Using Causal Inference in Field Development Optimization: Application to Unconventional Plays

Antoine Bertoncello¹ · Georges Oppenheim² ·
Philippe Cordier³ · Sébastien Gourvénec³ ·
Jean-Philippe Mathieu¹ · Eric Chaput¹ ·
Tobias Kurth⁴

Received: 9 June 2019 / Accepted: 10 December 2019 / Published online: 19 December 2019
© The Author(s) 2019, corrected publication 2020

Abstract In the current era of big data and machine learning, a strong focus exists on prediction and classification. In industrial applications, however, many important questions are not about prediction or classification; rather, they are causal: if I change A, what will happen to B? Traditional regression techniques such as machine learning optimize predictions based on correlations seen in the data and are not robust tools for epidemiologists and biostatisticians when evaluating the efficacy of new treatments or medications using observational data. Therefore, a set of statistical tools have been developed to go beyond correlations and aim to make inferences about causal relationships between variables. The goal of the present work is to apply one of these statistical tools, propensity score matching, in the oil and gas context, which is a novel application of the method. Two case studies are presented, one on proppant type and the other on lateral length, to determine their respective impacts on productivity.

Keywords Causality · Methods · Propensity score · Completion design · Shale plays · Data science

✉ Antoine Bertoncello
antoine.bertoncello@total.com

¹ Total S.A., Avenue Larribau, 64000 Pau, France

² Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, 5 Rue Georges Clemenceau, 91400 Orsay, France

³ Total S.A., 8 Avenue de la Vauve, 91120 Palaiseau, France

⁴ Institute of Public Health, Charité – Universitätsmedizin Berlin Charitéplatz 1, 10117 Berlin, Germany

1 Introduction

One increasingly popular approach is to use machine learning algorithms to predict the future behavior of a system (e.g., failure of a pump, production of a well) based on correlations found in the data. However, a prediction is sometimes not sufficient, and one may want to know what will happen if a specific factor is changed to minimize the failure rate or optimize production. This step is particularly difficult because it requires going beyond correlations and instead inferring a causal relationship between a variable A and outcome B.

The concept of causality is, in general, answered through experiments, such as randomized control trials, experimental design and simulation. However, in industrial settings, experiments are not feasible, and engineers and data scientists often only have access to observational data. In observational data, treatment allocation is not random and is influenced by known and unknown nontreatment factors (Rosenbaum 2017). Therefore, hidden biases that mask the true treatment effect may exist between the treated and control groups. Correcting for these biases is then necessary to estimate the causal connection between the treatment and outcome.

Epidemiologists and biostatisticians face similar problems when attempting to make inferences about the efficacy of a new medical procedure or medication using observational data (Stuart 2010). They have developed a range of statistical tools, such as propensity score matching (Rosenbaum and Rubin 1983), to account for the intrinsic differences between treated and untreated groups to make fair comparisons needed to make causal inference possible. These methods are now successfully used not only in epidemiology but also in econometrics, public health and political sciences. However, their use remains scarce in the energy world.

In this work, two case studies in which propensity score matching is applied to unconventional reservoir development are presented. In the first example, the approach is used to compare the effects of ceramic and sand proppants on productivity in the Marcellus play. In the second example, the approach is used to assess the impact of increasing lateral length on productivity in three gas fields: Haynesville, Marcellus, and Utica plays. In the final section, the benefits of such an approach is compared with answers obtained by machine learning techniques.

2 Presentation of the Marcellus Play

The Marcellus is a vast condensate and dry gas shale play in the eastern United States (Popova 2017). Well productivity is controlled by reservoir properties (pressure, porosity, thickness, source rock maturity) that follow large regional trends (Zagorski et al. 2011). Therefore, wells that are drilled close together (a few kilometers away) tend to be in rock of overall similar quality, whereas wells that are drilled far away from each other (tens to hundreds of kilometers apart) tend to be in rock of different quality.

The two most productive areas are in the northeast and southwest of the basin, where the porosity, thickness, and pressure are most favorable. First, horizontal wells were drilled around 2007, and more than 5000 of them are currently producing. Like

Fig. 1 Average 12-month gas production (MMCF) for wells stimulated with ceramic proppant and sand proppant. Wells with ceramic proppant produce 30% more gas during the first year than wells with sand proppant

any shale play, the reservoir permeability is very low. Horizontal drilling and hydraulic stimulation are required to fracture the rock and increase reservoir connectivity. Hence, the baseline variables impacting the productivity of a well are its location and depth in the basin (proxies for reservoir quality), lateral length, date of drilling, and parameters controlling fracture propagation and conductivity: the volume of injected water, weight of proppant, spacing between fracture stages, and type of proppant used.

