Ariel Azria

Lea Wu

Caleb Dimenstein

John Fitzgerald

**Final Project - Stage 2**

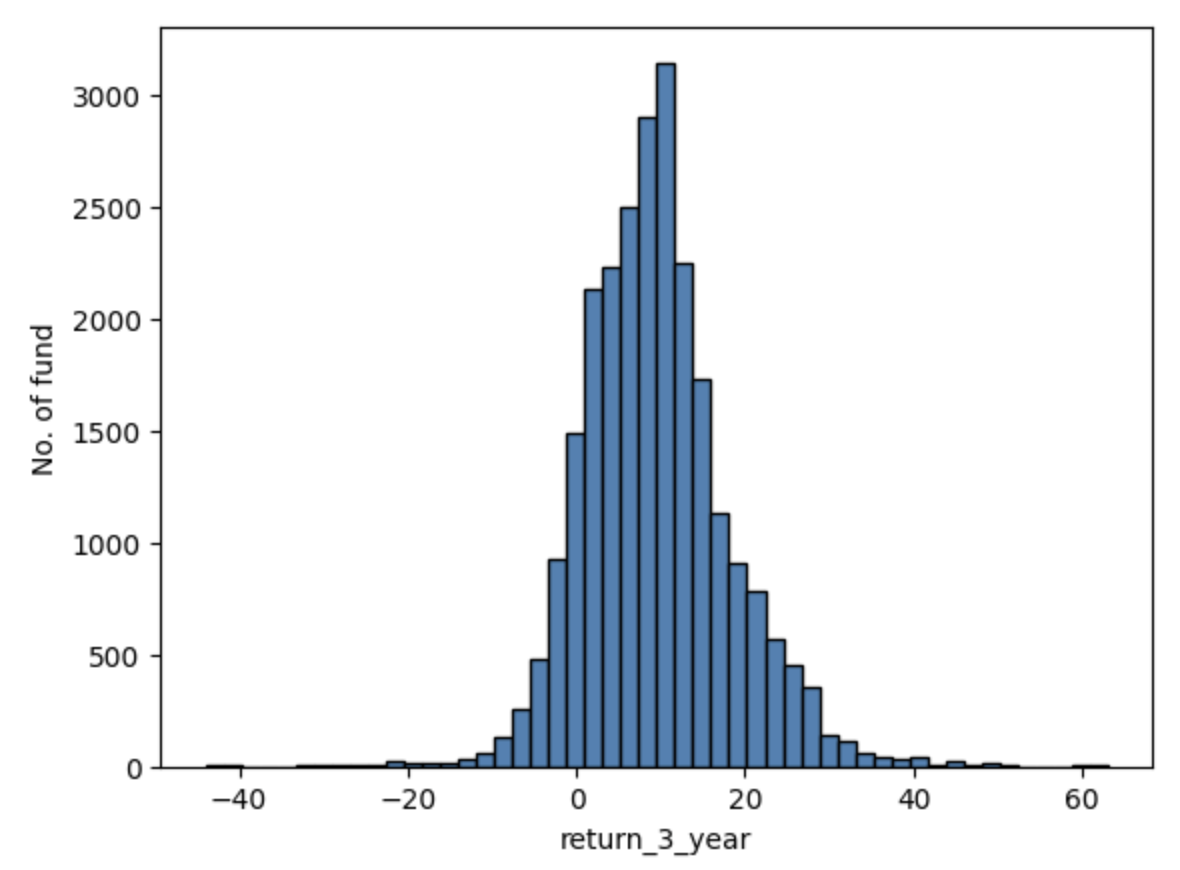
*Note: instead of the literature review, Professor Pamuksuz asked our group to focus on feature engineering. Thus, we've replaced the literature review portion of this assignment with a summary of work completed for feature engineering.*

**Feature Engineering**

1. **Rename, merge, and drop columns**

Our ESG fund data spanned from 2020 to 2022. Among 118 features, our primary variable of interest was the three-year return rate. To enhance the dataset's utility for our analysis, we undertook a thorough data-cleaning process. This process involved merging multiple datasets for continuity and renaming columns to ensure clarity and consistency. Furthermore, we eliminated features that either presented little to no predictive value or were impractical for model inclusion, such as the 'Asset Manager Name'. We also dropped data points with missing values in the three-year return rate.

Next, we assessed the distribution of our target variable using the 'Fitter' package and observed that it closely approximates a normal distribution:



1. **Feature transforming**
   1. **Categorical & datetime data**
      1. **ESG grade**

The data divided each fund into 5 grades (A, B, C, D, F) regarding compliance in various ESG industries. After initial modeling, we found that letter grades explained much of a fund's return. Hence, we transformed the grade into binary variables to record whether the fund had a good grade (i.e. in grade A or B), reducing the number of features while maintaining the ability to explain the causal effect.

