STAT31900 Introduction to Causal Inference

Final Project

ESG and Corporate Innovation: Evidence from an IV Approach

Seung Chul Lee

MS in Statistics

Winter 2022

Abstract

In this paper, I attempt to gain a causal understanding of the relationship between corporate ESG performance and innovative performance by using an instrumental variables (IV) approach. Under the LATE model with IPTW, becoming a high ESG firm lowers patents (patent citations) by 12.33% (16.87%). With the standard IV 2SLS model, increasing ESG score by 0.1 causes a firm to forgo about 3.88% (5.51%) of patents (patent citations). It is nominally translated into an average of 2.76 (1.75) patents (patent citations) using an IV model with the Poisson GLM. This is evidence of a competition for corporate resources between ESG activities and research and development.

Contents

1 Introduction

Environment, Social, and Governance (henceforth ESG) is a topic that has become increasingly popular in financial economics along with the hype about Socially Responsible Investment (SRI). Indeed, according to Broadbridge Financial Solutions, ESG assets will amount to $30 trillion by 2030[1]. To keep pace with the fast growing public interest, various aspects of ESG's effect on corporate finance has been covered in the existing literature. Scholars have debated the relationships including, but not limited to, ESG and a firm's ownership, financial risk, and firm value. (See Gillan et al. (2021) for a detailed review of the relevant literature). Despite the fact that most theories and empirical results remain inconclusive, research has been proliferating at an awesome speed. However, little is currently known about ESG's effect on a company's innovative caliber.

Innovation is another key theme, as it is essentially intertwined with economic growth and creation of jobs. In the famous words of Peter Drucker, "innovation is the specific instrument of entrepreneurship. The act that endows resources with a new capacity to create wealth." Hence, it is unquestionable that both concepts are exceptionally important when studying corporate behavior. A deeper understanding of the relationship will allow more informed firm management and economic policymaking to align incentives for the maximization of social welfare.

In this paper, I attempt to gain a causal understanding of the relationship between corporate ESG performance and innovative performance by using an instrumental variables (IV) approach. I employ both the Local Average Treatment Effect (LATE) method and the traditional econometric model to estimate the effect. As this is a final project for Introduction to Causal Inference course, I try to explore the pros and cons of each method. In particular, I study the impact of adding observable confounders to the model when the instrument may not be completely exogenous. I find that, in the absence of conditional independence, both LATE and econometric methods suggest that ESG score positively affects patent and citation counts. However, when controlling for confounders in the model, the effect of ESG on firm innovation is actually negative. This is an

[1] See https://www.broadridge.com/white-paper/asset-management/esg-and-sustainable-investment-outlook for the full article.

indication that estimation can be extremely misleading when exogeneity is not properly obtained in both methods. The results are robust to several different specifications of the structural equations, including a Poisson generalized linear model (GLM). The remainder of this paper is organized as follows. Section 2 develops and outlines the research design. Section 3 elaborates on the data collection process and provides summary statistics. Section 4 reports the empirical results. Section 5 concludes.

2 Research Design

There can be two competing hypotheses about the potential effect of ESG on corporate innovation. First, a high ESG performance may heighten a firm's innovative capacity. This idea can be reinforced by prior research that a high ESG performance lowers a firm's cost of capital. (El Ghoul et al., 2011; Chava, 2014; Ng and Rezaee, 2015, among others.) A decrease in the cost of capital will allow firms to allocate more funds to invest in projects that will potentially yield innovative outcome, as more investment opportunities become profitable. On the other hand, a high ESG performance may engender poor innovation efforts, as firms may be forgoing valuable investments to maintain a high ESG rating. (Hong et al., 2012; Di Giuli and Kostovetsky, 2014; Buchanan et al., 2018, among others.)

From the above discussion, I devise the following hypothesis to empirically test in this paper:

H_0: A high ESG performance will lead to more corporate innovation.

H_1 : A high ESG performance will lead to less corporate innovation.

In order to test the above hypothesis, I construct an instrumental variable model using both the semiparametric causal inference framework and the more traditional structural equations model. The instrument of my choice is firm visibility (Vis), which is measured as an indicator variable of whether the particular firm is a constituent of either the Dow Jones Industrials-30 index or the Nasdaq 100 index[2]. The rationale behind the instrument is that a more visible firm will

[2]Past studies have often employed the S&P 100 index, which has become unavailable on Wharton Research Data Services as of July 2020.

be more closely scrutinized by stakeholders, leading to a heightened level of attention paid to ESG metrics. However, there is little evidence to believe that stronger monitoring will affect corporate innovation. In fact, there is a wide variety of firms included in the indices. Both well-known technology firms, such as IBM mentioned in Section 3, and more traditional firms, such as Coca-Cola Company and Walmart Inc., belong to the visible group. These latter firms will not suddenly ramp up patent production just because they become constituents of the index, as patent development is closely related to the form of the business. Hence, the mere inclusion into a high-profile stock market index alone will not likely have a causal relationship with innovative performance.

