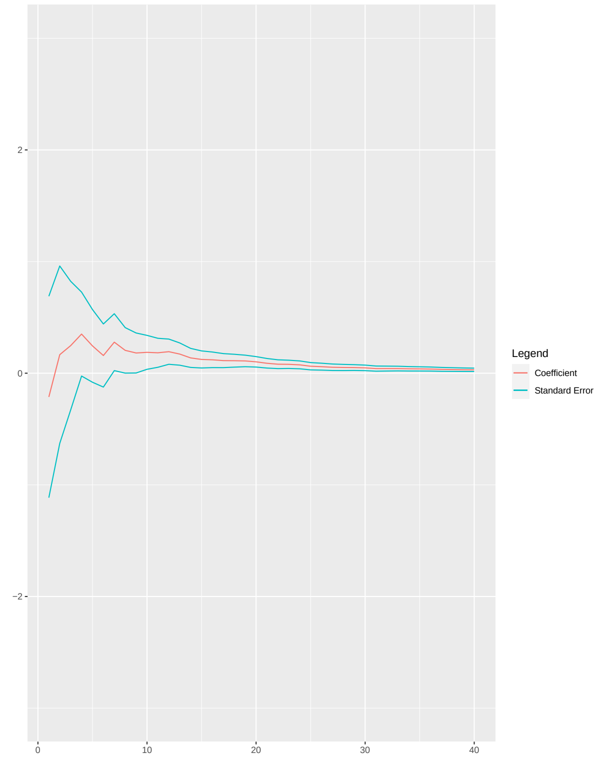
Table 4: DML Results at 20km

Model:	Corn	Soy
upwnd_well	-0.10	-0.08**
	(0.0618)	(0.025)

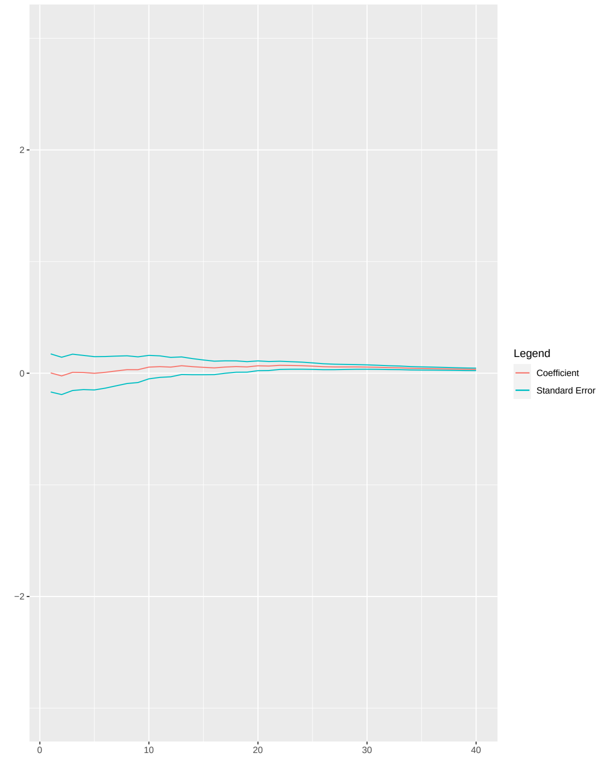
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Order of Code (Post to GitHub, but waiting for completion of paper)