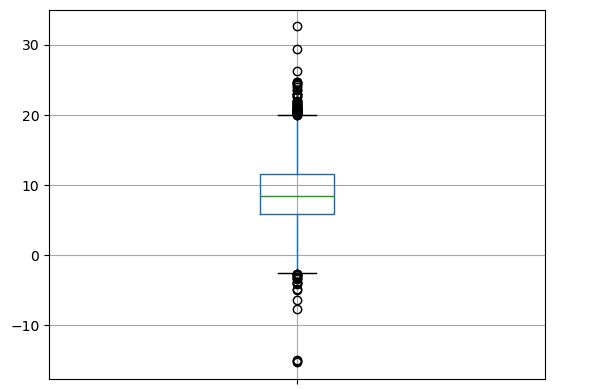
* + 1. **Inception date**

As ESG funds are a relatively new topic, the inception date is likely to affect the return rate. We transformed this date data into day counts until the data record date.

* 1. **Difference-in-difference data preparation**

The Difference-in-Differences (DiD) approach is a particularly fitting method for our topic. It allows us to explain the causal inference by investigating the impact of specific events such as sustainability status alterations and grading scores. To prepare our dataset for this technique, we constructed 'treatment' columns corresponding to each feature of interest. We also prepared ‘after’ columns to indicate whether an observation occurred after the events.

1. **Outlier identifying**

Financial performance: 3 years return rate

In our dataset, although the majority of funds exhibited a positive return rate, we identified outliers within our target variable. To enhance the robustness of our analysis and maintain data integrity, we excised these extreme values. This resulted in the exclusion of 61 observations from the original total of 9,929.

1. **Feature selection**
   1. **Correlation**

Our analysis revealed a significant degree of correlation between several features within the dataset. For example, the proportion of investment in the prison industry and the prison-free-grade of the fund are highly correlated. Weapon-related metrics and prison-related features are also highly correlated. Given the lack of access to the detailed grading rubric, we performed feature selection to avoid collinearity by considering correlation, variance-inflation-factor, and domain knowledge. Additionally, we plan to implement instrumental variables in our forthcoming research to address collinearity.

* 1. **Forward feature selection & Recursive feature elimination**

To reduce the number of features, we implemented 2 feature selection techniques. Through this process, the efficacy of each selection technique varied across different models. Consequently, we will select the most appropriate feature selection method based on the model employed. This tailored method ensures that the optimal number of features is retained to enhance each model's performance, thereby ensuring a balance between model complexity and predictive accuracy.

**Explainable AI Methods**

**LIME (Local Interpretable Model-Agnostic Explanations)**

**Overview**

Ribeiro, Singh, and Guestrin first proposed LIME in 2016 as a solution to explain predictions from black box models.[[1]](#footnote-0) The algorithm fits a surrogate model to the output of a black box model, allowing researchers to explain the black box model's output.[[2]](#footnote-1)

**Fundamental Principles**

LIME's creators posit that by understanding algorithms' decisions, humans will place more trust in machine learning techniques.[[3]](#footnote-2) To enable transparent machine learning models, LIME first generates a synthetic dataset set and obtains the corresponding predictions using the black box model researchers would like to explain. It then trains an interpretable model, such as linear regression, on this data set and employs feature selection to determine an interpretable representation of the data.[[4]](#footnote-3)

**Recent Advancements**

In 2021, researchers created a Bayesian extension of the base LIME model, which they called "BayLime."[[5]](#footnote-4) BayLime addresses some of LIME's fallbacks discussed in the "Limitations" section, allowing LIME to be used in a clinical setting.[[6]](#footnote-5)

**Applications**

LIME can be applied in various settings, such as detecting illness,[[7]](#footnote-6) modeling air quality,[[8]](#footnote-7) and creating a quantitative framework to predict foreign direct investments into Western European countries.[[9]](#footnote-8) For these authors' purpose, LIME could also be applied to explain which factors impact financial fund returns the most.

**Strengths**

LIME was one of the first locally interpretable models; in addition to shedding light on opaque models, it allows users to set the number of features they would like to include in the interpretable model.[[10]](#footnote-9)

**Limitations**

Two of LIME's greatest limitations are a "lack of consistency in repeated explanations of a single prediction"[[11]](#footnote-10) and a dearth of "robustness to kernel settings."[[12]](#footnote-11) For this reason, researchers posit that doctors should be weary of using LIME in clinical settings and propose BayLIME to mitigate these considerations.[[13]](#footnote-12)

**Explainable Boosting Machines (EBMs)**

**Overview**

Explainable Boosting Machines (EBMs) are a form of generalized additive model (GAM) that incorporates machine learning techniques such as bagging and boosting.[[14]](#footnote-13) It is a glassbox model, meaning that it is designed to allow users to easily interpret and explain its results.[[15]](#footnote-14)

**Fundamental Principles**

EBM's creators wanted users to be able to create models that quantified "the impact of each predictor."[[16]](#footnote-15) To do so, they studied GAMs, which are more accurate than generalized linear models, and combined them with bagged trees. They then released a package, interpretML, that allows users to easily implement EBMs and visualize each predictor's impact on the response variable.[[17]](#footnote-16)

**Recent Advancements**

At the end of 2023, researchers proposed a method to adapt EBMs to model scientific image data.[[18]](#footnote-17) After transforming the algorithm into a function adapted to images, researchers found their approach rivaled other algorithms for accuracy while maintaining intuitive explanations.[[19]](#footnote-18) Researchers remain heavily focused on EBMs, as many fields, such as financial services and healthcare, require explainable models.