Rather, I anticipate that the causal relations occur as depicted in Figure 1. If ESG performance has any causal effect on the level of corporate innovation, firm visibility will be an appropriate instrument. However, it is still possible that there are confounders which affect both firm visibility and corporate innovation. In this case, I can still obtain a causal estimate using conditional independence of firm visibility by including observable confounders. Although I have only specified financial health in particular, there can be many possible confounders affecting both the instrument and the response. For instance, an older, more established firm may be more visible and less innovative at the same time. The former comes from the fact that the firm has been in the market for a long time, whereas the latter will stem from decrease in profitable investment opportunities in the corporate lifecycle. Other factors, such as managerial ability, will be similar. Thus, I collect an assortment of relevant financial metrics and include them to the model. This includes return on asset, firm size, financial leverage, Tobin's q, cash flow ratio, firm age, dividend payout ratio, and R&D-to-sales ratio.

In this study, I try to obtain the average treatment effect (ATE) of having a high ESG performance on the degree of corporate innovation, given the observable confounders. To state it formally, it is

$$ATE = \mathbb{E}[Y(1) - Y(0)|\boldsymbol{X}].$$

where $Y(1)$ is the measure of corporate innovation for high ESG performing observations ("treated"),

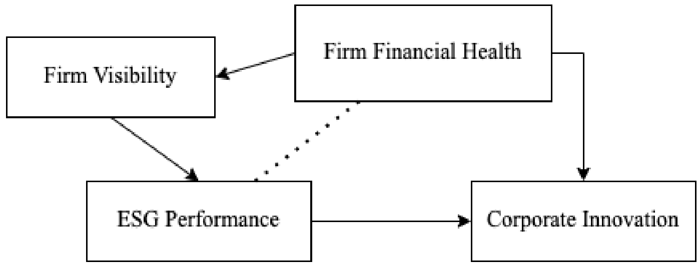

Figure 1. Conceptual Causal Relationships

$Y(0)$ is the measure of corporate innovation for low ESG performing observations ("not treated"), and $\boldsymbol{X}$ is the vector of observable confounders. This is under the assumption that

$$Y \perp\!\!\!\perp \boldsymbol{X}|Z.$$

where Z is my instrument, firm visibility. For the actual response variable, I use the number of patents registered and the number of citations on those patents as proxy variables. This is in line with the corporate finance literature regarding corporate innovation, such as Atanassov (2013). The number of patents represents the amount of corporate innovation, whereas the number of citations captures the quality of the generated patents.

The estimators that I use to find the estimand are the local average treatment effect (LATE) method developed in Rubin's causal framework. More specifically, I use the LATE method along with inverse probability of treatment weighting (IPTW) to attain conditional independence of the treatment and the instrument (Rubin, 1997; Hong, 2015). The formally stated estimator β is as follows:

$$\beta = \frac{\mathbb{E}[w_1 Y|Z=1] - \mathbb{E}[w_0 Y|Z=0]}{\mathbb{E}[w_1 D|Z=1] - \mathbb{E}[w_0 D|Z=0]} \tag{1}$$

where D is the treatment, w_1 is the weight for the treated group, and w_0 is the weight for the untreated group. In this model, $D_i = \mathbb{1}_{ESG_i > 0.5}$ because it is necessary to dichotomize the treatment variable in the RCM framework. The cut-off, which is the exact midpoint of the Refinitiv ESG score scale, yielded the highest correlation between the treatment variable and the instrument among different options. It also has an appealing intuitive property that one may view it as a pass-fail criterion in terms of ESG performance.

I also construct a standard econometric IV model, or a structural equations model, using two-stage least squares (2SLS) as follows:

$$D_{i,t-1} = \alpha_0 + \alpha_1 Z_{i,t-1} + \delta \boldsymbol{X}_{i,t-1} + \epsilon_D, \tag{2}$$

$$Y_{i,t} = \beta_0 + \beta_1 D_{i,t-1} + \gamma \boldsymbol{X}_{i,t-1} + \epsilon_{YD} \tag{3}$$

where Equation (2) is the first-stage regression and Equation (3) is the second-stage regression. Note that I use lagged variables to reduce simultaneity (or reverse causality) problems.

I now discuss the appropriateness of the model specified. The instrumental variables design is adequate under the following assumptions (Angrist et al., 1996):

- Exclusion condition.
- Exogeneity of the instrument.
- Efficiency of the instrument.
- Monotonicity assumption (for LATE).
- Homogeneity of effects or no correlation between effects of Z to D and D to Y (for standard IV).