The main function of a proppant is to maintain open the fractures generated during hydraulic fracturing (Liang et al. 2017). Operators can choose between two popular proppant types: sand and ceramic. Sand is cheaper to use but cannot withstand high closure stress. Ceramic proppants are designed to resist high closure stress and have higher thermal and chemical stability. They are, however, a more expensive alternative. It is therefore key for operators to accurately estimate the impact of a ceramic proppant on well productivity to see if the better performance, if any, is worth the extra cost.

A direct comparison between wells stimulated with ceramic and sand proppants shows that after 1 year, the first group produces 31% more gas than the second group (1.83 Bcf vs. 1.40 Bcf, Fig. 1). However, this comparison is misleading, because systemic differences in covariates exist between the two groups. A comparison of the two groups' characteristics reveals that well locations and fracture spacing differ between them. Wells stimulated with a ceramic proppant are located preferentially in the northern and southern parts of the play, where the geology is the most favorable (Fig. 2). Tighter fracture spacing (239 ft vs. 272 ft) is also observed in wells with a ceramic proppant. Consequently, the observed 31% increase in 1-year gas production appears to be caused not only by the difference in the proppant type but also by the differences in well locations and fracture spacing.

Fig. 2 In light blue, 5500 wells stimulated with sand proppant; in dark red, 182 wells stimulated with ceramic proppant. Wells with ceramic proppant tend to be located in the northern and southern parts of the field where the productivity is best

If direct comparison is not a solution, using multilinear regression techniques to adjust for the covariates (Gelman and Hill 2006) also leads to biased results. Gas production is regressed using the previously mentioned covariates and treatment variables. The treatment is expressed through a binary variable defined as 0 for a sand proppant and 1 for a ceramic proppant

$$Y = c + \alpha T + \sum_{i=1}^k \beta_i X_i. \quad (1)$$

where Y is the 1-year gas production, T is the treatment variable, and X represents the covariates. The coefficient α associated with the treatment variable is 272 (Table 1), which implies that changing the treatment of a well to a ceramic proppant while keeping the values of X constant increases the production by 272 MMCF, a gain in productivity of 19%. However, the robustness of this approach is questionable. The model poorly fits the data, with $r^2 = 0.49$. The relationship among productivity, covariates, and treatment may indeed not be linear, and interactions between variables, if existing, are not accounted for. Collinearity between variables may also bias causal estimation (Ma 2019). The last section of this paper briefly discusses the limitations of causal inference and the differences to more advanced regression techniques such as machine learning algorithms.

Table 1 Coefficients of linear regression between 1-year gas production and covariates

Variable	Regression coefficient	Standardized regression coefficient
Intercept	74,417	0
Treatment	272	0.04
Latitude	−720	−0.59
Longitude	586	1
Fracture spacing (ft)	−2.2	−0.22
Vertical depth (ft)	0.17	0.14
Completion date	0.14	0.08
Proppant intensity (lb/ft)	0.26	0.12
Lateral length (ft)	0.21	0.36

3 Potential Outcome Framework and Propensity Score Matching for Causal Inference

As seen in the previous example, causal inference from observational data is a difficult task because of the underlying differences in characteristics between treated and control groups. Rubin developed a framework called the “potential outcome framework” or “Neyman–Rubin causal model” to provide a way to quantify causal effects (Sekhon 2008).

In this framework, a data set is composed of individual study units, index i and a binary treatment T ($T_i = 1$ for treated units and $T_i = 0$ for untreated units, also called referent units). In our previous example, wells are the study units, treated units ($T_i = 1$) are wells with a ceramic proppant, and untreated units ($T_i = 0$) are wells with sand as proppant. For each treatment unit, an outcome Y can be observed: Y_i^1 for a treated unit and Y_i^0 for an untreated unit. For the ultimate causal contrast, we would like to know what the outcome for a treated unit would be had this unit not been treated (so-called counterfactual outcome). In practice, however, no single unit can be both treated and untreated at the same time. Only one Y_i^1 or Y_i^0 is observed (the factual outcome); the counterfactual outcome is never observed.