**Applications**

EBMs are not applicable in every instance. However, they should be favored over black-box models when users have access to the training data and are debugging, retraining, and improving a model's accuracy.[[20]](#footnote-19) EBMs have been used to predict traffic crashes near work-zones,[[21]](#footnote-20) predict geological slope failure,[[22]](#footnote-21) and have been applied in fields as diverse as education, manufacturing, health, and security.[[23]](#footnote-22)

**Strengths**

EBM's results' accuracy is often compared with Random Forests and XGBoost. They also have light memory usage on small to medium-sized datasets. These features make it a very attractive model to deploy in production.[[24]](#footnote-23)

**Limitations**

EBMS should not be used when a specific black-box model is required, nor when there is a need to understand a complex end-to-end pipeline.[[25]](#footnote-24) In addition, because of the additive component of GAMs, they can be quite computationally intensive on large datasets.[[26]](#footnote-25)

**Instrumental Variables**

**Overview**

An instrumental variable (IV) is a variable that affects the explanatory variables but only affects the response variable through its impact on the explanatory variables.[[27]](#footnote-26) Suppose a student hypothesizes that attending college is positively correlated with future wages. The regression equation *wages = ꞵ0 + ꞵ1 \* college + Ɛ* does not consider individual ability. People who learn more easily are more likely to go to college and, separately, might earn higher wages. Ability level impacts both the decision to attend university and wages – it is considered a “cofounder.”

Researchers would add an instrumental variable to their analysis to isolate the effect of ability on wages. This variable would impact going to college but not future wages. One such variable is the proximity between one’s pre-college residence and a college, as proximity to a college is not likely correlated with an individual’s ability but does influence whether they attend university.[[28]](#footnote-27) **Fundamental Principles**

According to Pearl (2000), variable Z is an instrumental variable “relative to the pair (X, Y)) if (i) Z is independent of all variables (including error terms) that have an influence on Y that is not mediated by X and (ii) Z is not independent of X."[[29]](#footnote-28)

**Recent Advancements**

Instrumental variables were first researched and proposed in the mid-twentieth century.[[30]](#footnote-29) They are widely used in econometrics and have recently been used to analyze panel data.[[31]](#footnote-30) In 2023, researchers released a widely-used Stata package, further facilitating IV’s use in econometrics and statistical research.[[32]](#footnote-31)

**Applications**

Instrumental variables are typically used to uncover causal relationships in the social sciences and are specifically useful when analyzing observational data or quasi-experiements.[[33]](#footnote-32)

To determine whether to use an IV in one’s analysis, one should reflect on the amount of unmeasured confounding there is between variables.[[34]](#footnote-33)

**Strengths & Limitations**

IVs are incredibly useful when attempting to reduce bias when evaluating causal relationships between variables.[[35]](#footnote-34) However, as seen in the example above, finding a valid instrumental variable that explains enough exogenous variation in data is incredibly challenging.[[36]](#footnote-35)

**Propensity Score Matching**

**Overview**

Propensity Score Matching (PSM) is a causal inference method applied to observational data without a controlled study. It is particularly useful in analyzing heterogeneous data, allowing users to isolate the treatment variable and gain insight into causal relationships. PSM can be applied to various contexts, such as evaluating the performance of ESG funds based on the weights of different factors.

**Fundamental Principles**

1. PSM aims to find similar observations within data, calculate propensity scores, and adjust the variables used in the model for optimal distribution within the chosen treatment and control groups. This process isolates the treatment variable from the study.
2. To apply PSM, the data must be divided into two groups with similar distributions based on the propensity score calculations. In the ESG fund example, similar funds can be identified and divided into categories based on their ESG factor weights, such as Fossil Free, Gun Free, Gender Equality, and Prison Free funds.
3. The treatment refers to the specific focus of the study, such as the weights of ESG factors in a fund, while the outcome variable of interest is the impact of that treatment on the results – in the case of ESG funds, their performance over a given period (e.g., 5 years).
4. PSM can be applied across a broad range of ESG factors or narrowed down to individual categories to gain more in-depth insights into the causal effects of specific factors on fund performance.

**Recent Advancements**

Propensity score matching estimators were first developed by Rosenbaum and Rubin in 1983.[[37]](#footnote-36) Recent advancements and applications include those in clinical research (ie, Radiology, Nephrology, and Cardiovascular Research)[[38]](#footnote-37),[[39]](#footnote-38),[[40]](#footnote-39) where an individual is assigned to a group of interest based on their propensity scores.