The exclusion assumption is likely to be satisfied as discussed above. The exogeneity assumption is not clear, but a conditional independence can be claimed with the inclusion of observable confounders, i.e., $Y \perp\!\!\!\perp \boldsymbol{X} | Z$. Efficiency of the instrument is a potential weakness of my design, as the correlation between the instrument and the treatment variable was merely 0.3. In my defense, it is very difficult to observe a high degree of correlation between any two variables in the corporate finance literature. Nevertheless, it is worthwhile to note that such lack of efficiency may inflate standard errors in the subsequent analyses. As for monotonicity, it has to be assumed with no concrete way to test it. However, this does not seem to be prohibitively strong: a firm that is a defier in this case will be a firm that has a high ESG performance when less visible but abruptly decreases ESG performance when they become more visible. Intuitively, it is quite unlikely that a firm will behave in this manner. Homogeneity of effects or no correlation of effects is obscure, and it is possible to have heterogeneous effects or correlated effects. This is mainly the reason I try both approaches as complements.

3 Data

I collect the necessary data from a variety of sources to assemble my data set of U.S. companies. Namely, I consult the following databases:

- CRSP-COMPUSTAT for basic financial and accounting variables, such as firm size, age, total assets, market value, etc.
- Thomson-Refinitiv ESG for ESG overall scores.
- Wharton Research Data Services (WRDS) U.S. Patent (Beta) for patent and citation counts.
- CRSP Indices for index constituency data.

I merge the CRSP-COMPUSTAT data to Refinitiv ESG data using 8-digit CUSIP number. I then add patent data based on GVKEY provided in the WRDS Patent Link table. Finally, index constituency dummy variables are created for each fiscal year and matched to the data set also using GVKEY. I remove firms that have missing values for the covariates, which includes total assets, leverage, Tobin's q, R&D-to-Sales ratio, return on assets (ROA), dividend payout ratio, EBITDA/Sales ratio, and ESG overall score. I also remove financial firms (SIC codes 6000-6999) and utilities (SIC codes 4000-4999), as they are influenced extensively by government regulations. I treat firm-year observations with missing patent or citation counts as zero, assuming these firms are not in the database because they did not file anything. All continuous variables are winsorized at the 1$^{\text{st}}$ and 99$^{\text{th}}$ percentiles to remove excessive leverage points.

After the above process, I am left with a panel of 14,625 firm-year observations. The time frame is from 2011 to 2019, as the beta version of the WRDS US Patent data set covers only this period. There are a total of 3,288 unique US companies included. Summary statistics of the data are provided in Table 1. The detailed definitions of each variable is provided in the appendix.

Table 1. Summary Statistics

	Mean	SD	Min	1Q	Median	3Q	Max
ESG	0.480	0.309	0.027	0.189	0.401	0.805	0.979
ESG_Dummy	0.495	0.500	0.000	0.000	0.000	1.000	1.000
Vis	0.085	0.279	0.000	0.000	0.000	0.000	1.000
Pat	38.517	257.474	0.000	0.000	1.000	9.000	9257.000
Cit	179.031	1558.004	0.000	0.000	0.000	14.000	75612.000
Size	8.203	1.781	1.797	7.121	8.235	9.396	12.298
Lev	0.595	0.256	0.029	0.427	0.596	0.765	1.722
ROA	0.082	0.187	-1.610	0.046	0.107	0.158	0.469
Q	1.750	1.659	0.073	0.795	1.254	2.111	11.845
Cashflow	-0.765	6.077	-59.088	0.094	0.182	0.317	0.789
lnAge	2.733	0.969	-0.003	2.113	2.946	3.475	4.059
Payout	0.257	0.668	-3.464	0.000	0.098	0.384	5.060
RDratio	0.021	0.056	0.000	0.000	0.000	0.019	0.857

The most notable aspect of the summary statistics is that the patents and citations display an extremely skewed distribution. Most of the firms take on the value of zero, whereas a handful of extraordinarily innovative firms have astoundingly large values. Namely, the maximum value of patent count is attained by IBM in 2017. A common approach for using the skewed patent and citation counts as the response is employing the log transformation after adding one to the counts (Atanassov, 2013, for example). In addition to following this fashion, I also try fitting a generalized linear model (GLM) with Poisson link. This may be more appropriate for count data as suggested by Agresti (2015) to check robustness.

Table 2. Contingency Table for Instrument and Treatment

	$Vis = 0$	$Vis = 1$
$ESG\_Dummy = 0$	7111	279
$ESG\_Dummy = 1$	6272	963

4 Empirical Results

4.1 Local Average Treatment Effect

Table 2 supplies the contingency table of the response, ESG_Dummy, and the instrument, Vis. For the sake of brevity, I exchange ESG_Dummy with D throughout this subsection. Note that when $Vis = 1$, the proportion of firms with high ESG scores is much higher (approximately 77.53%). This outcome is consistent with the theoretical prediction that a firm's visibility will entail more efforts toward ESG. However, the group with $Vis = 0$ is more evenly divided. That is, the probability of observing a high ESG performance in $Vis = 0$ group is almost equivalent to a fair coin toss. This makes intuitive sense but impairs our ability to extend the analysis across many firms. Assuming there are no defiers, the proportion of compliers in this setup is $0.775 - 0.469 = 0.531 - 0.225 = 0.306$. Because only a handful of firms belong to stock indices, the proportion of visible firms is quite low. In turn, the estimated proportion of compliers is also somewhat marginalized. Under the Rubin's Causal Model framework, the LATE is only applicable to this very small portion of the sample.