In a randomized experiment, if the treatment is randomly allocated, the causal effect of the treatment on the outcome is simply the difference in the outcomes of treated units to untreated units ($Y_i^1 - Y_i^0$) without further adjustment. The estimated average treatment effect (ATE) across all wells is then defined by Eq. 2

$$ATE = E(Y^1 - Y^0). \quad (2)$$

If a sufficient number of units are randomly assigned to receive treatment, the two treatment groups can be directly compared because their covariate patterns are similar. It can then be assumed that

$$E(Y^1) = E(Y^1 | T = 1) \quad (3)$$

$$E(Y^0) = E(Y^0|T = 0). \quad (4)$$

The ATE can be estimated as the difference in the means of the observed outcomes for each group as follows

$$\text{ATE} = E(Y|T = 1) - E(Y|T = 0). \quad (5)$$

As explained in the first part, we often have to work with observational data where the treatment is not assigned at random, as in a randomized experiment. In this case, Eq. 5 is no longer valid, and the ATE cannot be interpreted as the causal difference in the outcomes between the two treatment groups. A direct comparison between the two groups is not informative because their units are likely to be different (confounded). It is, however, possible to estimate the causal effect of the treatment on the outcome if the covariate patterns between treatment units are balanced. This can, for example, be achieved by using balancing scores, such as the propensity score.

3.1 Propensity Score Estimation

A propensity score is defined as the probability of receiving treatment given the observed covariate pattern, as follows (Rosenbaum and Rubin 1983)

$$\text{PS}(x) = \Pr(T = 1|X = x). \quad (6)$$

As the propensity score is a probability, its values range from 0 to 1.

In a randomized control trial, the probability of a unit receiving the treatment is the same for all units. With a fair coin flip, this probability is 0.5. In an observational study, however, the treatment assignment is not random, and the propensity score is not the same for each unit. In our example, each unit will have a different probability of treatment given its covariate pattern. Some units with a specific combination of covariates X (for instance, longer or deeper wells) may be more likely to receive treatment than others. As a consequence, the treated and untreated units are not directly comparable.

For a dichotomous treatment, the propensity score for each unit can be estimated by using a logistic regression model, where the dependent variable is the treatment status T (0, 1) and the independent variables are the covariates X of the units. Rosenbaum and Rubin (1983) demonstrated that the propensity scores balance the covariates X across treatment and control groups.

3.2 Covariate Selection for Propensity Score Estimation

A critical step for estimating a propensity score is to identify the relevant covariates X to include in the model that estimates the propensity score. In theory, any variable that (1) influences the outcome of interest and (2) is different between the treated and untreated groups should be considered unless it only affects the treatment (Myers et al. 2011). In addition to these general guidelines, Brookhart et al. (2007) recommended

including all variables related to the outcome. In practice, subject matter knowledge guides covariate selection.

3.3 Propensity Score Matching and Effect Estimation

Propensity scores can be used in several ways to balance the covariate patterns of the treatment groups (Andersen and Kurth 2018). An intuitive way is to match units with similar propensity scores (Stuart 2010), as this follows the framework of a (pseudo) randomized experiment with the exception that only the observed covariates incorporated in the propensity score model are balanced in the treated and untreated groups. Each treated unit is matched with a control unit based on similar propensity scores. The matched units are used to build a subsample of all units in which the propensity score distributions of the treated and untreated groups are similar. The similar propensity score distributions imply that the observed covariate pattern is, in expectation, similar between the two groups. The only difference between them is the treatment. Under the assumption of no further major covariate imbalances (for example, of unmeasured covariates), the causal treatment effect in the matched sample is simply the difference in the outcomes between the treated and untreated groups.

It is important to note that matching and comparison are achieved on a subset of the referent population that presents a propensity score distribution similar to the treated population. The matching process hence discards many units from the referent population. Consequently, estimation of the treatment effect through matching is often referred to as the ATT (average treatment effect in the treated), which may differ from the ATE. Methods such as inverse probability of treatment weighting (IPTW) can estimate the ATE directly but are not the focus of this paper.