**Applications**

Propensity score matching is best used when we want to know the effect of something without having random assignment. We would first have divided the treatment group and the control group, but instead of calculating the difference in effect between the treatment and control group we would calculate the propensity score (or probability) of being in the Treatment group based on the other confounding factors (everything else other than the variable being studied). This allows us to then compare the averages in the outcomes between the similar groups of confounding variables and correctly see the difference in outcomes between the treatment and control groups.

**Strengths & Limitations**

This will be best used when the distributions and covariates of the propensity scores of the treatment and new control group are similar. If they are very different, the comparisons might have too much bias. Also, if the distribution of propensity scores across the treatment and original control group are even, then this model might not be necessary. It also does not help when omitted variables affect both the outcome and whether the observation was treated (ie. unobserved confounders)

**Difference in Difference**

**Overview**

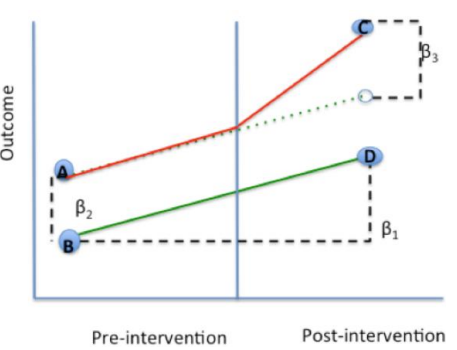
The difference-in-difference (DID) technique originated in the field of econometrics, but the

logic underlying the technique has been used as early as the 1850s by John Snow and is called the ‘controlled before-and-after study’ in some social sciences.[[41]](#footnote-40)

**Fundamental Principles**

𝑦 = 𝛽0 + 𝛽1𝑡𝑖𝑚𝑒 + 𝛽2𝑡𝑟𝑒𝑎𝑡𝑒𝑑 + 𝛽3𝑡𝑖𝑚𝑒 ∗ 𝑡𝑟𝑒𝑎𝑡𝑒𝑑 + 𝜀

Where 𝛽0 is the baseline average, 𝛽1 is the time trend in control group,𝛽2 is the difference between two groups pre-intervention,and 𝛽3 is the difference in changes over time[[42]](#footnote-41)



**Recent Advancements**

Recent advancements in DiD methods address cases with more than two time periods and different treatment timings. These new approaches isolate clean comparisons between treated and not-yet-treated groups, allowing for better estimations under treatment effect heterogeneity.

The possibility of parallel trends assumption being violated has led to the development of alternative methods. These methods ensure better power in pre-trend tests and remain valid under certain types of parallel trends violations.[[43]](#footnote-42)

**Applications**

Difference in Difference is used extensively in field experiments across disciplines.

It has, for example, been used extensively to study the impacts of various education reforms around the world. Examples include reforms of compulsory schooling and tracking (e.g., Meghir and Palme (2005), Pekkala Kerr, Pekkarinen, and Uusitalo (2013), Meghir, Palme, and Simeonova (2018)), education priority zones for disadvantaged schools (e.g., Bénabou, Kramarz, and Prost (2009)), subsidized child care (e.g., Havnes and Mogstad (2011)), and paid parental leave (e.g., Danzer and Lavy (2018)).[[44]](#footnote-43) It can also be used across fields such as medical policy research[[45]](#footnote-44) to understand the effects overtime of such policies.

**Strengths**

* Some of the strengths of the difference-in-difference model include:
  + Intuitive interpretation
  + Can use either individual or group-level data
  + Comparison groups can start at different levels of the outcome

**Limitations**

* Requires baseline data & a non-intervention group
* Must have two or more periods analyzed and thus cannot be point-in-time data
* Cannot use if the composition of groups pre/post change are not stable

**Bibliography**

[Alahmadi, Raed, Hamad Almujibah, Saleh Alotaibi, Ali E. A. Elshekh, Mohammad Alsharif, and Mudthir Bakri. “Explainable Boosting Machine: A Contemporary Glass-Box Model to Analyze Work Zone-Related Road Traffic Crashes.” *Safety* 9, no. 4 (December 2023): 83.](https://www.zotero.org/google-docs/?6LVFxX) https://doi.org/10.3390/safety9040083[.](https://www.zotero.org/google-docs/?6LVFxX)

Baek, S., Park, S. H., Won, E., Park, Y. R., & Kim, H. J. (2015). Propensity Score Matching: A Conceptual Review for Radiology Researchers. Korean Journal of Radiology, 16(2), 286–296. https://doi.org/10.3348/kjr.2015.16.2.286