Table 3. Prima Facie Effects

	$Y = log(Pat + 1)$	$Y = log(Cit + 1)$
(Intercept)	0.8256***	0.9267***
	(0.0191)	(0.0249)
D	1.0325***	1.1532***
	(0.0272)	(0.0354)

*, **, *** denote statistical significance at 10%, 5%, and 1% levels, respectively.

Table 4. Local Average Treatment Effect

	$Vis = 1$	$Vis = 0$
Panel A. Patent Counts		
$\bar{Y}_i$	3.117	1.1711
$\bar{D}_i$	0.7754	0.4687
LATE		6.3444***
SE		(0.3021)
Panel B. Citation Counts		
$\bar{Y}_i$	3.8276	1.2808
$\bar{D}_i$	0.7754	0.4687
LATE		8.3035***
SE		(0.4029)

*, **, *** denote statistical significance at 10%, 5%, and 1% levels, respectively.

I first run a simple linear regression of the treatment on the two response variables, reported in Table 3. The resulting *prima facie* effect of treatment for patent count is approximately 1.0325, obtained by regressing the response Y_i on the treatment D_i. As for citation count, the effect is approximately 1.1532. Note that both coefficient estimates are positive and statistically significant at the 1% level. This can be translated into an increase of $e^{1.0325}$ ($e^{1.1532}$) or approximately 1.03% (1.15%)[3] in patent (patent citation) counts. On the surface, corporate innovation performance and ESG performance seem positively correlated. Of course, it is naive to interpret the *prima facie* estimate as the causalilty without imposing strong assumptions. In Table 4, I present the results

[3] The percentage interpretation is possible due to the following relationship:

$$\begin{aligned}\Delta \log y &= \log(y + \Delta y) - \log y \\ &= \log(\frac{y + \Delta y}{y}) \\ &= \log(1 + \frac{\Delta y}{y}) \\ &\approx \frac{\Delta y}{y}\end{aligned}$$

of estimating the LATE without IPT weighting. Note that, much like the *prima facie* effects, the LATE is positive and significant. It indicates that becoming a high ESG performer has a more notable effect of 6.34% (8.30%) increase in patents (patent citations).

Table 5. LATE with IPTW

	$Y = \log(Pat + 1)$	$Y = \log(Cit + 1)$
(Intercept)	7.0936***	9.399***
	(1.7628)	(2.466)
D	-12.3274***	-16.8743***
	(4.2369)	(5.798)

*, **, *** denote statistical significance at 10%, 5%, and 1% levels, respectively.

However, as shown in Table 5, using IPTW to obtain conditional independence engenders a strikingly different result. The estimated causal effect becomes negative and significant, with a larger absolute value than predicted for the *prima facie* and unadjusted LATE estimators. This outcome demonstrates that there is a substantial bias in the *prima facie* effect and the unweighted local average treatment effect. Hence, on the surface, it may appear that the ESG performance has a positive correlation with innovation, but it is mainly due to confounding factors like firm size or financial health that are associated with both. When such confounders are controlled for, the effect of being a high ESG firm reduces patent (patent citations) by about 12.33% (16.87%).

4.2 Structural Equations Method

In this subsection, I use the structural equations method to complement the LATE approach. Similarly to the previous part, I first estimate the *prima facie* effect, without using instrumental variables. This is essentially equivalent to estimating Equation (3) as a stand-alone regression model. Table 6 presents the results of these regressions. Panel A uses $\log(Pat+1)$ as the response, whereas Panel B employs $\log(Cit+1)$. Column (1) uses ESG_{t-1} as the sole predictor. Column (2) includes $\boldsymbol{X}$ as additional regressors. Column (3) further includes industry fixed effects, based on two-digit SIC codes. Column (4) incorporates both industry and year fixed effects in the model. Table 6 provides the coefficient estimates and their standard errors in parentheses.