4 Application 1: Impact of Using Ceramic Proppant Instead of Sand Proppant in the Marcellus Play

4.1 Propensity Score Calculation

Based on the framework previously presented, the group of wells using sand proppant is defined as the control group ($n=5578$), and the group of wells using ceramic proppant is defined as the treated group ($n=189$). The variables used to compute the propensity score are those presented in Table 2 and have been selected based on the abovementioned rationale: the locations of the wells (latitude and longitude), depth of the reservoir, completion date, horizontal length of the wellbore, weight of proppant injected per foot of the horizontal wellbore, and fracture stage spacing. The outcome of interest is the cumulative gas production after 1 year. The propensity score is computed using a logistic regression model.

Table 2 Summary statistics of the Marcellus data set before matching

Average	Ceramic ($n = 189$)	Sand ($n = 5578$)	Standardized difference
1-year gas production (Bcf)	1.88	1.41	
Latitude	40.663	40.768	0.109
Longitude	−78.529	−78.56	0.016
Fracture spacing (ft)	239.2	273	0.552
Vertical depth (ft)	7004	6981	0.025
Completion date	14/11/2013	23/09/2013	0.075
Proppant intensity (lb/ft)	1447	1477	0.054
Lateral length (ft)	5225	5210	0.007
Propensity score	0.043	0.032	0.603

In the Marcellus play, well locations and fracture spacing differ between the two groups

As expected, the propensity score tends to be higher for wells with ceramic proppant (as they are defined as “treated”): the average propensity score for the treatment group is 0.043 compared with 0.032 for the untreated group. The differences in the distributions are shown in Fig. 4. This means that some wells have a higher chance of being stimulated with ceramic proppant (for instance, wells in the north and south of the field), and these wells are overrepresented in the treated group.

4.2 Application of Propensity Score Matching

To reduce the differences in the propensity score distributions, a 1-to-1 matching of treated and referent wells is performed. Each well in the treated group was matched with a well in the control group having the closest propensity score value. The matching process resulted in a new data set of 378 matched wells, 189 wells with ceramic proppant and 189 wells with sand proppant (Fig. 3). In this matched set, the treated and control groups have similar propensity score distributions (Fig. 4, Table 3), and the differences in their covariates are reduced (Table 3).

Comparing the outcomes between the two matched groups provides a more unbiased estimate of the proppant effect: wells with ceramic proppant produce only 10% (95% confidence interval of 5–19%) more than those with sand proppant (1.88 Bcf vs. 1.74 Bcf, Fig. 5), in contrast with the result of 30% (95% confidence interval of 17–45%) when comparing the population in the raw data. A decrease in the estimated proppant effect is expected after propensity score matching (compared with the raw data comparison) because this approach corrects the bias that wells with ceramic proppant tend to be in the most productive areas and have tighter stage spacing in general.

Fig. 3 In dark red, 189 wells stimulated with sand proppant. In light blue, 189 wells stimulated with ceramic proppant. Apart from the treatment type, the two families of wells have the same overall characteristics

Fig. 4 On the left, propensity score distributions for the treated and control groups before matching. On the right, propensity score distributions for the treated and control groups after matching

5 Application 2 with a Continuous Treatment Variable: Impact of Increasing Lateral Length in the Marcellus, Haynesville, and Utica Plays

The previous case study illustrates an application of propensity score matching to assess the effect of two proppant categories on productivity. However, many treatment variables are continuous and not categorical. In this section, a case study is presented to examine the impact of increasing lateral length on well productivity in three US gas plays: Utica, Marcellus, and Haynesville. This lateral length evaluation is important: drilling longer wells offers operational efficiency savings but also potential marginal productivity losses, as friction has a greater impact on longer wells. Knowing in

Table 3 Summary statistics of the Marcellus data set after matching

Average	Ceramic ($n = 189$)	Sand ($n = 189$)	Standardized difference
1-year gas production (Bcf)	1.88	1.73	
Latitude	40.663	40.73	0.05
Longitude	−78.529	−78.23	0.07
Fracture spacing (ft)	239.2	241	0.018
Vertical depth (ft)	7004	6991	0.015
Completion date	14/11/2013	23/12/2013	0.06
Proppant intensity (lb/ft)	1447	1452	0.009
Lateral length (ft)	5225	5343	0.05
Propensity score	0.043	0.043	0.005

Overall, lower standardized differences are observed compared with those in Table 1, indicating that the matching process reduces the differences between the control and treatment groups

Fig. 5 Well production after propensity score matching. Average 12-month gas production (MMCF) for wells stimulated with ceramic proppant and sand proppant. Wells with ceramic proppant produce 10% more gas during the first year than wells with sand proppant

advance how much increasing lateral length will impact well productivity is key to optimizing well profitability.