[Baiocchi, Michael, Jing Cheng, and Dylan S. Small. “Tutorial in Biostatistics: Instrumental Variable Methods for Causal Inference\*.” *Statistics in Medicine* 33, no. 13 (June 15, 2014): 2297–2340.](https://www.zotero.org/google-docs/?6LVFxX) https://doi.org/10.1002/sim.6128[.](https://www.zotero.org/google-docs/?6LVFxX)

Barnieh, L., James, M. T., Zhang, J., & Hemmelgarn, B. R. (2011). Propensity score methods and their application in nephrology research. Journal of Nephrology, 24(3), 256–262. https://doi.org/10.5301/jn.2011.6429

[Barr Kumarakulasinghe, Nesaretnam, Tobias Blomberg, Jintai Liu, Alexandra Saraiva Leao, and Panagiotis Papapetrou. “Evaluating Local Interpretable Model-Agnostic Explanations on Clinical Machine Learning Classification Models.” In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 7–12, 2020. https://doi.org/10.1109/CBMS49503.2020.00009.](https://www.zotero.org/google-docs/?6LVFxX)

[Baum, Christopher, Mark Schaffer, and Steven Stillman. “IVREG2: Stata Module for Extended Instrumental Variables/2SLS and GMM Estimation.” Statistical Software Components, July 30, 2023. https://econpapers.repec.org/software/bocbocode/S425401.htm.](https://www.zotero.org/google-docs/?6LVFxX)

[C3 AI. “LIME: Local Interpretable Model-Agnostic Explanations.” Accessed February 12, 2024. https://c3.ai/glossary/data-science/lime-local-interpretable-model-agnostic-explanations/.](https://www.zotero.org/google-docs/?6LVFxX)

[Caruna, Rich. “InterpretML: Explainable Boosting Machines (EBMs) by Microsoft Research.” n.d.](https://www.zotero.org/google-docs/?6LVFxX) https://people.orie.cornell.edu/mru8/orie4741/lectures/Tutorial4MadeleineUdellClass\_2020Dec08\_RichCaruana\_IntelligibleMLInterpretML\_EBMs\_75mins.pdf[.](https://www.zotero.org/google-docs/?6LVFxX)

Columbia University. (2022, June 7). Difference-in-Difference Estimation | Columbia University Mailman School of Public Health. Www.publichealth.columbia.edu. https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation

Deb, S., Austin, P. C., Tu, J. V., Ko, D. T., Mazer, C. D., Kiss, A., & Fremes, S. E. (2016). A Review of Propensity-Score Methods and Their Use in Cardiovascular Research. Canadian Journal of Cardiology, 32(2), 259–265. https://doi.org/10.1016/j.cjca.2015.05.015

Griffin, B. A., Schuler, M. S., Stuart, E. A., Patrick, S., McNeer, E., Smart, R., Powell, D., Stei, B. D., Schell, T. L., & Pacula, R. L. (2021). Moving beyond the classic difference-in-differences model: a simulation study comparing statistical methods for estimating effectiveness of state-level policies. BMC Medical Research Methodology, 21(1). https://doi.org/10.1186/s12874-021-01471-y

[Guler, Hakan, Derya Avcı, Mustafa Ulaş, and Tülay Omma. “Performance Comparison of Machine Learning Models Powered by SHAP and LIME Based Explainability Techniques on Diabetes Dataset.” SSRN Scholarly Paper. Rochester, NY, February 6, 2024. https://doi.org/10.2139/ssrn.4713039.](https://www.zotero.org/google-docs/?6LVFxX)

[Karmaker, Shamal Chandra, Munim Kumar Barai, Kanchan Kumar Sen, and Bidyut Baran Saha. “Effects of Remittances on Renewable Energy Consumption: Evidence from Instrumental Variable Estimation with Panel Data.” *Utilities Policy* 83 (August 1, 2023): 101614. https://doi.org/10.1016/j.jup.2023.101614.](https://www.zotero.org/google-docs/?6LVFxX)

[Lou, Yin, Rich Caruana, and Johannes Gehrke. “Intelligible Models for Classification and Regression.” In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 150–58. Beijing China: ACM, 2012. https://doi.org/10.1145/2339530.2339556.](https://www.zotero.org/google-docs/?6LVFxX)

[Maxwell, Aaron E., Maneesh Sharma, and Kurt A. Donaldson. “Explainable Boosting Machines for Slope Failure Spatial Predictive Modeling.” *Remote Sensing* 13, no. 24 (January 2021): 4991. https://doi.org/10.3390/rs13244991.](https://www.zotero.org/google-docs/?6LVFxX)

[Mora, Ricardo. “Instrumental Variables in Econometrics II.” *Department of Economics Universidad Carlos III de Madrid Master in Industrial Organization and Markets*, n.d.](https://www.zotero.org/google-docs/?6LVFxX)

[Pearl, Judea. *Causality: Models, Reasoning, and Inference*. Cambridge, U.K. ; New York: Cambridge University Press, 2000.](https://www.zotero.org/google-docs/?6LVFxX)

[Pokropek, Artur. “Introduction to Instrumental Variables and Their Application to Large-Scale Assessment Data.” *Large-Scale Assessments in Education* 4, no. 1 (February 24, 2016): 4.](https://www.zotero.org/google-docs/?6LVFxX) https://doi.org/10.1186/s40536-016-0018-2[.](https://www.zotero.org/google-docs/?6LVFxX)

ROSENBAUM, P. R., & RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41–55.