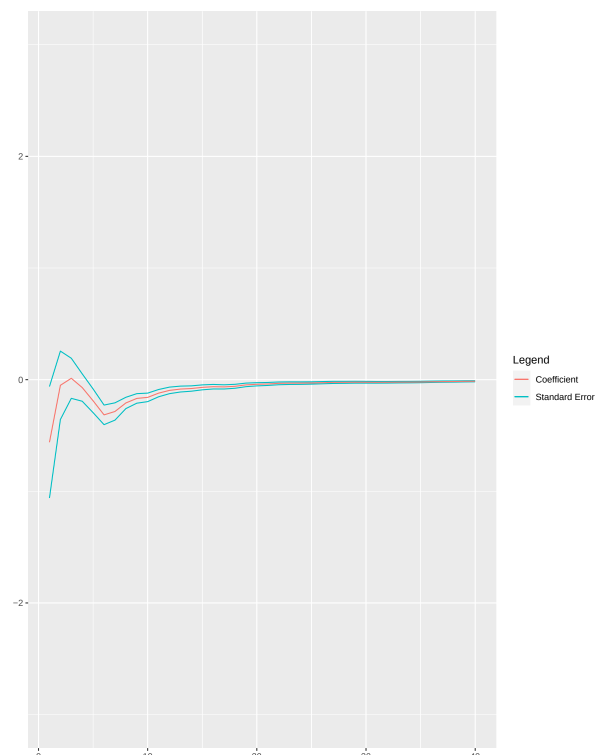
- create well production data
- randomForest for unidentified well classification
- K nearest neighbor to interpolate missing Spud Year
- calculate distance between county centroid and wells
- convert NARR wind data to a usable format
- create IDW wind measure from NARR raw
- match yearly wind to each well's spud year
- map upwind and downwind wells by wind direction
- subset mapped wells by km buffer distance
- baseline robustness
- baseline with COGD robustness
- DML robustness
- calculate crop loss
- output visuals