Table 6. Structural Equations - Prima Facie Effects

Variables	(1)	(2)	(3)	(4)
Panel A. Patent Counts				
ESG_{t-1}	2.0195***	1.2341***	0.5977***	0.6066***
	(0.043)	(0.0523)	(0.0468)	(0.0474)
ROA_{t-1}	-	0.8347***	0.0624	0.052
	-	(0.0865)	(0.0737)	(0.0739)
Q_{t-1}	-	0.2511***	0.1546***	0.1541***
	-	(0.0082)	(0.0072)	(0.0072)
Lev_{t-1}	-	-0.9164***	-0.3877***	-0.381***
	-	(0.0512)	(0.0453)	(0.0454)
$Size_{t-1}$	-	0.2654***	0.3974***	0.3927***
	-	(0.0101)	(0.009)	(0.0094)
$Payout_{t-1}$	-	-0.1134***	-0.0256*	-0.026*
	-	(0.0182)	(0.015)	(0.015)
$Cashflow_{t-1}$	-	0.0185***	0.0072***	0.0072***
	-	(0.0025)	(0.0021)	(0.0021)
$lnAge_t$	-	0.1299***	0.1019***	0.1008***
	-	(0.0139)	(0.0117)	(0.0117)
$RDratio_{t-1}$	-	11.6464***	5.0786***	5.0632***
	-	(0.2503)	(0.2226)	(0.223)
Fixed Effects	None	None	Industry	Industry & Year
Panel B. Citation Counts				
ESG_{t-1}	2.334***	1.1008***	0.3268***	0.5747***
	(0.0561)	(0.0699)	(0.0668)	(0.0644)
ROA_{t-1}	-	1.2923***	0.293***	0.0364
	-	(0.1154)	(0.1051)	(0.1005)
Q_{t-1}	-	0.3083***	0.2011***	0.1828***
	-	(0.011)	(0.0103)	(0.0098)
Lev_{t-1}	-	-1.2265***	-0.6499***	-0.51***
	-	(0.0683)	(0.0646)	(0.0617)
$Size_{t-1}$	-	0.3911***	0.5546***	0.4237***
	-	(0.0135)	(0.0128)	(0.0128)
$Payout_{t-1}$	-	-0.1384***	-0.0407*	-0.042**
	-	(0.0243)	(0.0214)	(0.0204)
$Cashflow_{t-1}$	-	0.0159***	0.0022	0.003
	-	(0.0033)	(0.003)	(0.0028)
$lnAge_t$	-	0.152***	0.118***	0.0948***

(cont.)

Variables	(1)	(2)	(3)	(4)
	-	(0.0186)	(0.0166)	(0.0159)
$RDratio_{t-1}$	-	12.3326***	5.2326***	5.0192***
	-	(0.3342)	(0.3176)	(0.3032)
Fixed Effects	None	None	Industry	Industry & Year

*, **, *** denote statistical significance at 10%, 5%, and 1% levels, respectively.

Note that the *prima facie* effects estimated using the continuous ESG score with the confounding variables are also positively significant. This is consistent with the result I find with the LATE. Controlling for other covariates and fixed effects reduces the magnitude of the coefficient estimate, which is expected due to collinearity. I now estimate Equations (2) and (3) to obtain the standard IV estimator, the result of which is provided in Table 7. Again, as with the outcome of LATE estimation with IPTW, the coefficient estimates for ESG_{t-1} become negatively significant. This suggests that the homogeneity assumption may not be severely violated, as both LATE and standard IV offer qualitatively equivalent results. Of course, the estimates themselves differ in terms of interpretation. Focusing on column (4) as the full model, the regression coefficient has the meaning that the rather unlikely jump from ESG overall score of 0 (the absolute minimum) to 1 (the absolute maximum) will decrease number of patents (patent citations) by around 38.79% (55.10%). To make inference in a more realistic manner, an increase of 0.1 points in overall ESG score causes a decrease in patents (patent citations) of approximately 3.88% (5.51%). From the above, I find a strong evidence for H_1, in which case I may conclude that ESG causes firms to forgo innovative research funding.

Table 7. Structural Equations - IV Effects

Variables	(1)	(2)	(3)	(4)
Panel A. Patent Counts				
ESG_{t-1}	8.7464***	-33.9535***	-39.4477***	-38.7907***
	(0.3501)	(9.1677)	(14.1562)	(13.591)
ROA_{t-1}	-	7.0992***	2.1058**	2.3025**
	-	(1.703)	(0.8945)	(0.9315)

(cont.)