Roth, J., Sant’Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What’s trending in difference-in-differences? A synthesis of the recent econometrics literature. Journal of Econometrics. https://doi.org/10.1016/j.jeconom.2023.03.008

[Sargan, J. D. “The Estimation of Economic Relationships Using Instrumental Variables.” *Econometrica* 26, no. 3 (1958): 393–415.](https://www.zotero.org/google-docs/?6LVFxX) https://doi.org/10.2307/1907619[.](https://www.zotero.org/google-docs/?6LVFxX)

Schwerdt, G., & Woessmann, L. (2020, January 1). Chapter 1 - Empirical methods in the economics of education (S. Bradley & C. Green, Eds.). ScienceDirect; Academic Press. https://www.sciencedirect.com/science/article/abs/pii/B978012815391800001X.

[Schug, Daniel, Sai Yerramreddy, Rich Caruana, Craig Greenberg, and Justyna Zwolak. “Extending Explainable Boosting Machines to Scientific Image Data.” *NIST*, November 30, 2023. https://www.nist.gov/publications/extending-explainable-boosting-machines-scientific-image-data.](https://www.zotero.org/google-docs/?6LVFxX)

[Singh, Devesh. “Foreign Direct Investment and Local Interpretable Model-Agnostic Explanations: A Rational Framework for FDI Decision Making.” *Journal of Economics, Finance and Administrative Science* ahead-of-print, no. ahead-of-print (January 1, 2023). https://doi.org/10.1108/JEFAS-05-2021-0069.](https://www.zotero.org/google-docs/?6LVFxX)

[Tasioulis, Thomas, and Kostas Karatzas. “Reviewing Explainable Artificial Intelligence Towards Better Air Quality Modelling.” In *Advances and New Trends in Environmental Informatics 2023*, edited by Volker Wohlgemuth, Dieter Kranzlmüller, and Maximilian Höb, 3–19. Progress in IS. Cham: Springer Nature Switzerland, 2024.](https://www.zotero.org/google-docs/?6LVFxX) https://doi.org/10.1007/978-3-031-46902-2\_1[.](https://www.zotero.org/google-docs/?6LVFxX)

Torres-Reyna, O. (2015). Differences-in-Differences (using Stata). https://www.princeton.edu/~otorres/DID101.pdf

[“‘Why Should I Trust You?’ | Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.” Accessed February 12, 2024. https://dl.acm.org/doi/abs/10.1145/2939672.2939778.](https://www.zotero.org/google-docs/?6LVFxX)

[Zafar, Muhammad Rehman, and Naimul Khan. “Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability.” *Machine Learning and Knowledge Extraction* 3, no. 3 (September 2021): 525–41. https://doi.org/10.3390/make3030027.](https://www.zotero.org/google-docs/?6LVFxX)

[Zhao, Xingyu, Wei Huang, Xiaowei Huang, Valentin Robu, and David Flynn. “BayLIME: Bayesian Local Interpretable Model-Agnostic Explanations.” In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, 887–96. PMLR, 2021.](https://www.zotero.org/google-docs/?6LVFxX) https://proceedings.mlr.press/v161/zhao21a.html[.](https://www.zotero.org/google-docs/?6LVFxX)