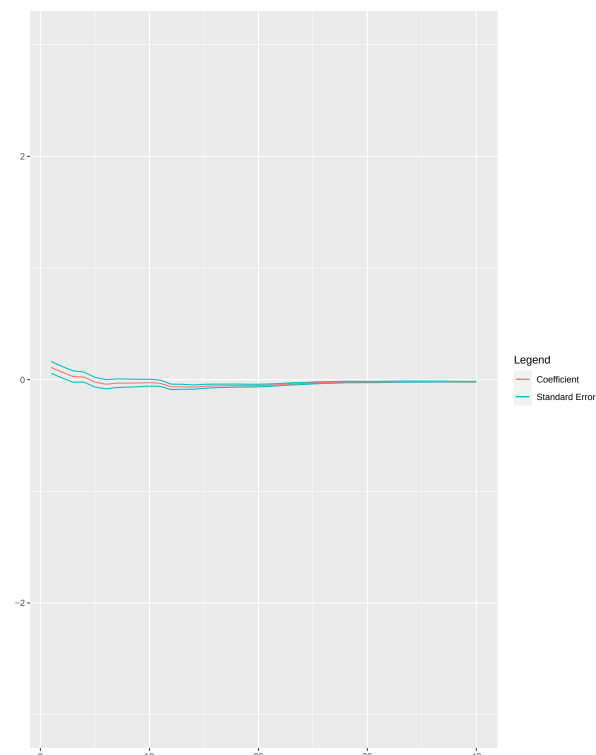
(a) Corn ATE for Upwind Wells



(b) Corn ATE for Downwind Wells

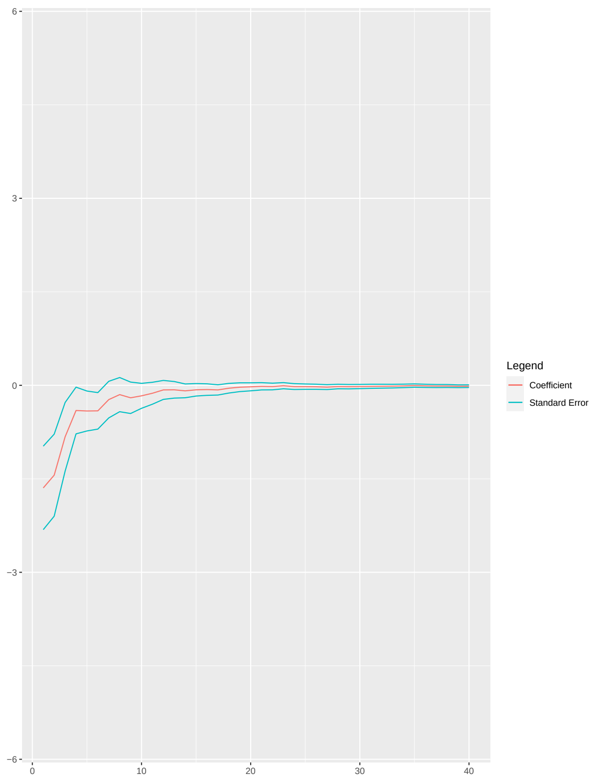


(c) Soy ATE for Upwind Wells

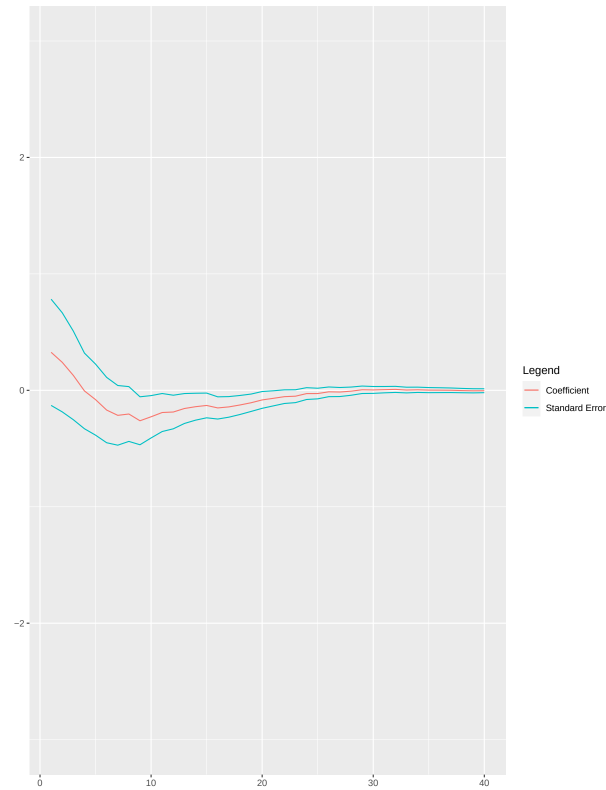


(d) Soy ATE for Downwind Wells