Variables	(1)	(2)	(3)	(4)
Q_{t-1}	-	1.2628***	0.8799***	0.9142***
	-	(0.2675)	(0.2615)	(0.2669)
Lev_{t-1}	-	-4.324***	0.3242	0.2367
	-	(0.9333)	(0.4106)	(0.3815)
$Size_{t-1}$	-	3.9733***	4.9255***	5.0429***
	-	(0.9673)	(1.6015)	(1.6051)
$Payout_{t-1}$	-	0.2566*	0.4491**	0.4575**
	-	(0.1409)	(0.1992)	(0.1968)
$Cashflow_{t-1}$	-	-0.0234	-0.0184	-0.0192
	-	(0.0178)	(0.0174)	(0.0171)
$lnAge_t$	-	2.8618***	2.6502***	2.6261***
	-	(0.7157)	(0.9045)	(0.8747)
$RDratio_{t-1}$	-	36.6339***	16.9401***	15.9661***
	-	(6.6589)	(4.4848)	(4.0686)
Fixed Effects	None	None	Industry	Industry & Year
Panel B. Citation Counts				
ESG_{t-1}	11.4471***	-47.0628***	-57.1957***	-55.1009***
	(0.4676)	(12.539)	(20.3317)	(19.1918)
ROA_{t-1}	-	9.867***	3.2282**	3.2167**
	-	(2.3292)	(1.2847)	(1.3153)
Q_{t-1}	-	1.693***	1.243***	1.2569***
	-	(0.3659)	(0.3755)	(0.3769)
Lev_{t-1}	-	-5.8907***	0.3726	0.363
	-	(1.2765)	(0.5898)	(0.5387)
$Size_{t-1}$	-	5.4665***	7.0589***	6.9953***
	-	(1.323)	(2.3002)	(2.2665)
$Payout_{t-1}$	-	0.3681*	0.6411**	0.6413**
	-	(0.1927)	(0.2861)	(0.2779)
$Cashflow_{t-1}$	-	-0.0414*	-0.0345	-0.0343
	-	(0.0244)	(0.025)	(0.0241)
$lnAge_t$	-	3.8914***	3.7785***	3.6635***
	-	(0.9789)	(1.299)	(1.2352)
$RDratio_{t-1}$	-	46.5348***	22.2708***	20.4271***
	-	(9.1075)	(6.4413)	(5.7453)
Fixed Effects	None	None	Industry	Industry & Year

*, **, *** denote statistical significance at 10%, 5%, and 1% levels, respectively.

As a further check, I try using the binary variable D to build upon the LATE result. The results are reported in Table 8. Excluding column (1), all the coefficient estimates for D are negative. This is again consistent with the previous findings. However, column (3) and (4) with fixed effects display extremely large standard errors and thus statistically insignificant estimates. The main reason for this finding is that, because D is binary, D is inevitably highly correlated with the fixed effects. Therefore, the regression coefficient obtains a large standard error due to identifiability issues. To elaborate, D and the fixed effects are all binary variables in nature. If a firm belongs to $D = 1$ group for multiple time periods (which is very likely), then it will enlarge both the correlation of D with the industry that this particular firm belongs to. A similar argument can easily be formulated for the serial correlation in time. This is not a problem with the continuous ESG score, as it captures more granular fluctuations in the score. Then, the correlation structure with the fixed effects will be less acute, as there is some variation within the firm over time and dispersion of scores in firms within the same industry. Because I observe a negatively significant coefficient without the fixed effects, I argue that the insignificance of coefficients in columns (3) and (4) is not an evidence against my previous findings.

Table 8. Structural Equations - IV with Dummy

Variables	(1)	(2)	(3)	(4)
Panel A. Patent Counts				
D_{t-1}	6.3444***	-36.2961**	-68.9921	-92.2465
	(0.3021)	(18.448)	(79.735)	(141.0008)
ROA_{t-1}	-	8.9821**	1.1018	3.6407
	-	(4.1319)	(2.0607)	(5.8783)
Q_{t-1}	-	1.6097**	1.6246	2.1851
	-	(0.6779)	(1.6944)	(3.0941)
Lev_{t-1}	-	-5.9434**	2.1044	1.8525
	-	(2.5524)	(3.0523)	(3.6691)
$Size_{t-1}$	-	5.5942**	11.2159	15.5916
	-	(2.6439)	(12.4261)	(23.1235)
$Payout_{t-1}$	-	0.1279	0.5128	0.7357
	-	(0.2256)	(0.7045)	(1.2397)
$Cashflow_{t-1}$	-	-0.0093	0.003	-0.0033
	-	(0.0297)	(0.0481)	(0.065)
$lnAge_t$	-	3.7391**	5.194	6.9037
	-	(1.7914)	(5.8469)	(10.3447)
$RDratio_{t-1}$	-	45.1404***	22.7594	27.6951
	-	(16.7863)	(20.8682)	(34.9975)
Fixed Effects	None	None	Industry	Industry & Year

(cont.)

Variables	(1)	(2)	(3)	(4)
Panel B. Citation Counts				
D_{t-1}	8.3035***	-50.31**	-100.0326	-131.0332
	(0.4029)	(25.4073)	(115.3089)	(200.1387)
ROA_{t-1}	-	12.4769**	1.7724	5.1177
	-	(5.6906)	(2.98)	(8.3438)
Q_{t-1}	-	2.1739**	2.3227	3.0621
	-	(0.9337)	(2.4504)	(4.3919)
Lev_{t-1}	-	-8.1353**	2.9538	2.6582
	-	(3.5153)	(4.4141)	(5.208)
$Size_{t-1}$	-	7.7131**	16.1795	21.9794
	-	(3.6413)	(17.9701)	(32.8219)
$Payout_{t-1}$	-	0.1897	0.7333	1.0364
	-	(0.3107)	(1.0189)	(1.7597)
$Cashflow_{t-1}$	-	-0.0218	-0.0035	-0.0118
	-	(0.041)	(0.0695)	(0.0922)
$lnAge_{t}$	-	5.1075**	7.4668	9.7397
	-	(2.4671)	(8.4555)	(14.6834)
$RDratio_{t-1}$	-	58.3257**	30.7084	37.0877
	-	(23.1188)	(30.1786)	(49.676)
Fixed Effects	None	None	Industry	Industry & Year

*, **, *** denote statistical significance at 10%, 5%, and 1% levels, respectively.