5.1 Data Set Presentations

This case study focuses on the areas of three plays that produce dry gas. As of early 2019, 833 dry gas wells were drilled in the Utica play in Ohio, Pennsylvania, and West

Virginia. The lateral lengths of the wells range from approximately 3000 ft to more than 12,000 ft. The most common lateral length is 7000 ft. A total of 4176 dry gas wells were drilled in the Haynesville play, located in Texas and Louisiana. The lateral lengths of the wells range from less than 2000 ft to more than 11,000 ft, with the most common length being approximately 4000 ft. In the Marcellus play, information on 4176 dry gas wells is available. The lateral lengths range from less than 1000 ft to more than 15,000 ft. The most common lateral length is approximately 4500 ft.

5.2 Propensity Score Estimation and Matching

The methodology follows the work of Imbens on multivalued treatment (Imbens 2000). For each play, a set of wells with commonly occurring lateral lengths defines the control group: wells from 4800 ft to 5200 ft in Utica, 4000 ft to 4500 ft in Haynesville, and from 4500 ft to 5000 ft in the Marcellus play. Then, for each play, four treated groups are defined, each based on increasing lateral length. A propensity score is calculated for each group/control using a logistic regression model. The variables entered in the model were factors that influence the locations of the wells (latitude and longitude), depth of the reservoir, completion date, weight of proppant injected per foot of the horizontal wellbore, fracture stage spacing, and proppant type. The outcome of interest is 1-year cumulative gas production. Matching is achieved through a nearest neighbor approach of the propensity score values. A summary of the data is presented in Table 4 and displayed in Fig. 6.

5.3 Comparison of Productivity

The production of the wells belonging to each treated group is compared with the production of the wells belonging to the control group in Fig. 6. A result of 120% means that the treated group produces 20% more than the control group. This comparison is performed before and after propensity score matching.

Before matching, the propensity scores significantly differ between the treated and control groups. A linear increase in lateral length does not lead to a linear increase in productivity for the Utica and Haynesville plays. In the Utica play, productivity doubles between 5000 ft and 8000 ft and then plateaus to 9000 ft. In the Haynesville play, productivity increases by 20% between 4000 ft and 5000 ft, decreases slightly at approximately 5500 ft, and increases slightly again at 6000 ft. The productivity increase in the Marcellus play follows a more linear trend, with an 86% increase between 4000 ft and 8000 ft.

After propensity score matching, the calculated trends show a clear linear behavior for all three plays ($r^2 = 0.98$, $r^2 = 0.97$, and $r^2 = 0.98$ for Utica, Haynesville, and Marcellus, respectively). In the Utica play, productivity increases by 87% between 5000 ft and 9000 ft. In the Haynesville play, productivity increases by 92% between 3000 ft and 6000 ft. In the Marcellus play, productivity increases by 68% between 5000 ft and 8000 ft. Assuming a linear behavior, doubling the lateral length leads to an 85% productivity increase in the Marcellus play, a 92% productivity increase in the Haynesville play, and a 96% productivity increase in the Utica play. It is important

Table 4 Summary statistics for the Utica, Haynesville, and Marcellus plays

	Lateral length (ft)	Unmatched data		Matched data	
		Ratio production (%)	Prop. score cont./treat.	Ratio production (%)	Prop. score cont./treat.
Utica group					
Control (<i>n</i> = 70)	4800–5200	100		100	
Treatment 1 (<i>n</i> = 34)	6000–6500	155	0.502/0.302	125	0.502/0.502
Treatment 2 (<i>n</i> = 37)	7000–7500	176	0.653/0.255	151	0.653/0.623
Treatment 3 (<i>n</i> = 33)	8000–8500	212	0.673/0.204	160	0.672/0.631
Treatment 4 (<i>n</i> = 31)	9000–9500	215	0.711/0.174	187	0.710/0.709
Haynesville group					
Control (<i>n</i> = 598)	4000–4500	100		100	
Treatment 1 (<i>n</i> = 79)	3000–3500	75	0.196/0.106	83	0.196/0.197
Treatment 2 (<i>n</i> = 192)	5000–5500	117	0.394/0.194	122	0.394/0.380
Treatment 3 (<i>n</i> = 65)	5500–6000	108	0.433/0.06	149	0.433/0.423
Treatment 4 (<i>n</i> = 70)	6000–6500	120	0.364/0.076	160	0.354/0.338
Marcellus group					
Control (<i>n</i> = 876)	4500–5000	100		100	
Treatment 1 (<i>n</i> = 585)	3500–4000	80	0.410/0.380	91	0.410/0.410
Treatment 2 (<i>n</i> = 431)	6000–6500	129	0.351/0.310	122	0.351/0.352
Treatment 3 (<i>n</i> = 240)	7000–7500	143	0.322/0.185	130	0.322/0.313
Treatment 4 (<i>n</i> = 131)	8000–8500	149	0.241/0.113	153	0.241/0.234