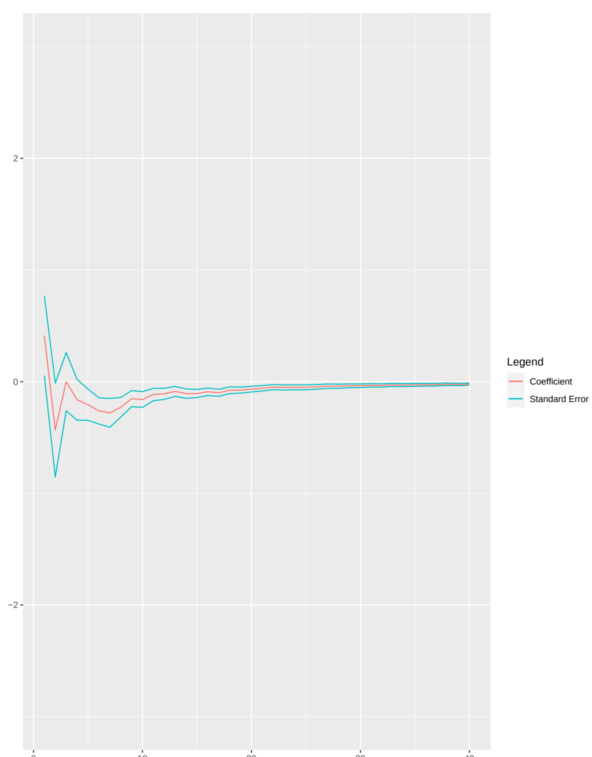
Figure 4: Baseline ATE Results



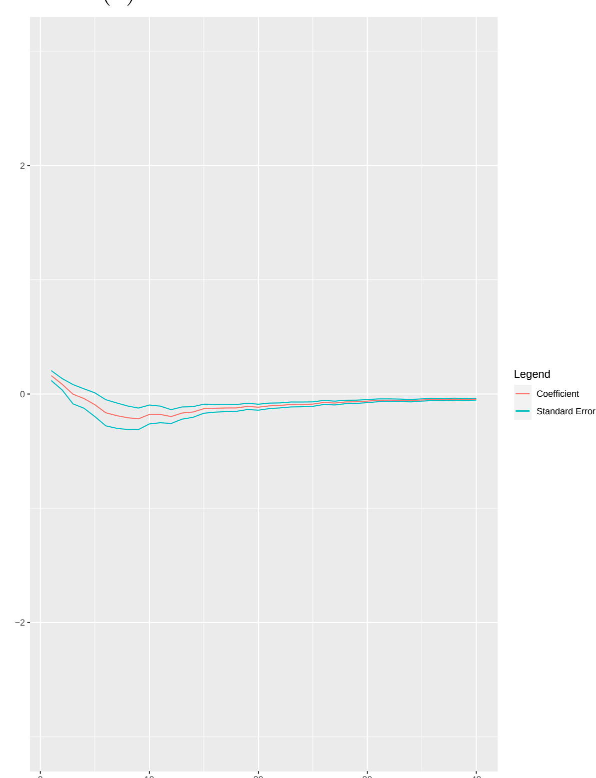
(a) Corn ATE for Upwind Wells



(b) Corn ATE for Downwind Wells

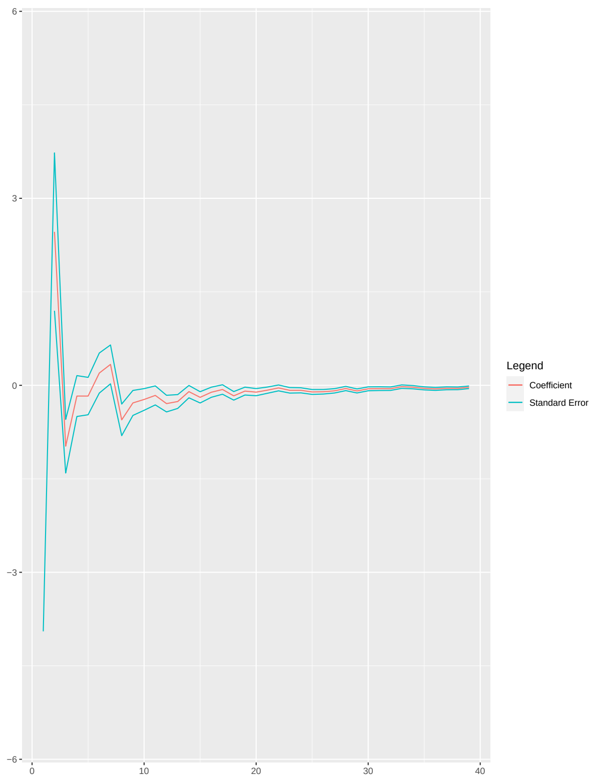


(c) Soy ATE for Upwind Wells