1. [“‘Why Should I Trust You?’ | Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,” accessed February 12, 2024, https://dl.acm.org/doi/abs/10.1145/2939672.2939778.](https://www.zotero.org/google-docs/?Vm5Omb) [↑](#footnote-ref-0)
2. [“LIME: Local Interpretable Model-Agnostic Explanations,” C3 AI, accessed February 12, 2024, https://c3.ai/glossary/data-science/lime-local-interpretable-model-agnostic-explanations/.](https://www.zotero.org/google-docs/?H0Wlzf) [↑](#footnote-ref-1)
3. [“‘Why Should I Trust You?’ | Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.”](https://www.zotero.org/google-docs/?YVpONU) [↑](#footnote-ref-2)
4. [Muhammad Rehman Zafar and Naimul Khan, “Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability,” *Machine Learning and Knowledge Extraction* 3, no. 3 (September 2021): 525–41, https://doi.org/10.3390/make3030027.](https://www.zotero.org/google-docs/?PQd2XK) [↑](#footnote-ref-3)
5. [Xingyu Zhao et al., “BayLIME: Bayesian Local Interpretable Model-Agnostic Explanations,” in *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence* (Uncertainty in Artificial Intelligence, PMLR, 2021), 887–96, https://proceedings.mlr.press/v161/zhao21a.html.](https://www.zotero.org/google-docs/?Gxq2Px) [↑](#footnote-ref-4)
6. [Nesaretnam Barr Kumarakulasinghe et al., “Evaluating Local Interpretable Model-Agnostic Explanations on Clinical Machine Learning Classification Models,” in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 2020, 7–12, https://doi.org/10.1109/CBMS49503.2020.00009.](https://www.zotero.org/google-docs/?ZShGgp) [↑](#footnote-ref-5)
7. [Hakan Guler et al., “Performance Comparison of Machine Learning Models Powered by SHAP and LIME Based Explainability Techniques on Diabetes Dataset,” SSRN Scholarly Paper (Rochester, NY, February 6, 2024), https://doi.org/10.2139/ssrn.4713039.](https://www.zotero.org/google-docs/?zVZp6e) [↑](#footnote-ref-6)
8. [Thomas Tasioulis and Kostas Karatzas, “Reviewing Explainable Artificial Intelligence Towards Better Air Quality Modelling,” in *Advances and New Trends in Environmental Informatics 2023*, ed. Volker Wohlgemuth, Dieter Kranzlmüller, and Maximilian Höb, Progress in IS (Cham: Springer Nature Switzerland, 2024), 3–19, https://doi.org/10.1007/978-3-031-46902-2\_1.](https://www.zotero.org/google-docs/?IzgEAd) [↑](#footnote-ref-7)
9. [Devesh Singh, “Foreign Direct Investment and Local Interpretable Model-Agnostic Explanations: A Rational Framework for FDI Decision Making,” *Journal of Economics, Finance and Administrative Science* ahead-of-print, no. ahead-of-print (January 1, 2023), https://doi.org/10.1108/JEFAS-05-2021-0069.](https://www.zotero.org/google-docs/?wqQLbg) [↑](#footnote-ref-8)
10. [Zafar and Khan, “Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability.”](https://www.zotero.org/google-docs/?rhd4tY) [↑](#footnote-ref-9)
11. [“‘Why Should I Trust You?’ | Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.”](https://www.zotero.org/google-docs/?5uK4Tl) [↑](#footnote-ref-10)
12. [“‘Why Should I Trust You?’ | Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.”](https://www.zotero.org/google-docs/?XzK0zr) [↑](#footnote-ref-11)
13. [Zhao et al., “BayLIME.”](https://www.zotero.org/google-docs/?xoBFit) [↑](#footnote-ref-12)
14. [Rich Caruna, “InterpretML: Explainable Boosting Machines (EBMs) by Microsoft Research,” https://people.orie.cornell.edu/mru8/orie4741/lectures/Tutorial4MadeleineUdellClass\_2020Dec08\_RichCaruana\_IntelligibleMLInterpretML\_EBMs\_75mins.pdf.](https://www.zotero.org/google-docs/?TPztEM) [↑](#footnote-ref-13)
15. [Caruna.](https://www.zotero.org/google-docs/?SAj6p9) [↑](#footnote-ref-14)
16. [Yin Lou, Rich Caruana, and Johannes Gehrke, “Intelligible Models for Classification and Regression,” in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD ’12: The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing China: ACM, 2012), 150–58, https://doi.org/10.1145/2339530.2339556.](https://www.zotero.org/google-docs/?ZszB4U) [↑](#footnote-ref-15)
17. [Caruna, “InterpretML: Explainable Boosting Machines (EBMs) by Microsoft Research.”](https://www.zotero.org/google-docs/?qN1OlC) [↑](#footnote-ref-16)
18. [Daniel Schug et al., “Extending Explainable Boosting Machines to Scientific Image Data,” *NIST*, November 30, 2023, https://www.nist.gov/publications/extending-explainable-boosting-machines-scientific-image-data.](https://www.zotero.org/google-docs/?bywqVW) [↑](#footnote-ref-17)
19. [Schug et al.](https://www.zotero.org/google-docs/?J57fwt) [↑](#footnote-ref-18)
20. [Caruna, “InterpretML: Explainable Boosting Machines (EBMs) by Microsoft Research.”](https://www.zotero.org/google-docs/?O7VnAe) [↑](#footnote-ref-19)