Fig. 6 Left: cross plot between production (1-year cumulative production) and lateral length for the Utica, Haynesville, and Marcellus plays. Right: the orange squares represent the average production increase for the treated groups before matching the data. The black dots represent the average production increase for the treated groups after matching the data. Propensity score matching enables the identification of a clear linear trend between lateral length and productivity that was not visible in the raw data

to note that these trends may not be valid for ranges of lateral length outside of those presented.

The variation between plays could be attributed to differences in the reservoirs. The main developed gas areas in Utica and Haynesville are highly over-pressured, while the central part of the Marcellus play has normal pressure trends (Zagorski et al. 2011).

6 Discussion

6.1 Interpretation of the Propensity Score Matching Results

Propensity score matching was applied to answer two causal questions, one to compare the effects of ceramic and sand proppants on productivity in the Marcellus play and the other to assess the impact of increasing lateral length on productivity in three gas fields. In the examples, we show that the use of propensity score matching leads to different results for raw data and regression. Thus, we believe that the utilization of this approach has substantial advantages in the context of analyzing a large amount of observed data in the oil and gas production industry.

Several aspects must be considered when interpreting the results of our study. First, unlike randomized controlled studies, propensity score methods address imbalances in the covariate patterns between treated and untreated groups only for observed covariates. Unmeasured or unmeasurable information is not considered. However, these methods reflect the real-world setting in which the oil and gas industry has to operate. Second, variable selection for computing the propensity score depends on subject matter knowledge and the assumed causal association between covariates and the treatment-outcome effect. Third, the estimated effects in the propensity score-matched sample are only directly applicable to that sample. Whether the effect is similar in the unmatched sample depends on assumptions about whether the effect is biased or modified by other factors (Kurth et al. 2006).

Fourth, the influence on unmeasured information depends on the underlying causal structure among the treatment, confounding factors, and outcome. As in all causal inference estimates from observational research, we assume that we can correct the influence of measured and unmeasured information by the measured covariates. For our specific example, information on the clay type, a parameter of reservoir quality that varies spatially, is not available. We have assumed that matching on location information allows one to correct for the impact of this parameter on the causal treatment effect.

6.2 Other Propensity Score Utilizations

While matching on the propensity score is likely the most intuitive method for using this tool to balance the measured covariate structures of treated and untreated units, other methods are available. These include regression adjustment with the propensity score and weighted approaches. One weighted approach is the so-called inverse probability of treatment weighting (IPTW) estimator (Austin and Stuart 2015). The concept of this approach differs from propensity score matching, as the causal effect is estimated not only within the distribution of the propensity score in which matched pairs could be identified, but also in the entire distribution. For our example, the causal question

is what would happen if all wells were treated with the ceramic proppant in contrast to none. The estimated effect is different from the propensity score-matched analysis if the treatment effect of those that could not be matched differs from those that could be matched. It means there is a heterogeneity of treatment effects across the propensity score distribution (Kurth et al. 2006).

Regarding propensity score matching, an ideal way of comparing units would be to find a perfect copy of a unit but with the opposite treatment (or not treated), for instance, wells in the same area with the same completion design but with different proppant types. This approach does not work well in practice, especially when the covariate pattern is high-dimensional. Thus, when using only exact matches, only a few units can be matched and many are unmatched, leading to biased results (Rosenbaum and Rubin 1983; Kurth et al. 2006).