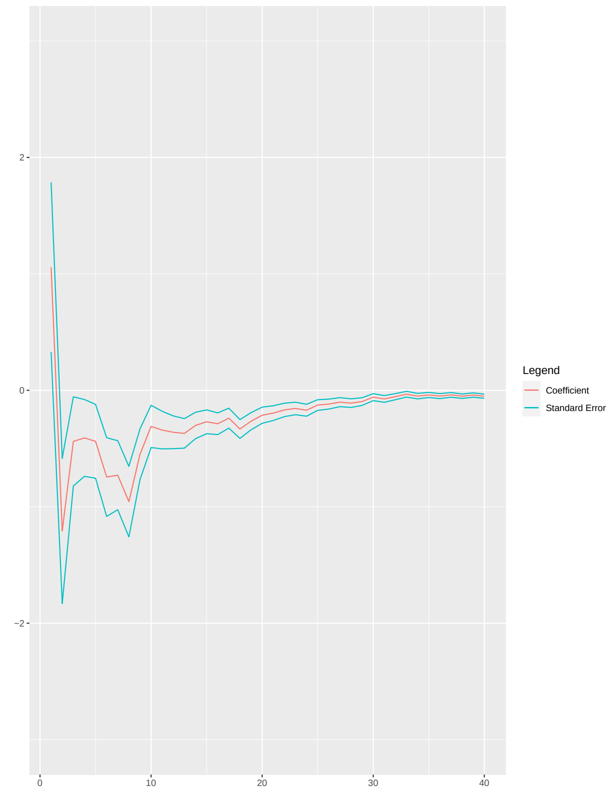


(d) Soy ATE for Downwind Wells

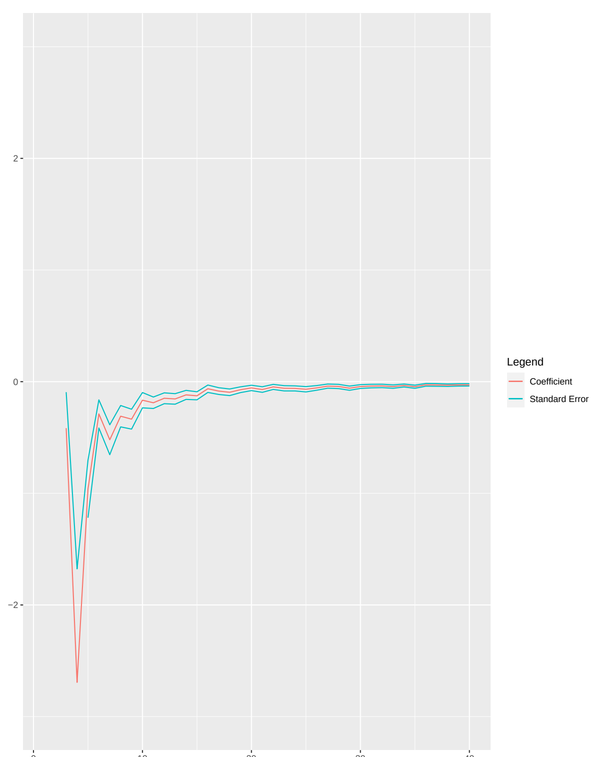
Figure 5: DML alg2 ATE Results



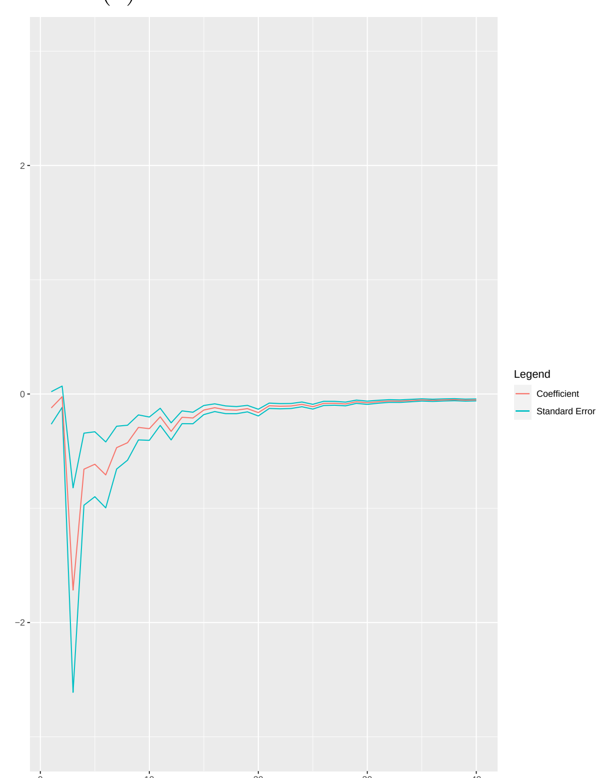
(a) Corn ATE for Upwind Wells



(b) Corn ATE for Downwind Wells