4.3 Robustness Checks

As briefly mentioned in Section 3, I construct a Poisson GLM to test the robustness of my results. Since structural equations model is quite sensitive to the regression specification and the underlying assumptions of least squares, it is useful to try a variation of the model. Also, using a GLM relaxes linear model assumptions such as constant variance of the error term, as the link function can be chosen separately from the linear relationship. This is particularly useful for the data at hand because both patents and patent citations are count data. I have taken the logarithm after adding one to these variables, but this may not produce a response that is both linear in predictors and homoskedastic.

Table 9. IV with Generalized Linear Model

Variables	(1)	(2)	(3)	(4)
Panel A. Patent Counts				
ESG_{t-1}	13.1229***	-39.5494***	-29.4171**	-27.6554***
	(0.5774)	(11.1036)	(11.4655)	(10.7344)
ROA_{t-1}	3.8238***	8.5299***	1.5732*	1.4372*
	(0.1651)	(2.1269)	(0.8508)	(0.8358)
Q_{t-1}	-	1.4615***	0.7372***	0.7662***
	-	(0.3247)	(0.2098)	(0.208)
Lev_{t-1}	-	-5.0116***	0.8116*	0.7933**
	-	(1.0908)	(0.4197)	(0.3899)
$Size_{t-1}$	-	5.0488***	4.2345***	4.1904***
	-	(1.1686)	(1.2972)	(1.267)
$Payout_{t-1}$	-	0.2503	0.3671**	0.3681**
	-	(0.1744)	(0.1658)	(0.1607)
$Cashflow_{t-1}$	-	-0.0473*	-0.047***	-0.0453***
	-	(0.0258)	(0.0159)	(0.0152)
$lnAge_t$	-	3.5215***	2.2758***	2.1742***
	-	(0.8783)	(0.7427)	(0.7015)
$RDratio_{t-1}$	-	38.5196***	14.8509***	13.85***
	-	(7.9776)	(3.6606)	(3.2539)
Fixed Effects	None	None	Industry	Industry & Year
Panel B. Citation Counts				
ESG_{t-1}	11.8412***	-26.4942***	-26.3124**	-17.4589**
	(0.7134)	(8.9985)	(11.0914)	(7.9401)
ROA_{t-1}	2.899***	9.2716***	6.1174***	2.0029**
	(0.2188)	(1.7095)	(1.2991)	(0.8162)
Q_{t-1}	-	0.9327***	0.5113**	0.7071***
	-	(0.2739)	(0.2129)	(0.159)
Lev_{t-1}	-	-4.8558***	-0.284	0.1258
	-	(0.8821)	(0.4827)	(0.4043)
$Size_{t-1}$	-	3.6264***	3.8536***	3.0796***
	-	(0.9397)	(1.2488)	(0.9331)
$Payout_{t-1}$	-	0.0489	0.2268	0.2387*
	-	(0.129)	(0.1626)	(0.1324)
$Cashflow_{t-1}$	-	-0.0332	-0.0663***	-0.0546***
	-	(0.0235)	(0.0179)	(0.014)
$lnAge_t$	-	2.2258***	1.6994**	1.1341**
	-	(0.762)	(0.7492)	(0.5507)
$RDratio_{t-1}$	-	28.9305***	14.2445***	11.8069***
	-	(6.4423)	(3.5129)	(2.5063)
Fixed Effects	None	None	Industry	Industry & Year

(cont.)

Variables	(1)	(2)	(3)	(4)

*, **, *** denote statistical significance at 10%, 5%, and 1% levels, respectively.

Table 9 presents the outcome of the IV regression with a Poisson GLM. Note that the outcome is very similar to that of the original linear 2SLS model. The coefficient estimates are smaller in absolute value, which stems from the fact that the Poisson GLM models the actual counts, and no longer has the interpretation of percentage change. That is, using the values of column (4) as the full model, increase of 0.1 in ESG score leads to a decrease of approximately 2.76 (1.75) patents (patent citations). This is another way that the GLM has a more intuitive appeal for this data.

Table 10. Black et al. Endogeneity Test

Variable	$D = 0$	$D = 1$
Panel A. Patent Counts		
(Intercept)	0.7931***	1.5998***
	(0.0146)	(0.0235)
Vis_{t-1}	0.861***	1.9411***
	(0.0749)	(0.0644)
Panel B. Citation Counts		
(Intercept)	0.8718***	1.7446***
	(0.02)	(0.0297)
Vis_{t-1}	1.452***	2.5187***
	(0.1027)	(0.0814)

*, **, *** denote statistical significance at 10%, 5%, and 1% levels, respectively.