6.3 Use of Other Methods

Causal inference is an approach in which the underlying assumed causal structure determines the analysis of the data. Depending on the causal structure, many other methods enable the interpretation of causal effects, such as G-estimation (Murray et al. 2017) and the target trial approach (Hernan and Robins 2016). In contrast to estimating a causal effect, the domain of prediction serves the purpose of using available information at a given time to predict an outcome of interest without making inferences about causality, and thus does not require knowledge about the causal structure of the data. Machine learning algorithms are increasingly used to optimize prediction (Ma 2019). The application of machine learning methods to address a causal question can result in biased estimates when the variables used for prediction are the consequences of at least two causes, also called collider bias (Munafo et al. 2018).

Interesting attempts have been made to combine the fields of causal inference and machine learning (Blakely et al. 2019) to include processing steps before the final (causal) estimation. These methods may be useful for future applications in industry settings.

7 Conclusions

Propensity score analyses are useful and intuitive tools to account for imbalances in the observed covariate patterns between treated and untreated groups, and allow causal questions to be addressed. This approach is increasingly common in biomedical sciences, but it is barely used in an industrial context. This work presents an example of how to apply the methodology for analyzing oil and gas field data. In the first case study, the application of propensity score matching showed that the effect of ceramic proppant on productivity was substantially lower than that initially assumed. In the second case study, the approach deciphered a clear linear relationship between well lateral length and increased productivity. The assessed effects of different completion designs and well architectures could then be integrated with economic and cost data to identify optimum field development plans.

Future work should also focus on developing other methods of causal inference by coupling uncertainty assessment with propensity score matching to characterize the confidence level of the results.

Acknowledgements The authors thank Total Management for their support of this work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Andersen LW, Kurth T (2018) Propensity scores—a brief introduction for resuscitation researchers. *Resuscitation* 125:66–69. <https://doi.org/10.1016/j.resuscitation.2018.02.003>
- Austin PC, Stuart EA (2015) Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* 34:3661–3679. <https://doi.org/10.1002/sim.6607>
- Blakely T, Lynch J, Simons K, Bentley R, Rose S (2019) Reflection on modern methods: when worlds collide—prediction, machine learning and causal inference. *Int J Epidemiol*. <https://doi.org/10.1093/ije/dyz132>
- Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T (2007) Variable selection for propensity score models. *Am J Epidemiol* 163:1149–1156. <https://doi.org/10.1093/aje/kwj149>
- Gelman A, Hill J (2006) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge
- Hernan MA, Robins JM (2016) Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 183:758–764. <https://doi.org/10.1093/aje/kwv254>
- Imbens G (2000) The role of the propensity score in estimating dose–response functions. *Biometrika* 87:706–710. <https://doi.org/10.1093/biomet/87.3.706>
- Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, Robins JM (2006) Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol* 163:262–270. <https://doi.org/10.1093/aje/kwj047>
- Liang F, Sayed M, Al-Muntasheri GA, Chang FF, Li L (2017) A comprehensive review on proppant technologies. *Petroleum* 2:26–39. <https://doi.org/10.1016/j.petlm.2015.11.001>
- Ma YZ (2019) Quantitative geosciences: data analytics, geostatistics, reservoir characterization and modeling, English edn. Springer, Cham
- Munafo MR, Tilling K, Taylor AE, Evans DM, Davey Smith GD (2018) Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* 47:226–235. <https://doi.org/10.1093/ije/dyx206>
- Murray EJ, Robins JM, Seage GR, Freedberg KA, Hernan MA (2017) A comparison of agent-based models and the parametric G-formula for causal inference. *Am J Epidemiol* 186:131–142. <https://doi.org/10.1093/aje/kwx091>
- Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, Joffe MM, Glynn RJ (2011) Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol* 174:1213–1222. <https://doi.org/10.1093/aje/kwr364>
- Popova O (2017) Marcellus shale play. U.S. Energy Information Administration. U.S. Department of Energy, Washington, DC
- Rosenbaum P (2017) Observation and experiment. Harvard University Press, Cambridge
- Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55. <https://doi.org/10.1093/biomet/70.1.41>

- Sekhon JS (2008) The Oxford handbook of political methodology. Oxford University Press, Oxford
- Stuart EA (2010) Matching methods for causal inference: a review and a look forward. *Stat Sci* 25:1–21. <https://doi.org/10.1214/09-sts313>
- Zagorski W, Bowman D, Emery M, Wrightstone G (2011) An overview of some key factors controlling well productivity in core areas of the appalachian basin marcellus shale play. In: AAPG (American association of petroleum geologists) 2011 annual convention. AAPG, Houston, Texas, US, pp 1–24