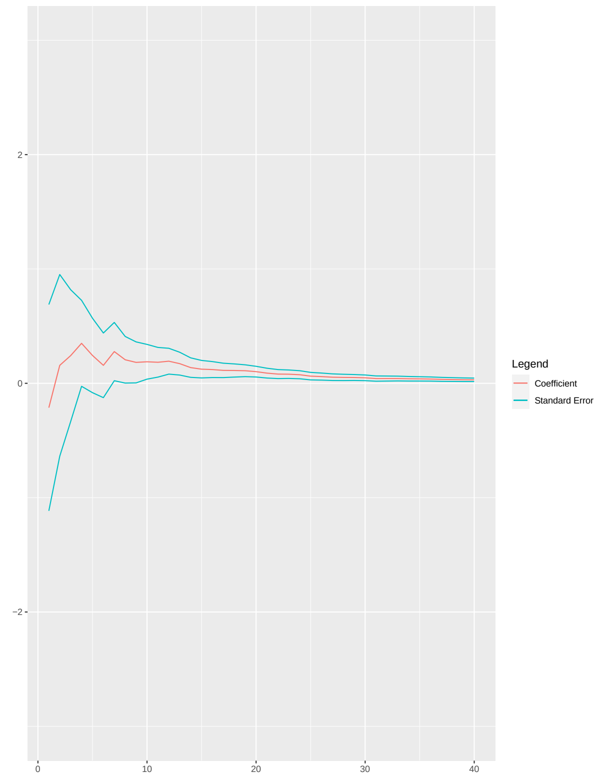


(c) Soy ATE for Upwind Wells

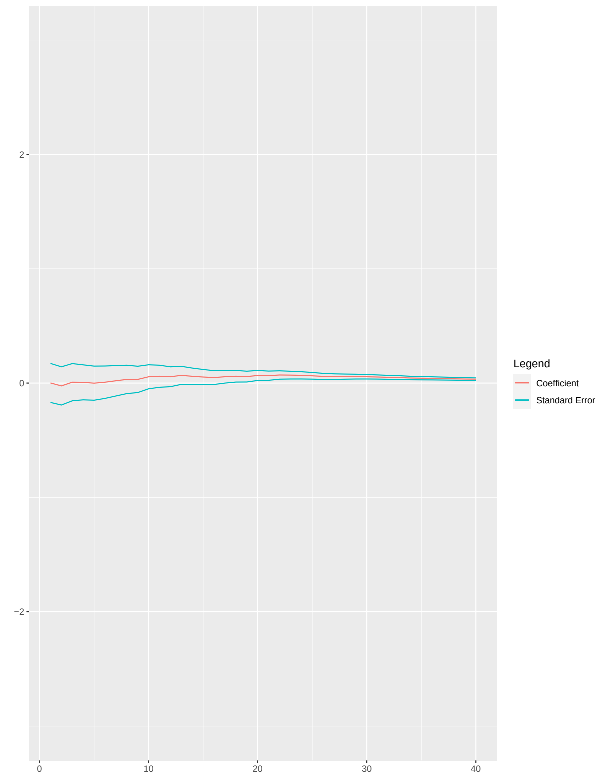


(d) Soy ATE for Downwind Wells

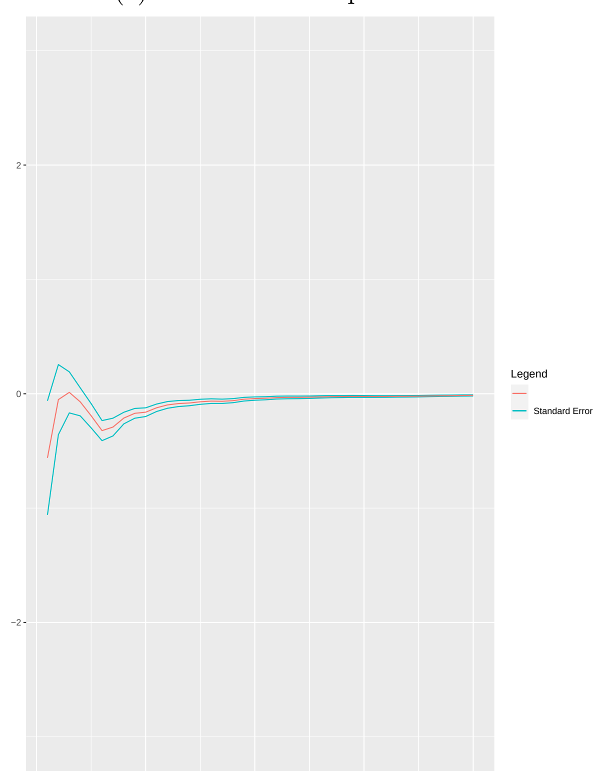
Figure 6: DML alg1 ATE Results