21. [Raed Alahmadi et al., “Explainable Boosting Machine: A Contemporary Glass-Box Model to Analyze Work Zone-Related Road Traffic Crashes,” *Safety* 9, no. 4 (December 2023): 83, https://doi.org/10.3390/safety9040083.](https://www.zotero.org/google-docs/?Rpzjp6) [↑](#footnote-ref-20)
22. [Aaron E. Maxwell, Maneesh Sharma, and Kurt A. Donaldson, “Explainable Boosting Machines for Slope Failure Spatial Predictive Modeling,” *Remote Sensing* 13, no. 24 (January 2021): 4991, https://doi.org/10.3390/rs13244991.](https://www.zotero.org/google-docs/?16xpQ6) [↑](#footnote-ref-21)
23. [Alahmadi et al., “Explainable Boosting Machine.”](https://www.zotero.org/google-docs/?38dqCY) [↑](#footnote-ref-22)
24. [Caruna, “InterpretML: Explainable Boosting Machines (EBMs) by Microsoft Research.”](https://www.zotero.org/google-docs/?5bqm8J) [↑](#footnote-ref-23)
25. [Caruna.](https://www.zotero.org/google-docs/?VCU26b) [↑](#footnote-ref-24)
26. [Caruna.](https://www.zotero.org/google-docs/?H5XAni) [↑](#footnote-ref-25)
27. [Michael Baiocchi, Jing Cheng, and Dylan S. Small, “Tutorial in Biostatistics: Instrumental Variable Methods for Causal Inference\*,” *Statistics in Medicine* 33, no. 13 (June 15, 2014): 2297–2340, https://doi.org/10.1002/sim.6128.](https://www.zotero.org/google-docs/?SNd49A) [↑](#footnote-ref-26)
28. [Ricardo Mora, “Instrumental Variables in Econometrics II,” *Department of Economics Universidad Carlos III de Madrid Master in Industrial Organization and Markets*, n.d.](https://www.zotero.org/google-docs/?1JsfJ5) [↑](#footnote-ref-27)
29. [Judea Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge, U.K. ; New York: Cambridge University Press, 2000); Artur Pokropek, “Introduction to Instrumental Variables and Their Application to Large-Scale Assessment Data,” *Large-Scale Assessments in Education* 4, no. 1 (February 24, 2016): 4, https://doi.org/10.1186/s40536-016-0018-2.](https://www.zotero.org/google-docs/?AwTqOr) [↑](#footnote-ref-28)
30. [J. D. Sargan, “The Estimation of Economic Relationships Using Instrumental Variables,” *Econometrica* 26, no. 3 (1958): 393–415, https://doi.org/10.2307/1907619.](https://www.zotero.org/google-docs/?oAyJaV) [↑](#footnote-ref-29)
31. [Shamal Chandra Karmaker et al., “Effects of Remittances on Renewable Energy Consumption: Evidence from Instrumental Variable Estimation with Panel Data,” *Utilities Policy* 83 (August 1, 2023): 101614, https://doi.org/10.1016/j.jup.2023.101614.](https://www.zotero.org/google-docs/?JUBccu) [↑](#footnote-ref-30)
32. [Christopher Baum, Mark Schaffer, and Steven Stillman, “IVREG2: Stata Module for Extended Instrumental Variables/2SLS and GMM Estimation,” Statistical Software Components, July 30, 2023, https://econpapers.repec.org/software/bocbocode/S425401.htm.](https://www.zotero.org/google-docs/?CVn4SV) [↑](#footnote-ref-31)
33. [Pokropek, “Introduction to Instrumental Variables and Their Application to Large-Scale Assessment Data.”](https://www.zotero.org/google-docs/?IH3Fzn) [↑](#footnote-ref-32)
34. [Baiocchi, Cheng, and Small, “Tutorial in Biostatistics.”](https://www.zotero.org/google-docs/?BDHdus) [↑](#footnote-ref-33)
35. [Pokropek, “Introduction to Instrumental Variables and Their Application to Large-Scale Assessment Data.”](https://www.zotero.org/google-docs/?FEzjo2) [↑](#footnote-ref-34)
36. [Pokropek.](https://www.zotero.org/google-docs/?zRbZdd) [↑](#footnote-ref-35)
37. "The central role of the propensity score in observational studies for ...." https://academic.oup.com/biomet/article/70/1/41/240879. [↑](#footnote-ref-36)
38. "Propensity score matching: a conceptual review for radiology ...." https://pubmed.ncbi.nlm.nih.gov/25741190[/](https://pubmed.ncbi.nlm.nih.gov/25741190/). [↑](#footnote-ref-37)
39. "Propensity score methods and their application in nephrology ...." https://pubmed.ncbi.nlm.nih.gov/21404223/. [↑](#footnote-ref-38)
40. "A Review of Propensity-Score Methods and Their Use in ... - PubMed." https://pubmed.ncbi.nlm.nih.gov/26315351/. [↑](#footnote-ref-39)
41. "Difference-in-Difference Estimation | Columbia Public Health." https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation. [↑](#footnote-ref-40)
42. "Differences-in-Differences." https://www.princeton.edu/~otorres/DID101.pdf.. [↑](#footnote-ref-41)
43. "What's trending in difference-in-differences? A synthesis of the ...." https://www.sciencedirect.com/science/article/pii/S0304407623001318. [↑](#footnote-ref-42)
44. "Empirical methods in the economics of education - ScienceDirect." https://www.sciencedirect.com/science/article/abs/pii/B978012815391800001X. [↑](#footnote-ref-43)
45. "Peer Review reports - BMC Medical Research Methodology." https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-021-01471-y/peer-review. [↑](#footnote-ref-44)