Finally, I check the viability of the IV design using the Black et al. endogeneity test. Two separate regressions of the response on the instrument is run for the non-treated group ($D = 0$) and the group with ($D = 1$). The results are shown in Table 10. All coefficients on Vis_{t-1} is significant, indicating difference in treatment effects for each group. Hence, there is endogeneity in the setting and IV approach is necessary.

5 Conclusion and Discussion

From the study, I find that ESG performance has an adverse causal effect on corporate innovation. Under the LATE model with IPTW, becoming a high ESG firm lowers patents (patent citations) by 12.33% (16.87%). With the standard IV 2SLS model, increasing ESG score by 0.1 causes a firm to forgo about 3.88% (5.51%) of patents (patent citations). It is nominally translated into an average of 2.76 (1.75) patents (patent citations) using an IV model with the Poisson GLM. This is consistent with claims that the allocation of capital towards enhancing ESG scores creates a competition for finite corporate resources, as suggested by Hong et al. (2012), Di Giuli and Kostovetsky (2014), Buchanan et al. (2018), and others.

One very important caveat is that my results cannot be an evidence towards the adverse nature of ESG activity. Emphasis, and relevant regulations, on corporate ESG practices is inevitable for maximizing the aggregate social utility. However, it does provide a key insight that policymakers should be take notice of: such a trade-off between ESG and innovation exists within the firm. Hence, a wise and effective policy will take this factor into account and minimize the adverse impact on innovation. Moreover, this research casts doubt upon the current ESG scoring scheme. Currently, none of the major ESG ratings incorporate factors such as registering patents that are environmentally friendly or socially helpful. In order to encourage innovation in this direction, it would be sensible to include such criteria.

Lastly, another key takeaway is that my results are drastically different from a naive estimation of the *prima facie* effect, which reemphasizes the necessity of a proper, careful approach in causal inference. This lesson is perhaps of the most significance for the purpose of this course.

Appendices

Table A1. Variable Definitions

Variable	Definition	Description
ESG	Thomson/Refinitiv ESG Score	Continuous variable between 0 and 1.
ESG_Dummy	$= \mathbb{1}_{ESG>0.5}$	Binary variable.
Vis	$= \mathbb{1}_{\text{included in stock index}}$	Binary variable.
Pat	Number of patents registed	Nonnegative integer variable.
Cit	Number of citations on registered patents	Nonnegative integer variable.
Size	Firm size = log(Total Assets)	Continuous variable.
Lev	Financial leverage $= \frac{\text{Total Liabilities}}{\text{Total Assets}}$	Continuous variable between 0 and 1.
ROA	Return on Asset $= \frac{\text{Operating Income}}{\text{Total Assets}}$	Continuous variable between -1 and 1.
Q	Tobin's $q = \frac{\text{Market Capitalization+Long-term Debt}}{\text{Total Assets}}$	Continuous variable.
Cashflow	Cash flow ratio $= \frac{\text{EBITDA}}{\text{Total Revenue}}$	Continuous variable.
lnAge	Logarithm of age = log(Firm age)	Continuous variable.
Payout	Dividend payout ratio $= \frac{\text{Cash Dividends}}{\text{Net Income}}$	Continuous variable between 0 and 1.
RDratio	R&D-to-sales ratio $= \frac{\text{R\&D Expenses}}{\text{Total Revenue}}$	Continuous variable between 0 and 1.

References

Agresti, A. (2015). *Foundations of linear and generalized linear models*. Wiley.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

Atanassov, J. (2013). Do hostile takeovers stifle innovation? evidence from antitakeover legislation and corporate patenting. *The Journal of Finance*, 68(3):1097–1131.

Buchanan, B., Cao, C. X., and Chen, C. (2018). Corporate social responsibility, firm value, and influential institutional ownership. *Journal of Corporate Finance*, 52(C):73–95.

Chava, S. (2014). Environmental externalities and cost of capital. *Management Science*, 60(9):2223–2247.

Di Giuli, A. and Kostovetsky, L. (2014). Are red or blue companies more likely to go green? politics and corporate social responsibility. *Journal of Financial Economics*, 111(1):158–180.

El Ghoul, S., Guedhami, O., Kwok, C. C., and Mishra, D. (2011). Does corporate social responsibility affect the cost of capital? *Journal of Banking and Finance*, 35(9):2388–2406.

Gillan, S. L., Koch, A., and Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, 66:101889.

Hong, G. (2015). Causality in a social world: Moderation, meditation and spill-over.

Hong, H., Kubik, J. D., and Scheinkman, J. A. (2012). Financial constraints on corporate goodness. Working Paper 18476, National Bureau of Economic Research.

Ng, A. C. and Rezaee, Z. (2015). Business sustainability performance and cost of equity capital. *Journal of Corporate Finance*, 34(C):128–149.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, 127(8):757.