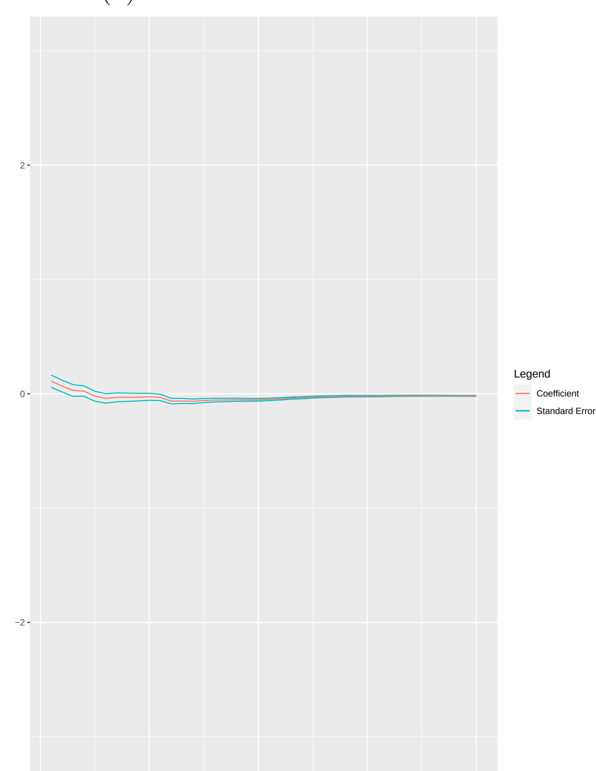
(a) Corn ATE for Upwind Wells



(b) Corn ATE for Downwind Wells

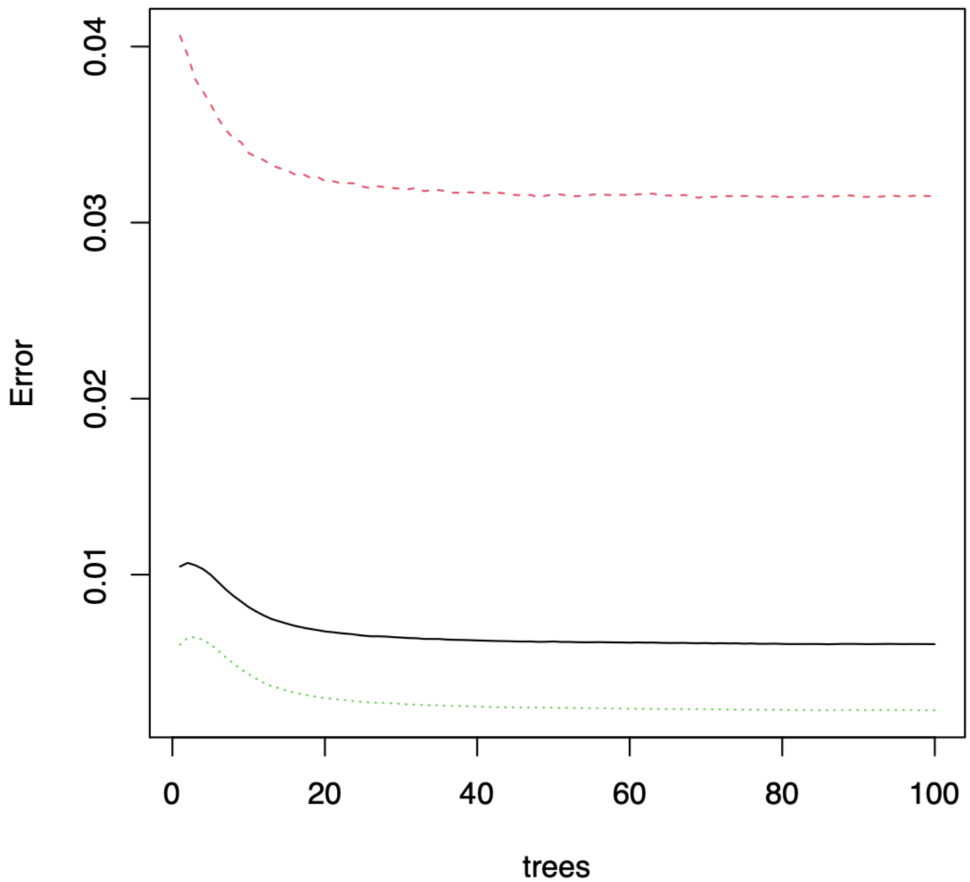


(c) Soy ATE for Upwind Wells



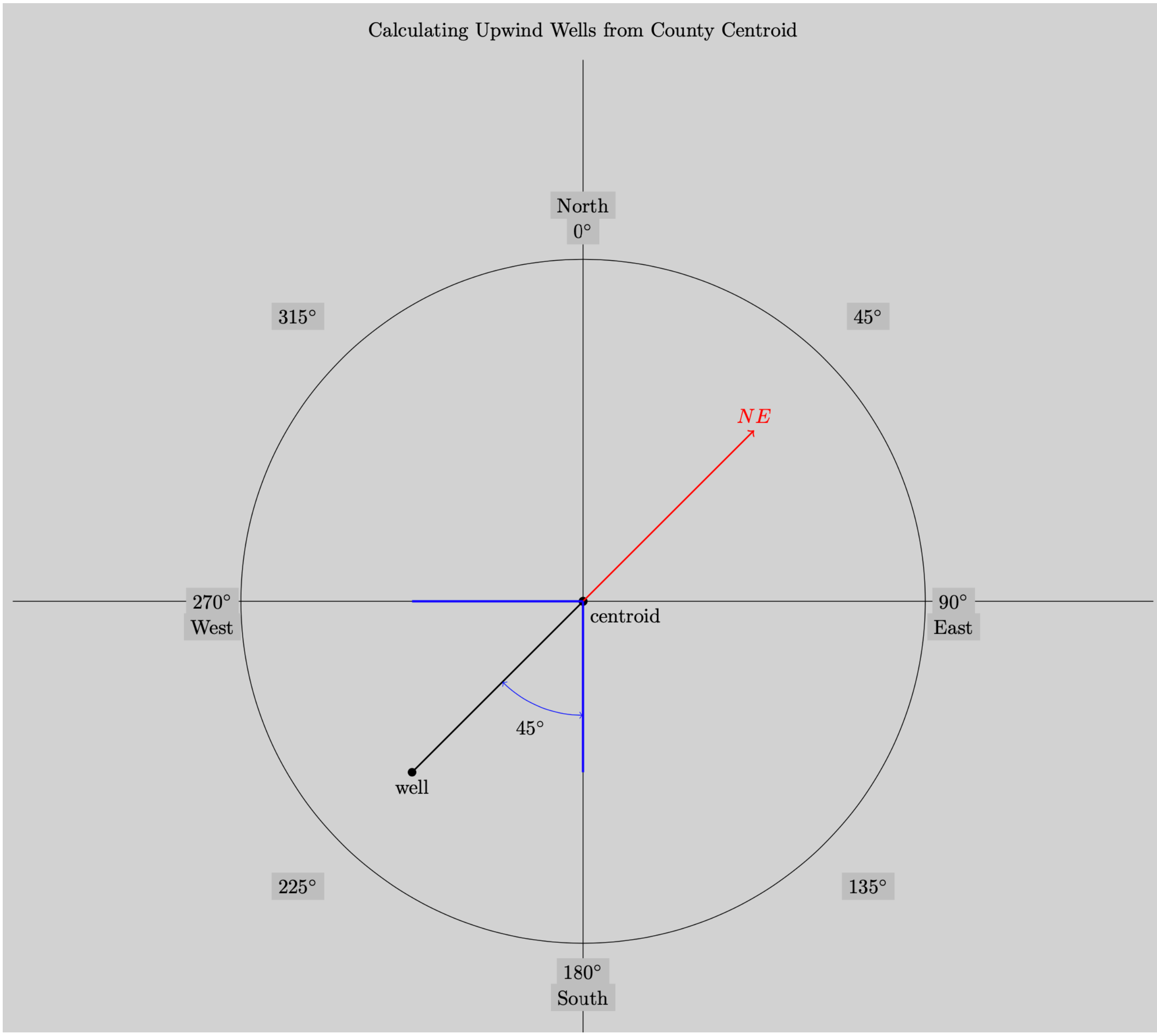
(d) Soy ATE for Downwind Wells

Figure 7: Baseline with COGD ATE Results



(a) Prediction Error for Classification of Unknown/Directional Wells

Figure 8: Random Forest Out of Bag Prediction Error



(a) Visualization of upwind wells treatment area defined

Figure 9: Upwind Wells Treatment Definition